SRI LANKA INSTITUTE OF INFORMATION TECHNOLOGY

Malabe

Data Warehousing And Business Intelligence

IT 3021

Name : **Jayasekara M.D.S**

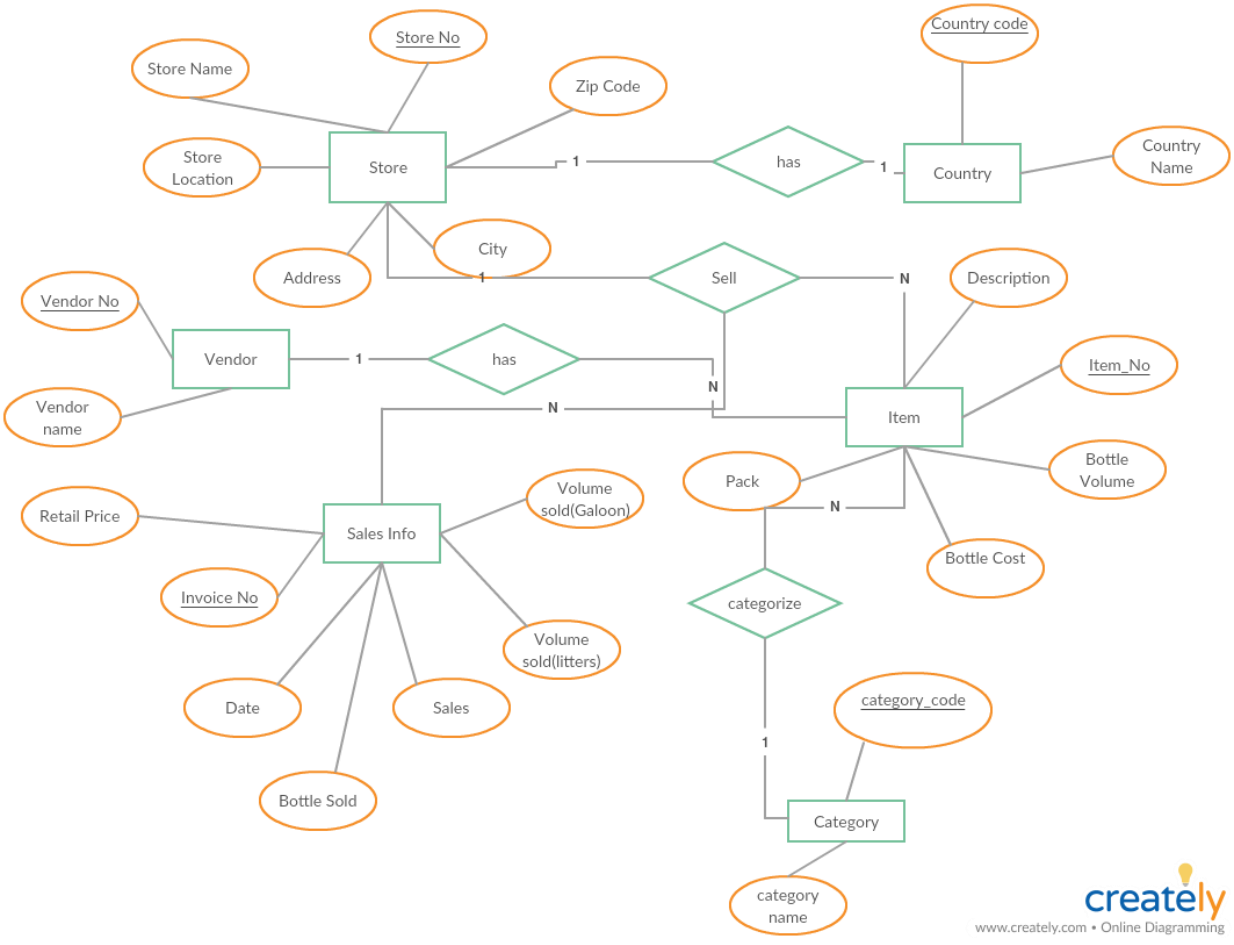
Reg No : **IT 131 12 660**

Batch : **DS-2018**

Data set selection

This dataset contains the spirits purchase information of Iowa Class “E” liquor licenses by product and date of purchase from January 1, 2012 to current. The dataset can be used to analyze total spirits sales in Iowa of individual products at the store level.

