Sri Lanka Institute of Information Technology

Data Warehousing and Business Intelligence IT3021

B.Sc. (Hons) in Information Technology
DATA SCIENCE

Contents

Declaration

I declare that this project report or part of it was not a copy of document done by any organization, university and other institute or a previous student project at SLIIT and was not copied from the internet or other resources.

Student Detail

Registration Number : IT19080840

Name in Full : K.Kovishwakarunya

Batch : Y3S1.15 Weekday

Acknowledgement

I would like to express my sincere gratitude to the Lecturer in charge Ms.Lumini Wickramasinghe , Mr. Oshada Senaweera for the constant guidance and supervision provided to complete this assignment succesfully.

A special gratitude goes to Mr.Sheron Dinushka for guiding me with industrial best practices and knowledge while completing this assignment.

Finally, a special thanks to Sri Lanka Institute of Information Technology for taking measures to provide continous support in carrying out this specialization amidst this pandemic situation too.

1.0 Data Source for Assignment 2

The Data Warehouse created for Assignment 1 is the main data source used to carry out the necessary steps of the task. SQL Server Management Studio 2016 and SQL Server Data tools 2015 (Visual Studio) are the major tools used in Source data set compilation and ETL orchestration in order to create the Data Warehouse.

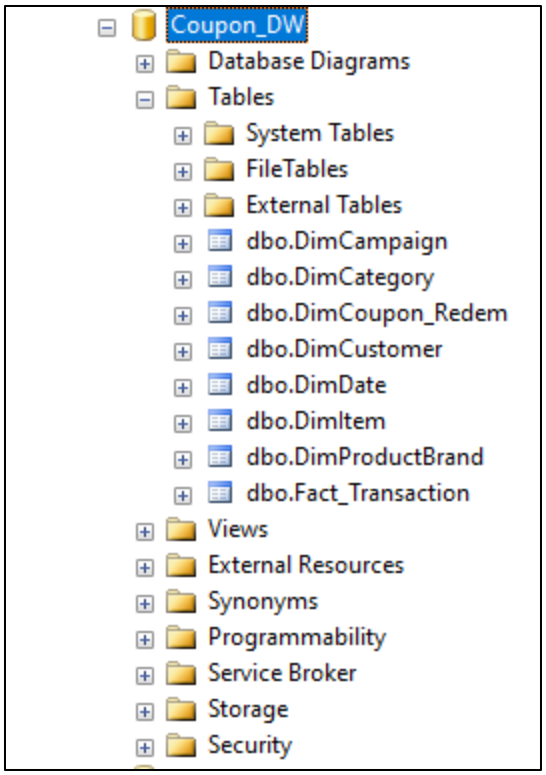


Figure 1 Data Source table structure

The Data Warehouse represents the data of a Brick-and-Mortar Retail shop that uses discount marketing as their major strategy in carrying out successful business operations. By analyzing the business entities, a snowflake schema was designed in order to carry out efficient Data Warehousing and Business Intelligence principles in the business process.

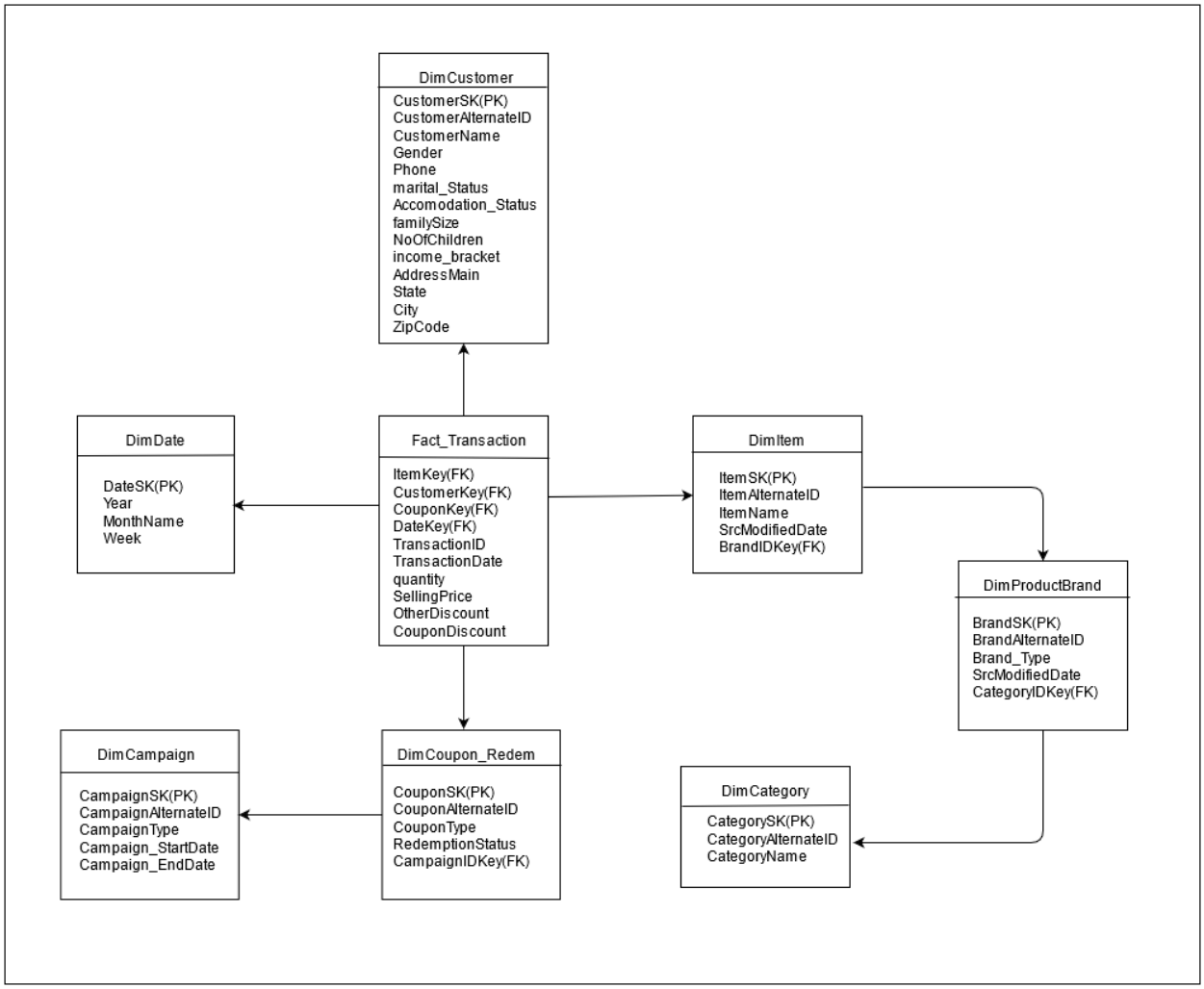


Figure 2 Datawarehouse SNOWFLAKE schema

The schema consists of one fact table and seven-dimension table including date dimension. It is observed that the entities are in a normalized form because of the snowflake design.

Hierarchical implementations are found in this schema
1. DimProductBrand and DimCategory are hierarchichal dimensions of DimItem
2. DimCustomer has a Customer address wise hierarchy.

DimCustomer table is **a slowly changing dimension** with historical attributes and changing attributes where Type 2 and Type 1 implementations are being enforced, respectively. Transaction of a customer for a particular date is considered as the **grain** of the Fact_Transaction fact table.