In this data set contain several sources, Excel, CSV, text files and data base files. Data set describe about “Sales details about Liquor in various stores(location) in various country which provided by several vendors over time since 2012 ”

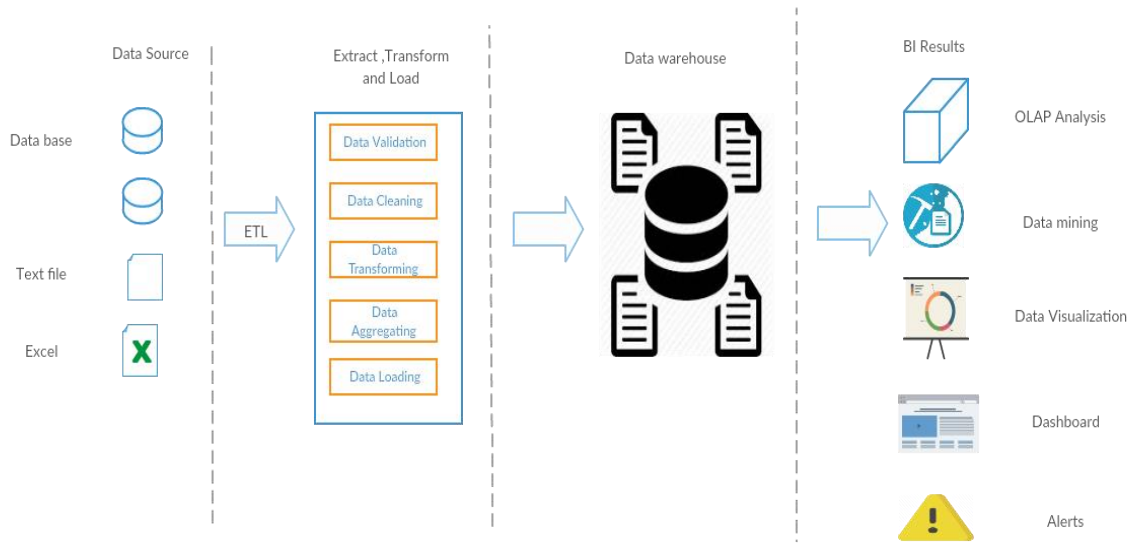


Preparation of Data Sources

There are multiple data sources which describe about liquor sales.

Source Name	Type	contains
Category	csv	Category Code and name
Item	Text	Liquor details
Store	Excel	Store details and Location
Vendor	csv	Vendor number and name
Date	Data base file	Date details
Country	Text	Country name and number
Liquor Sales	Excel	Sales Details

Solution Architecture



Data sources

Data source mean which sources contain data related to problem.In here ,There is multiple data sources. DB source, Excel, text files and CSVs .

Extract, Transform and Load

Extract

Data extraction about read data from data source systems. In this solution also read data from data sources and validate data to check accuracy of data.

Transform

Above extracted data store in multiple sources.Then going through some transformation can get all data to single place .In here we some transformation task to develop proper ETL.Sort ,Join, derive column and data conversions are some of them. Also In here we do data cleansing to improve quality of data.

Load

Load data to relevant destination. In our solution load data to Fact table.

Data warehouse

Using dimension model create data warehouse. There are two types of schema can use to create data warehouse. In our scenario I use snow flake schema to create dimension model. And here use some dimension modelling technique like slowly changing dimension, surrogate key generation hierarchy and fact tables.

Data warehouse design & development

Schema Diagram

Using Fact table and dimension table can create dimension model. There are 2 types of dimension model mostly use for data ware house design.