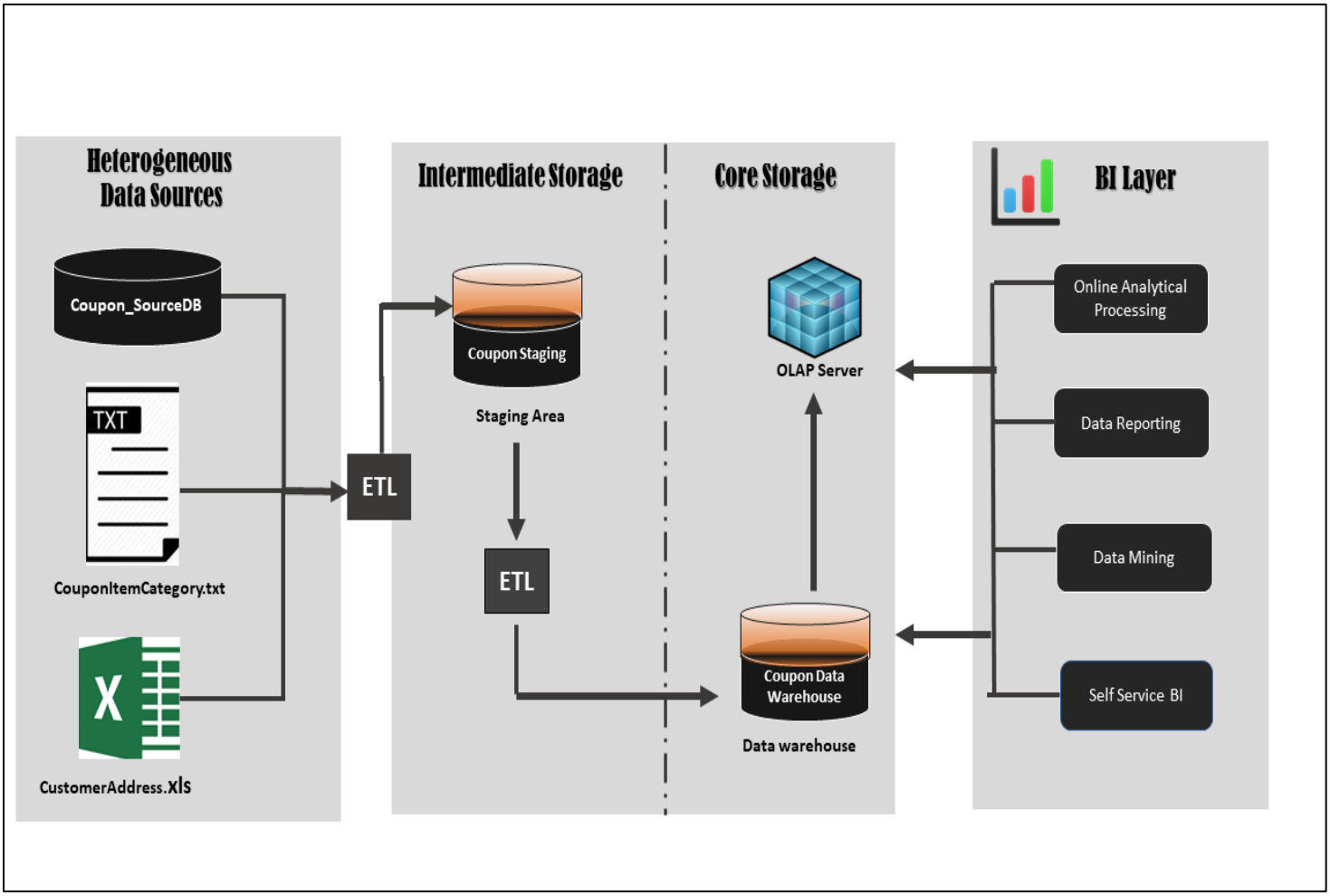


Figure 3 Solution Architecture

The datawarehouse created is used as a source to create an OLAP Cube which helps to carry out efficient business intelligence principles improving the performance of business.

2.0 SSAS Cube Implementation

SQL server data tools 2015 was used to create an ***Analysis Services Multidimensional and Data Mining Project*** in order to create the Cube Structure with the Sourcedata.(Datawarehouse)

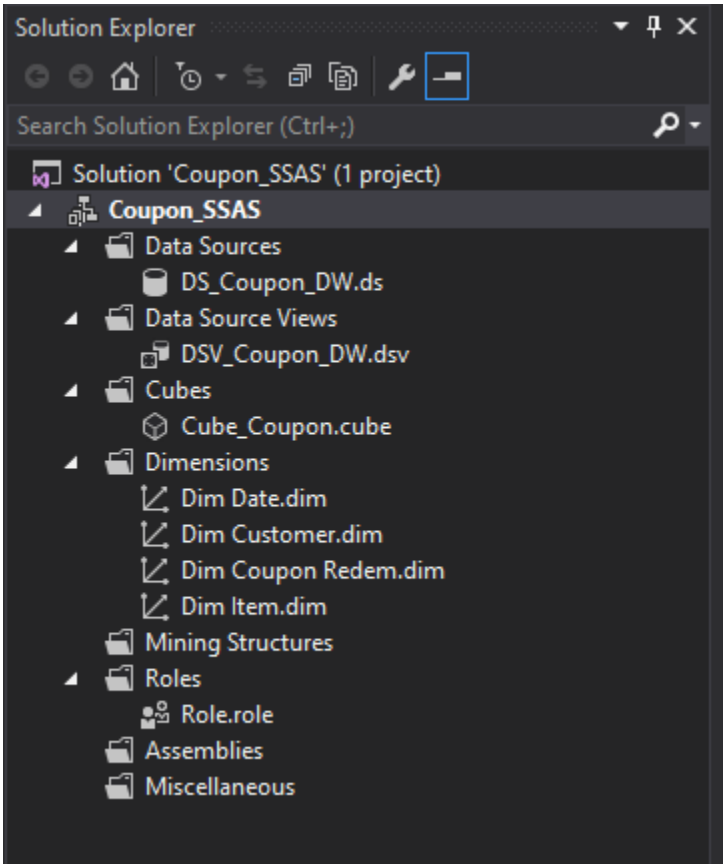


Figure 4 Cube structure Solution explorer

In order to create a working SSAS cube as shown above , sequence of tasks were carried out such as

1. Data source Creation**. (DS_Coupon_DW)**
2. Data Source View Creation**.(DSV_Coupon_DW)**
3. Cube creation**.(Cube_Coupon)**

2.1 Data Source View

Once the data source was created a data source view was created by making use of the Source data warehouse Coupon_DW.

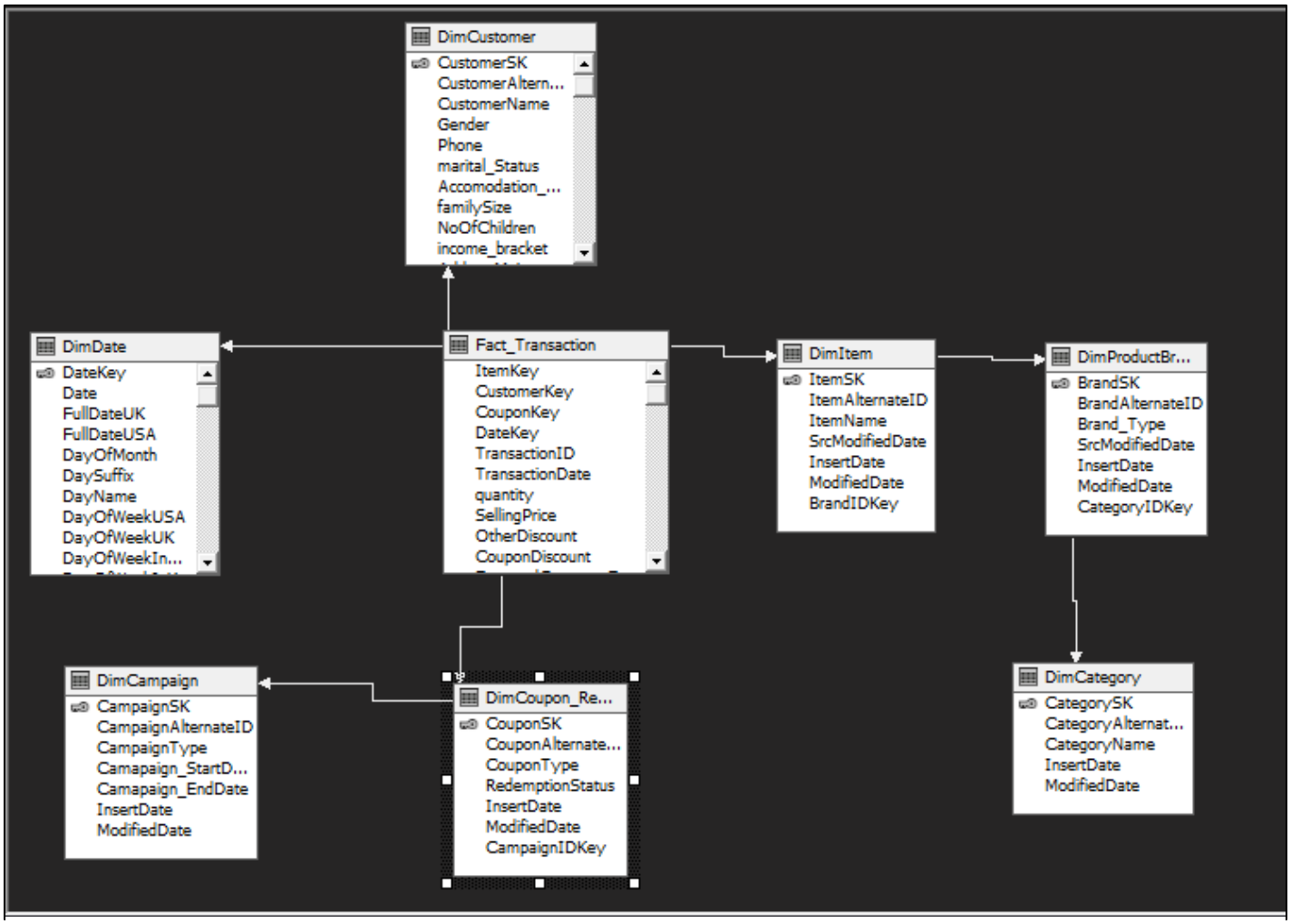


Figure 5 Data Source view (DSV_Coupon_DW)

Relationships among the facts and dimensions were manually created with th use of Surrogate keys of dimensions and Foreign keys in the fact table.

2.2 Cube Structure

As the final step the cube structure **(Cube_Coupon)** is designed by using the data source view created in the previous step.

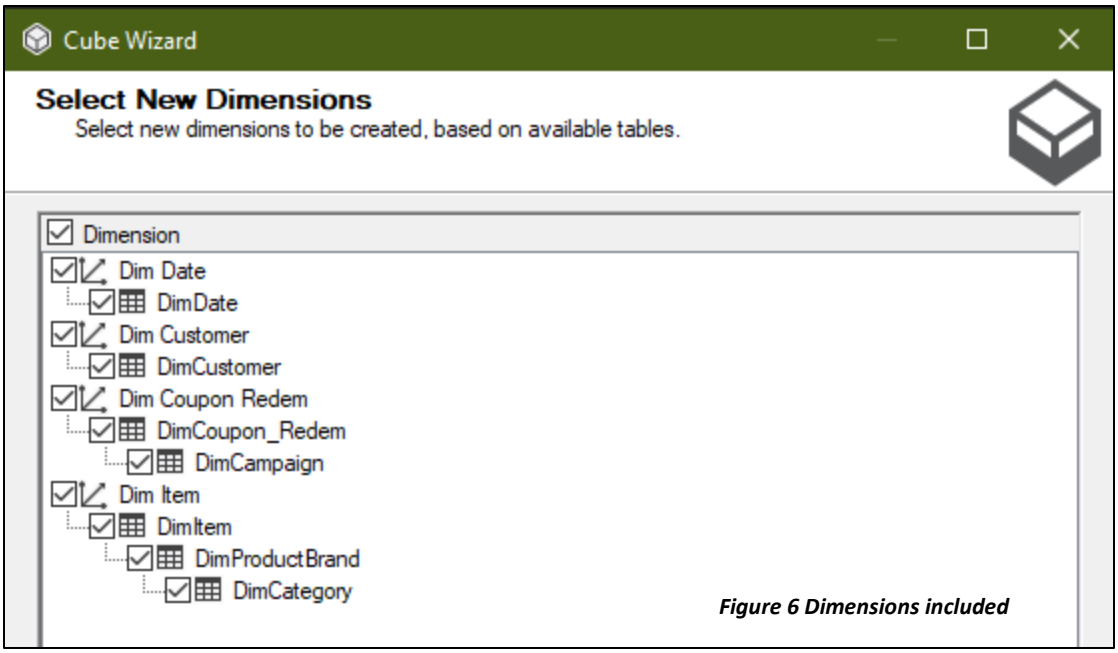


Figure 6 Dimensions included

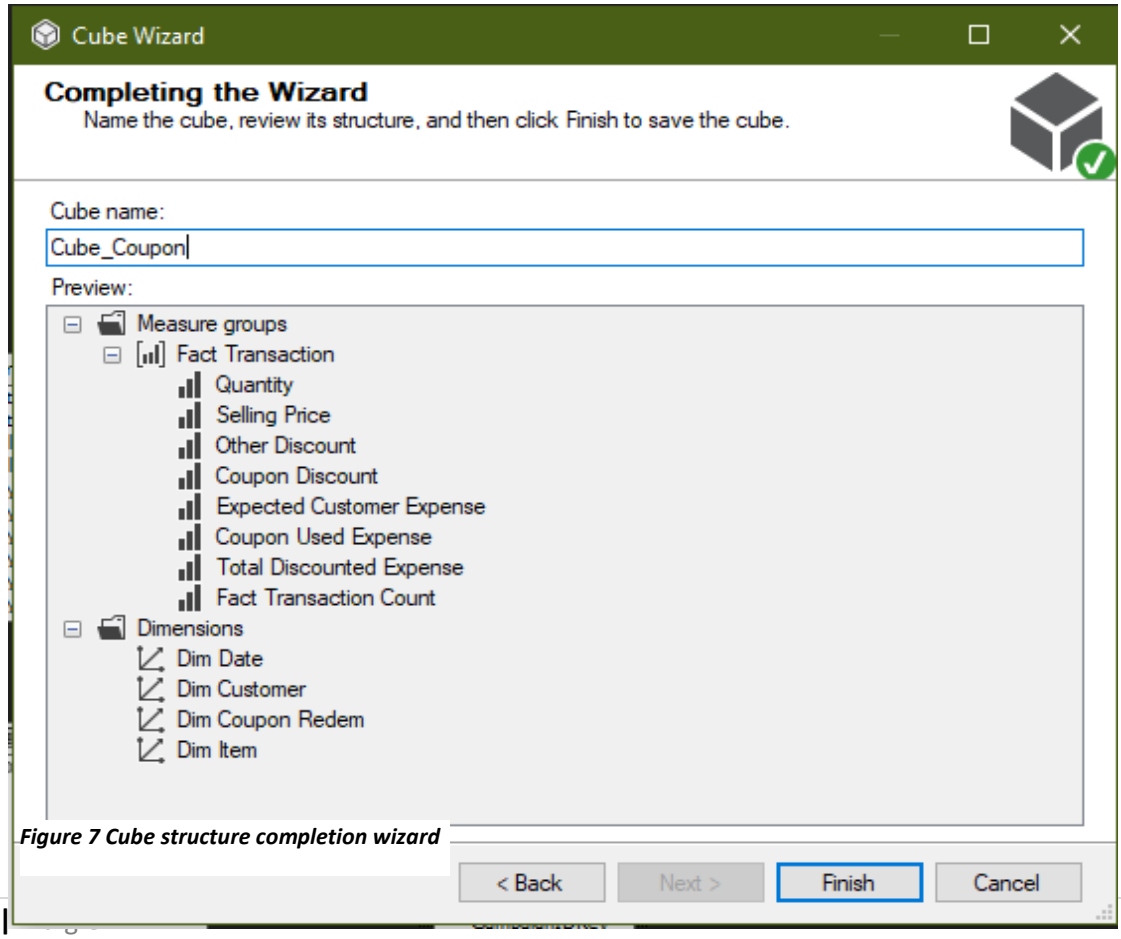


Figure 7 Cube structure completion wizard

The final cube structure created is as shown below. 7 main dimensions and a fact table are used in creating this cube structure

Dimensions and Fact tables used – DimCustomer , DimDate , DimCampaign , DimCoupon_Redemption , DimCategory , DimProductBrand , DimItem and Fact_Transaction.