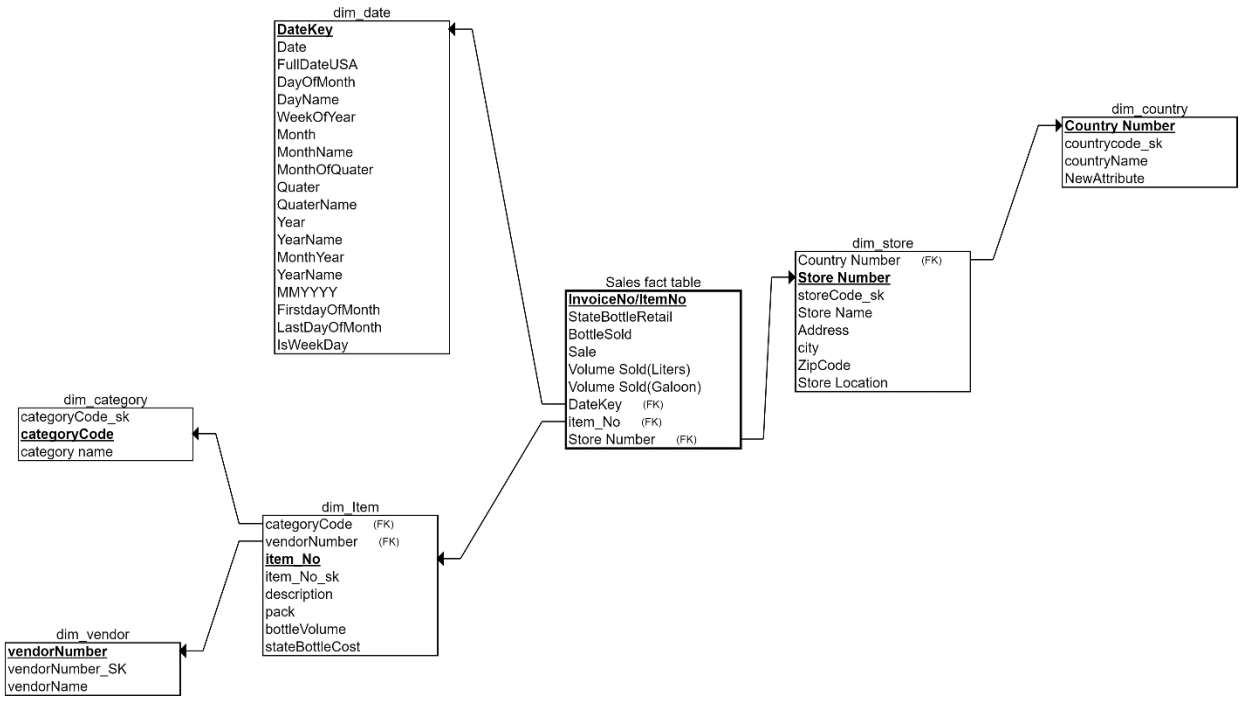
1. Star Schema
2. Snow Flake Schema

In our scenario Snow Flake Schema is most suitable one. Because In here some dimension have sub dimension and hierarchies.

e.g: country → Store Location

Assumption

- Vendor and Category are sub dimension of Item dimension
- There is hierarchy in store dimension, country → Store Location
- Date is another dimension

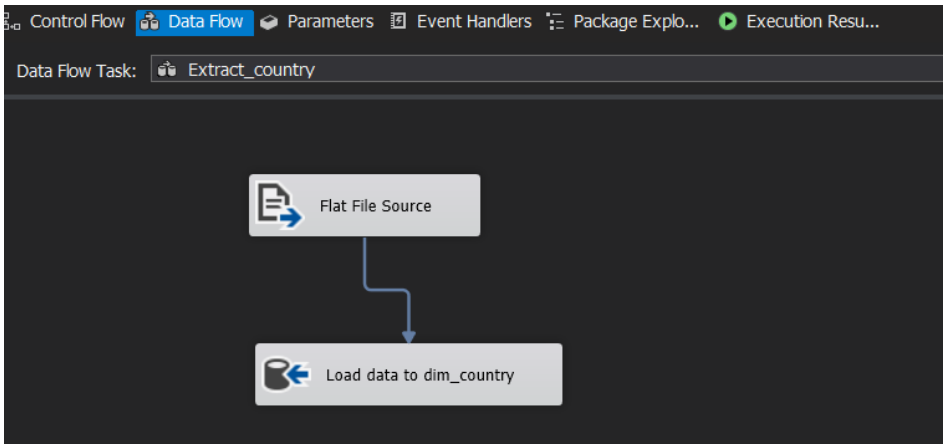


ETL development

Using MS SQL Data tool, SQL Server Integration Service can develop above mention Data Warehouse Mode. In here to get better result we use some transformation task