The 4 dimensions DimCustomer , DimDate , DimCoupon_Redemption ,and DimItem are directly connected with the fact table whereas DimCategory , DimProductBrand and DimCampaign are hierarchichal implementations.

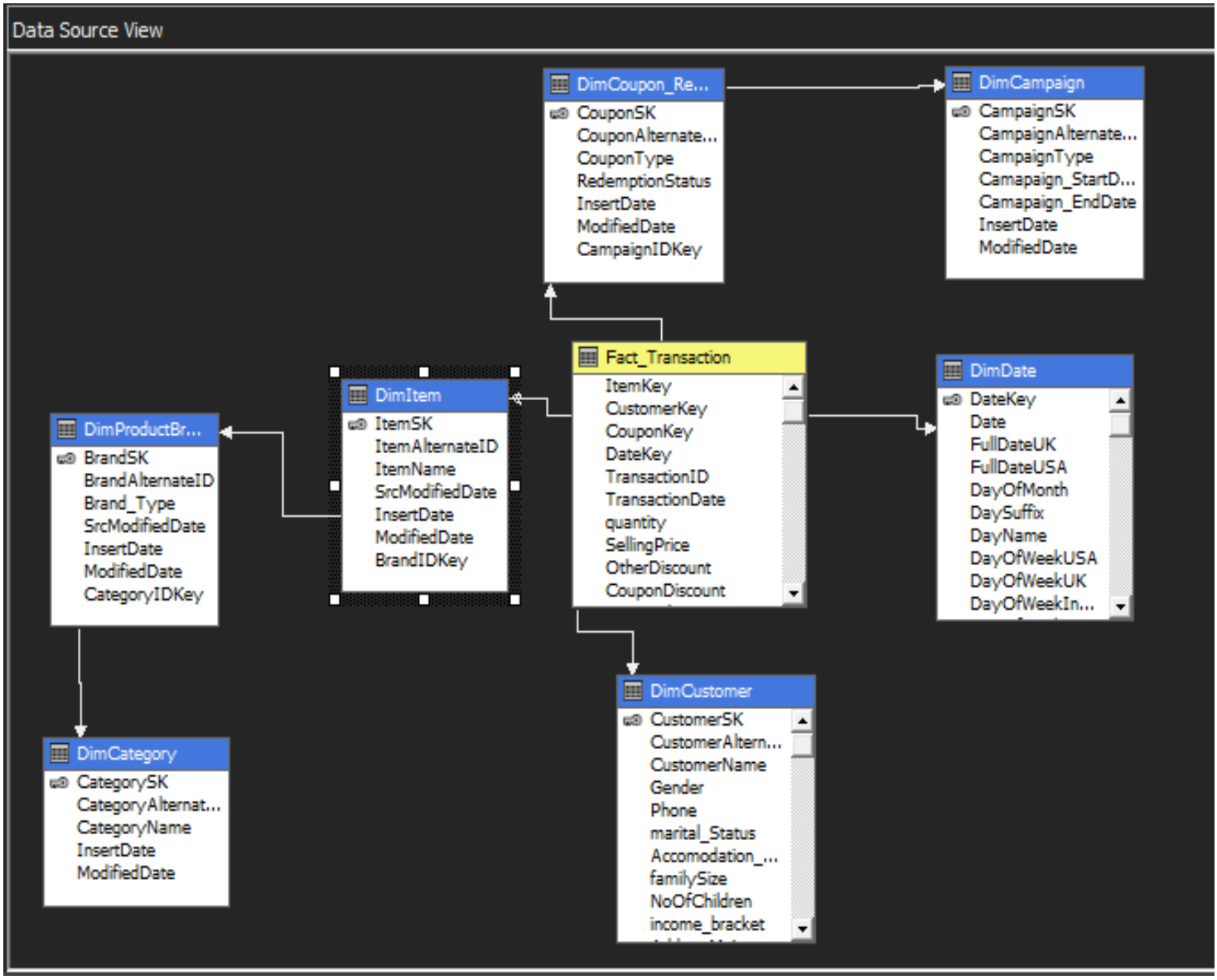


Figure 8 Cube (Snowflake Structure)

2.3 Hierarchies.

According to the data loaded in datawarehouse, the built snowflake schema reflects many hierarchical relationships between data in datawarehouse. Hierarchies are created for DimCustomer , DimItem and DimDate dimensions respectively. The hierarchies created are mainly based on location of the Customers , Category details of Items sold in the retail shop and Date of business processes.In addition hierarchical relationships between two dimensions DimCampaign and DimCouponRedem also exists.The main purpose of hierarchy creation is to provide the ability to drill down to detailed level data and to roll up to create aggregated data thus making business analysis purposes efficient.

Hierarchies are built in the **Dimension Structure** of the dimensions.

a. Date hierarchy.

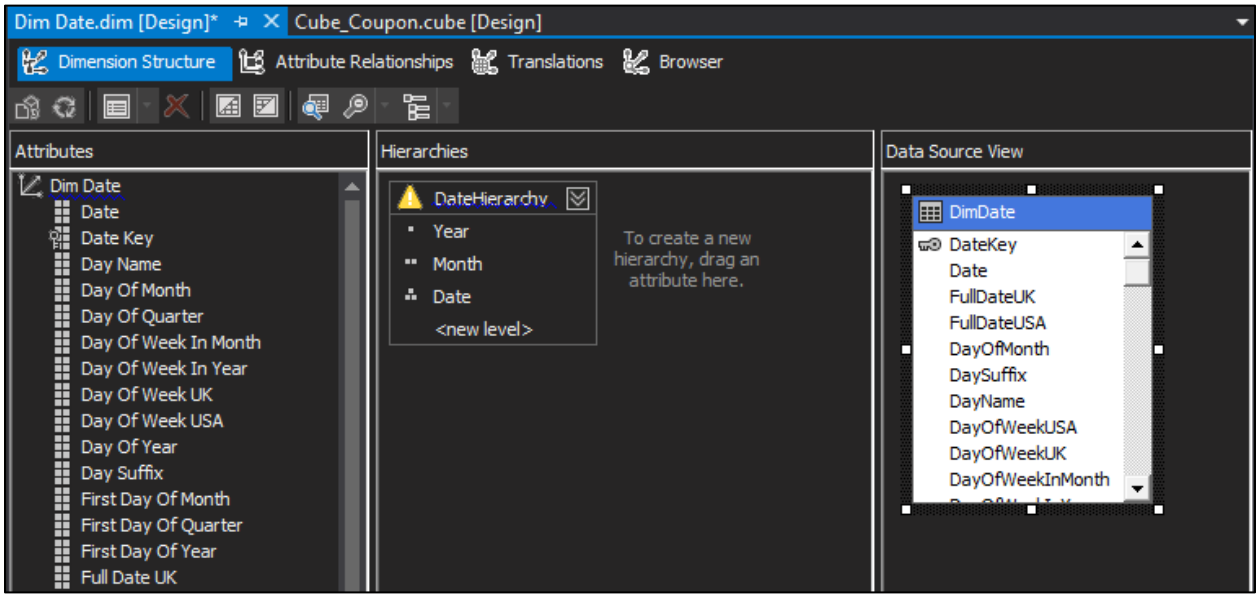
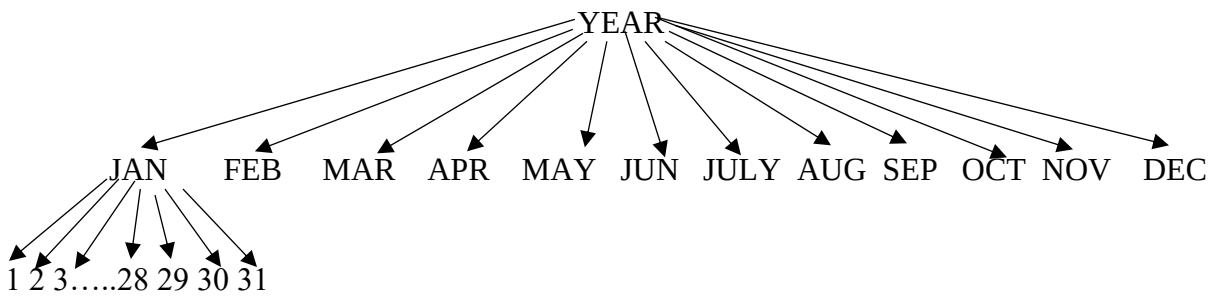


Figure 9 Date hierarchy

As per requirement Year , Month , Datewise hierarchy is created to analyse data in a more detailed manner with respect to date.



b. Location hierarchy.

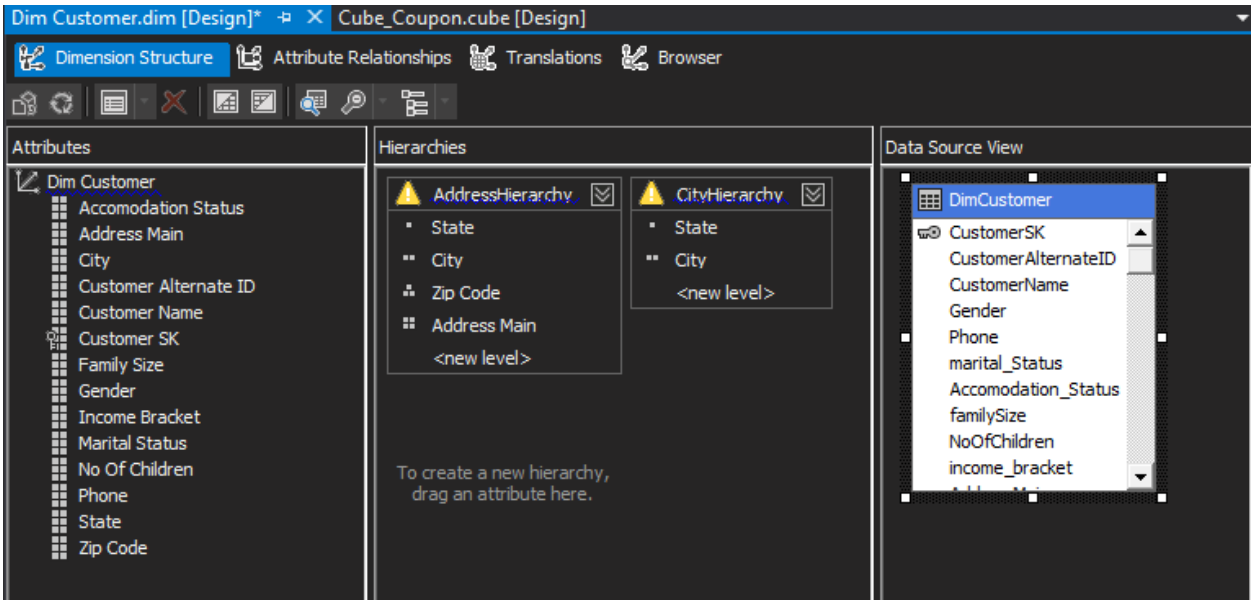
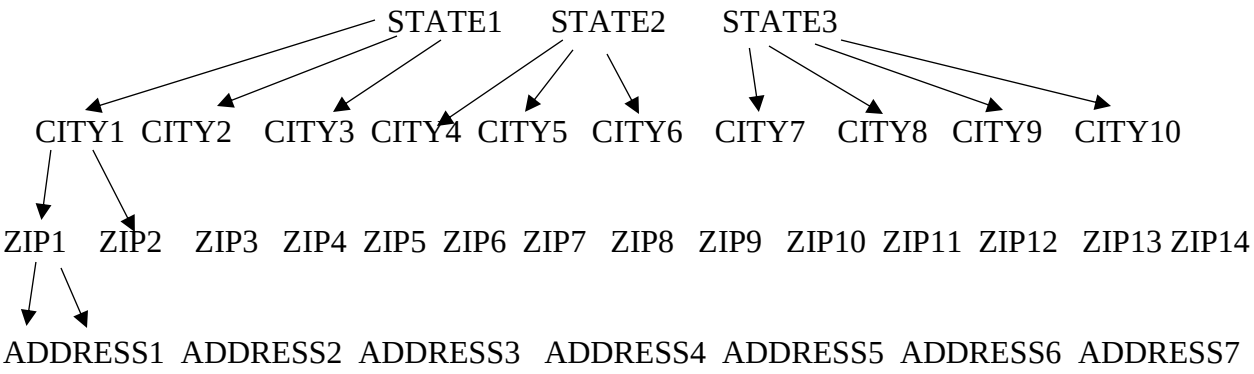


Figure 10 Location hierarchy

Customer location is analysed in many aspects as per business requirements. Therefore 2 location hierarchies are created in the same dimension providing two different levels of detail.



c. Product hierarchy.

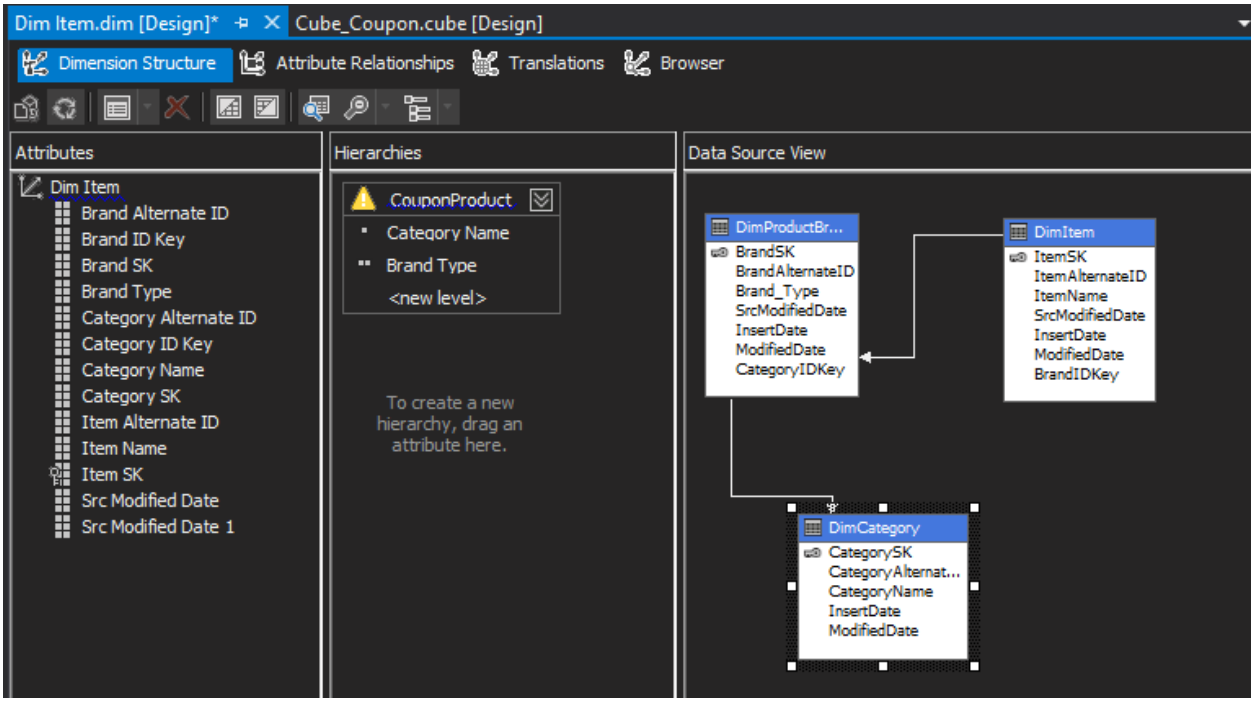
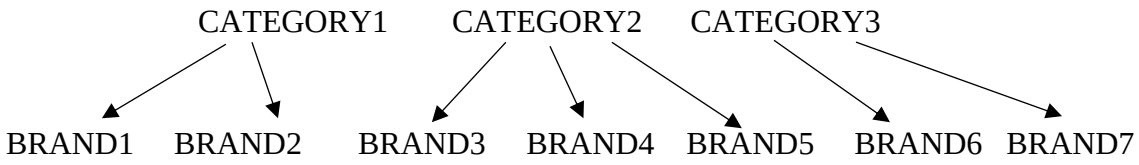


Figure 11 Product Hierarchy

Products in retail shop are analysed in terms of category and brands where each item belongs to.



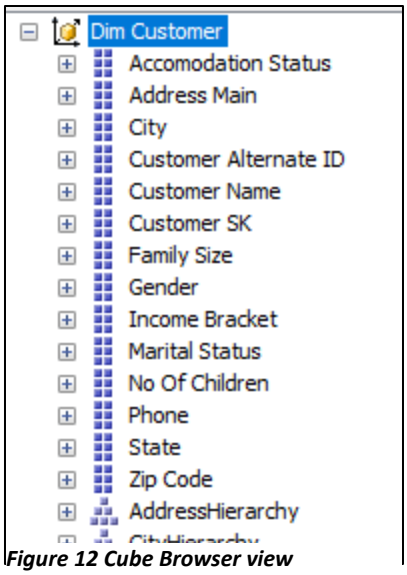


Figure 12 Cube Browser view

Once the hierarchies are created they can be viewed in the cube browser as shown in

2.4 KPI Creation.

KPI (Key Performance Indicator) shows the amount of progress that is done to achieve a business goal. The fact table Fact_Transaction is mainly used to create the KPIs here.

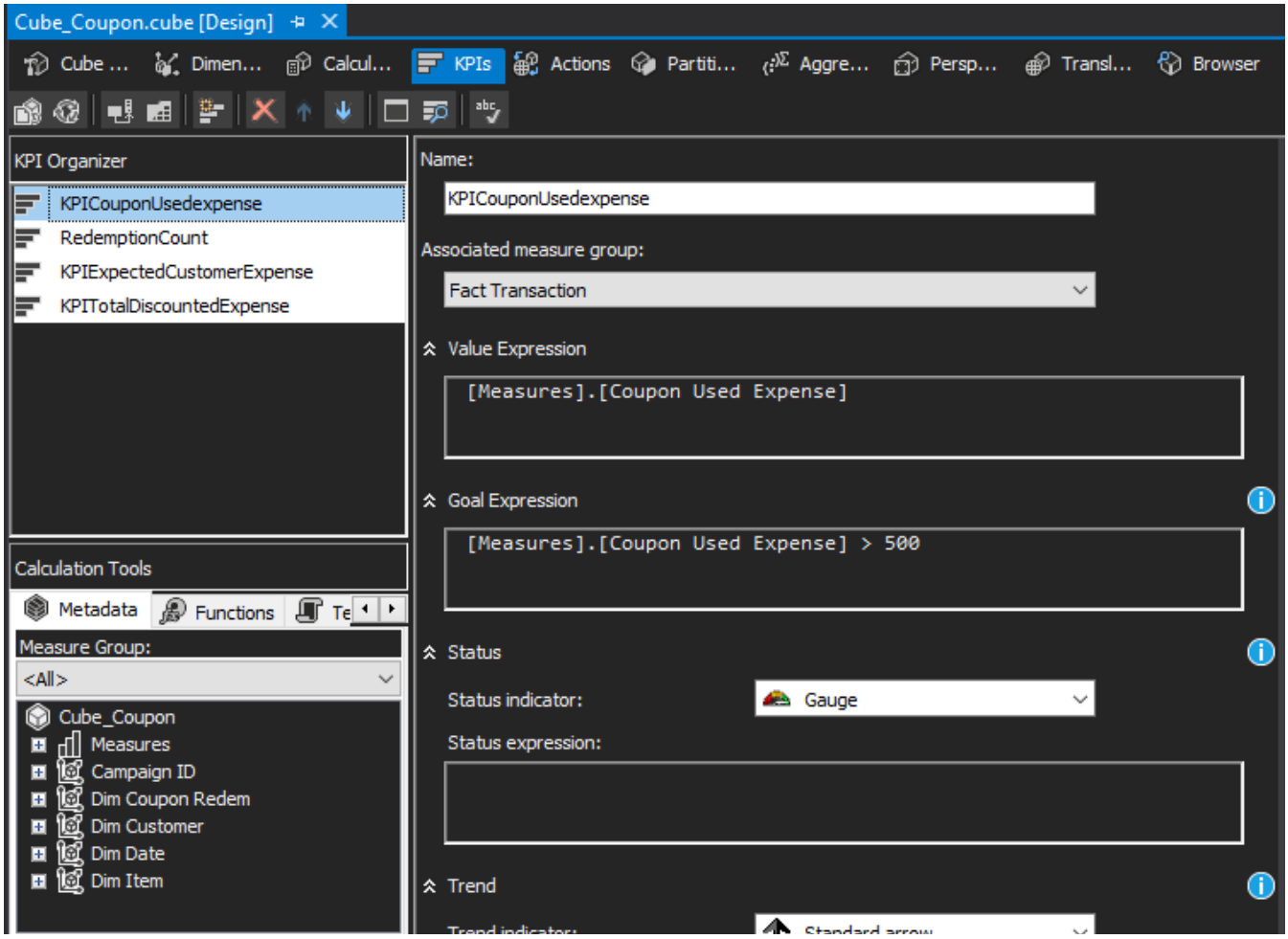


Figure 13 KPI creation wizard

KPI	GOAL	Description
KPICouponUsedExpense	`[Measures].[Coupon Used Expense] > 500`	If goal is false then there is no expected coupon usage.
RedemptionCount	`[Dim Coupon Redem].[Redemption Status] > 200`	If goal is true then the coupon marketing target successful.
KPITotalDiscountedExpense	`[Measures].[Total Discounted Expense] > 5000`	If goal is true business discount transactions successful,

2.5 Role.

Roles are created to provide control access to users in order to ensure who can do what. Role has a set of customised control access options to permit or deny permissions to objects within a particular database or cube. In this step Full control (Administrator) permission is provided to the role.

Figure 14 User role

2.6 Cube deployment.

Once the sub tasks are over finally cube is deployed in order to carry out the analysis purpose.

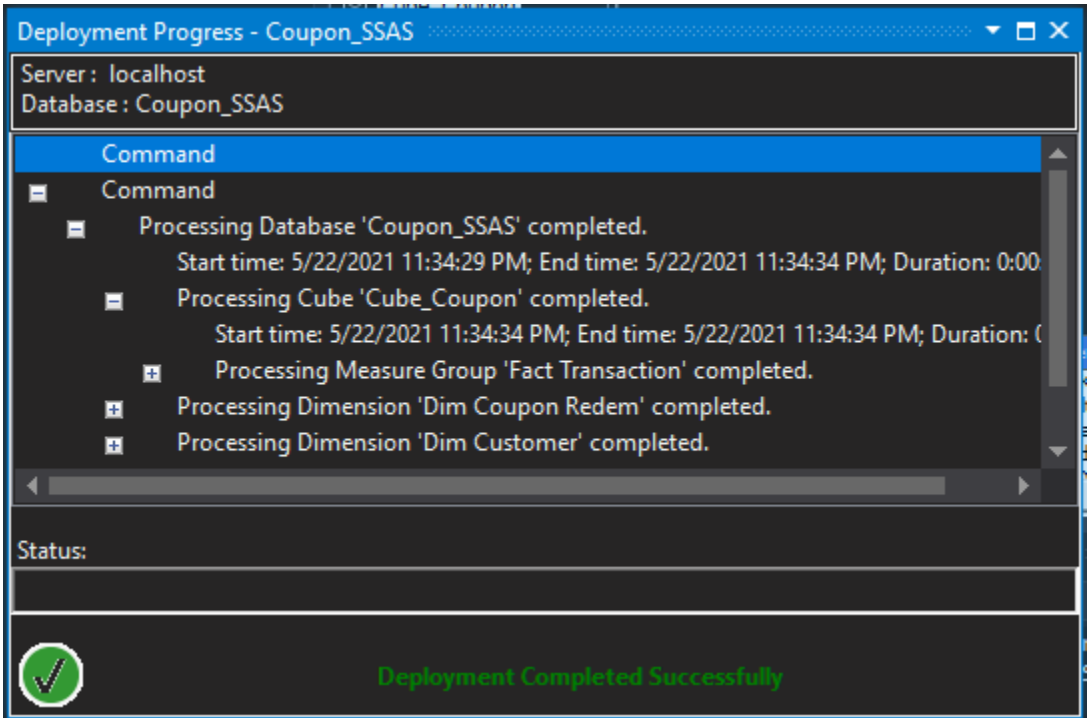


Figure 15 Cube deployment

At the end of completeion of cube strucure, Cube_Coupon is created successfully under the provided valid impersonation information credentials.

3.0 OLAP Operation Demonstration.

Once the cube is deployed succesfully,the deployed cube can be viewes in SQL Server Analysis service of SQL Server Management Studio.

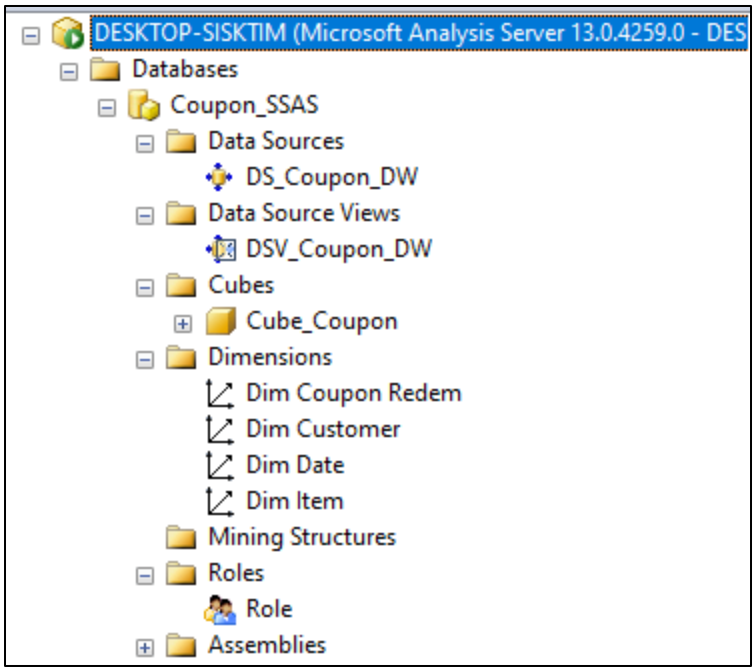


Figure 16 SSAS Cube Structure

This Cube Structure can be used to browse data in SQL Server Analysis Service where queries can be passed as MDX query for visualization processes.

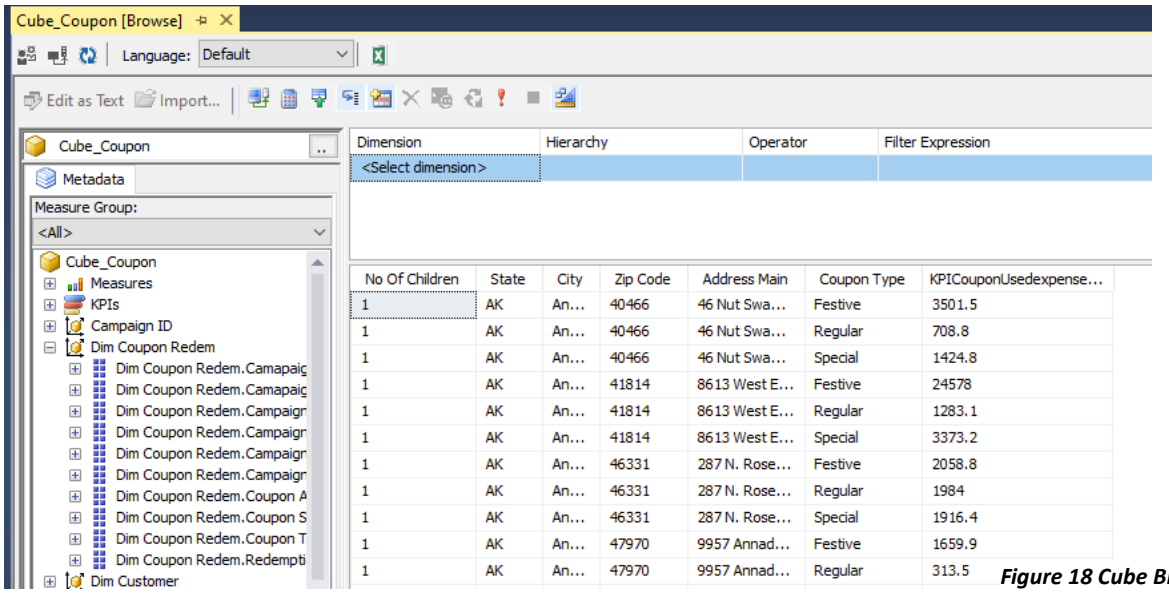


Figure 18 Cube Browsing

There are some main OLAP Operations demonstrated through Excel visualizations

- Slice Operations
- Dice Operations
- Pivot Operations
- Drill-Down Operations
- Roll-Up Operations

To demonstrate all of these operations , Excel workbook was connected to the cube via Data tab. Data from Cube is considered as the data source and is obtained via Analysis services.

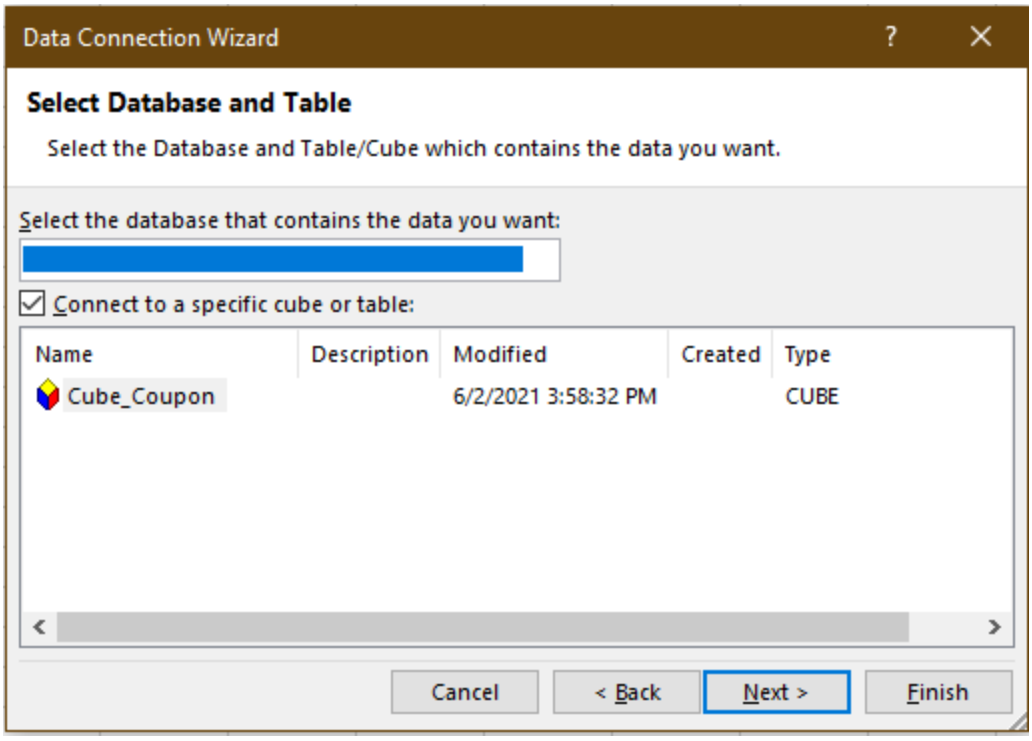


Figure 19 Cube Connection wizard in Excel

Pivot Table, Pivot charts and PowerView Dashboards are used to demonstrate the above mentioned OLAP operations.

NOTE – All operations are demonstrated in separate excel sheets of the Excel Workbook RetailStoreReports.xlsx.

For all the operations Pivot charts and tables can be generated using Pivot Chart fields which was obtained from the Cube accessed via Analysis Service of Data tab.

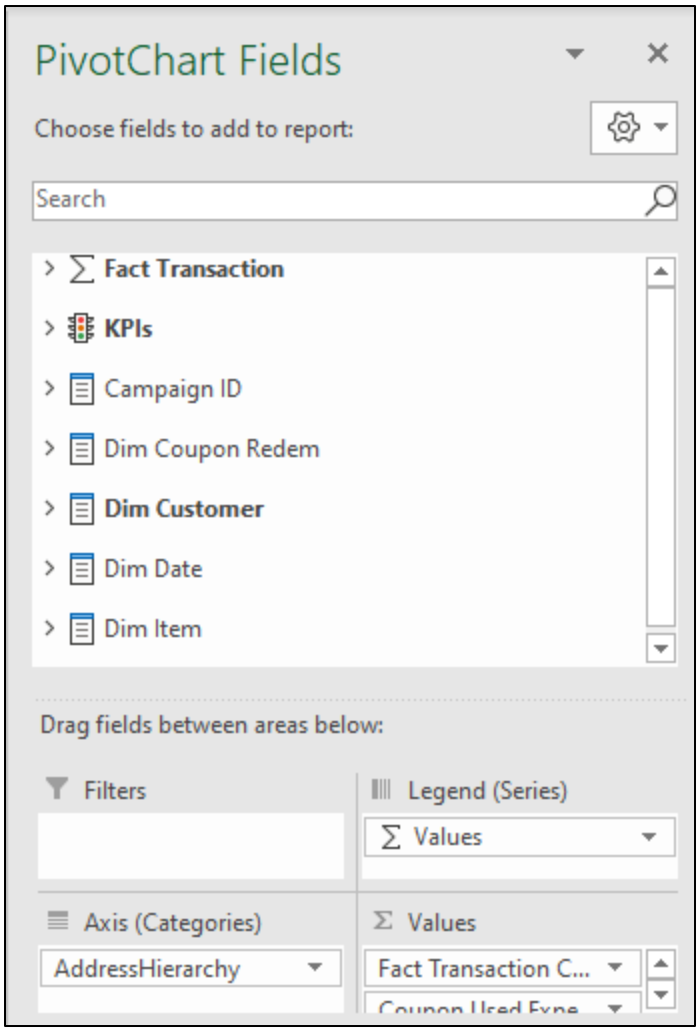


Figure 20 Pivot Chart Fields

3.1 SLICE OLAP Operation.

In Slice a single dimension is selected from OLAP cube which represents a Sub cube creation. Slice operation changes the overview section in accordance with the passed dimension. In simple terms the original view is sliced to get another different view.
In order to demonstrate the SLICE operation Pivot Chart and Pivot Table was used along with a Slicer.

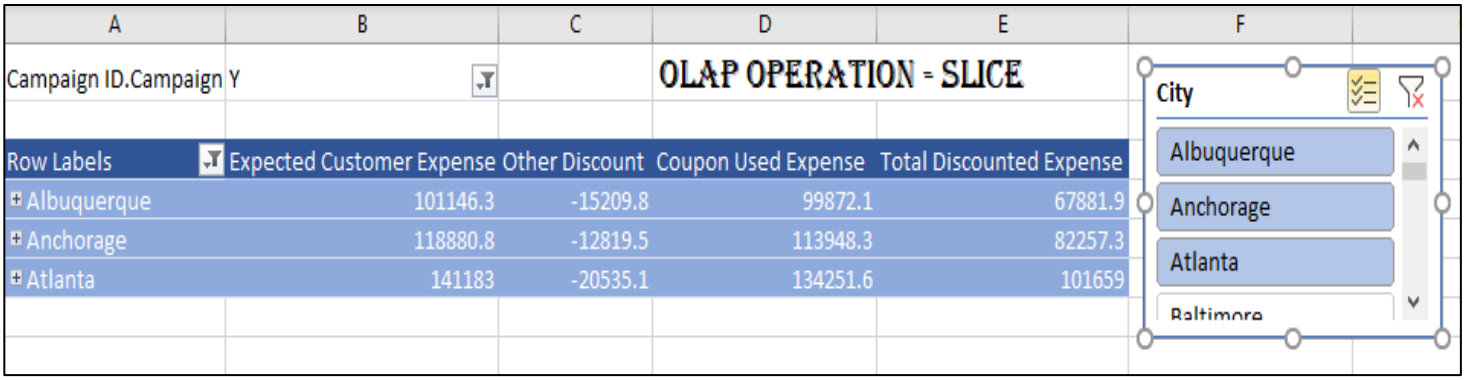


Figure 21Pivot Table - Slice Operation

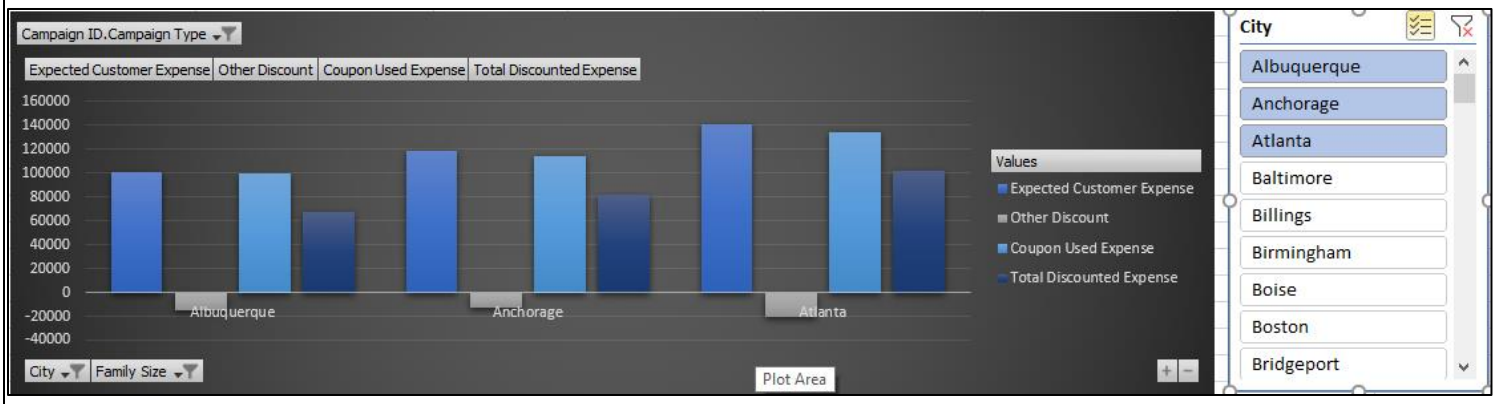


Figure 22Pivot Chart - SLice Operation