Step 01: Data Extraction

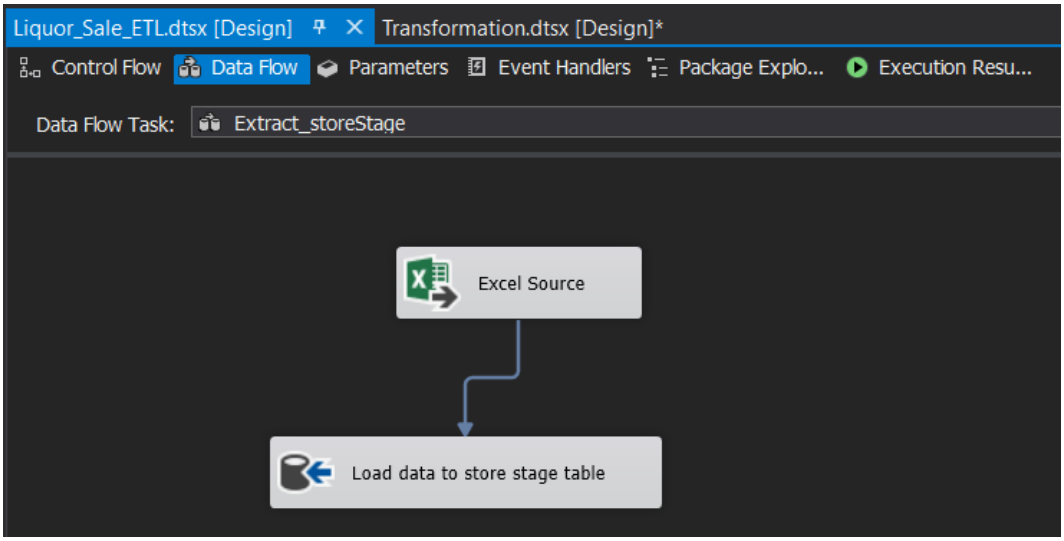
Extract data to country dimension



In here source of dim_country is text file. Therefore, we use flat file source as input and load data to dim_country

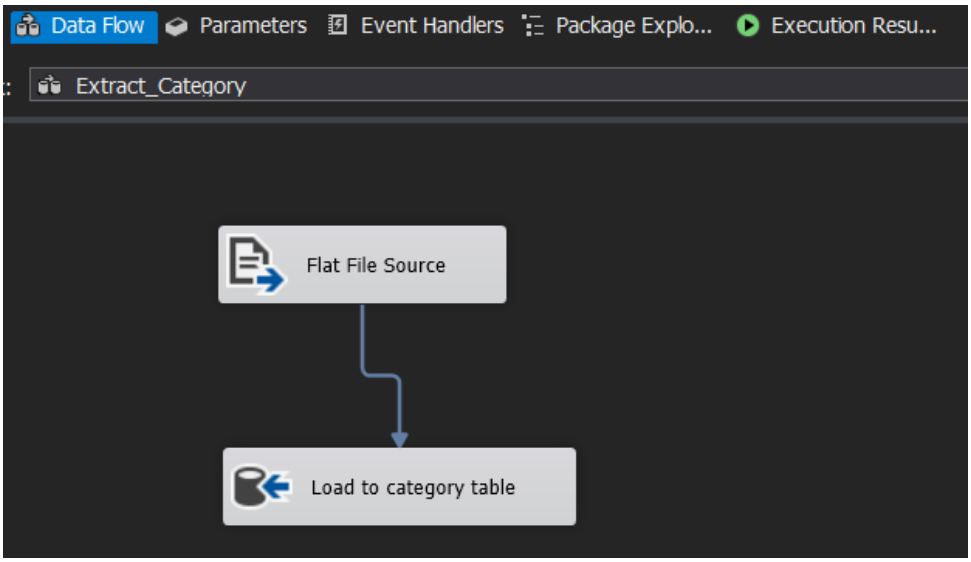
Extract data to store stage

According to dimension model, dim_store has hierarchy. So it should join with dim_country.Then before doing joining transformation should extract store data from source. For that we use store stage table.



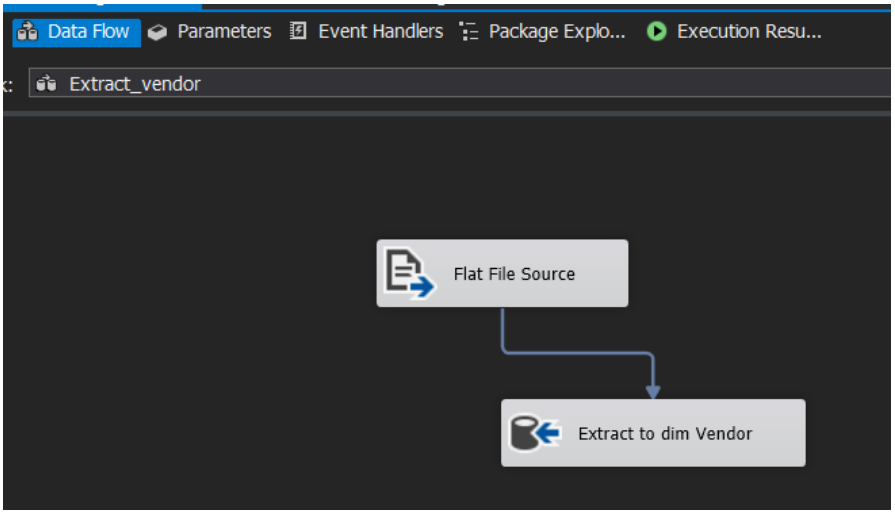
Extract data to dim_category

There is no extended attribute in this dimension.Then we can directly extract to dim_category.



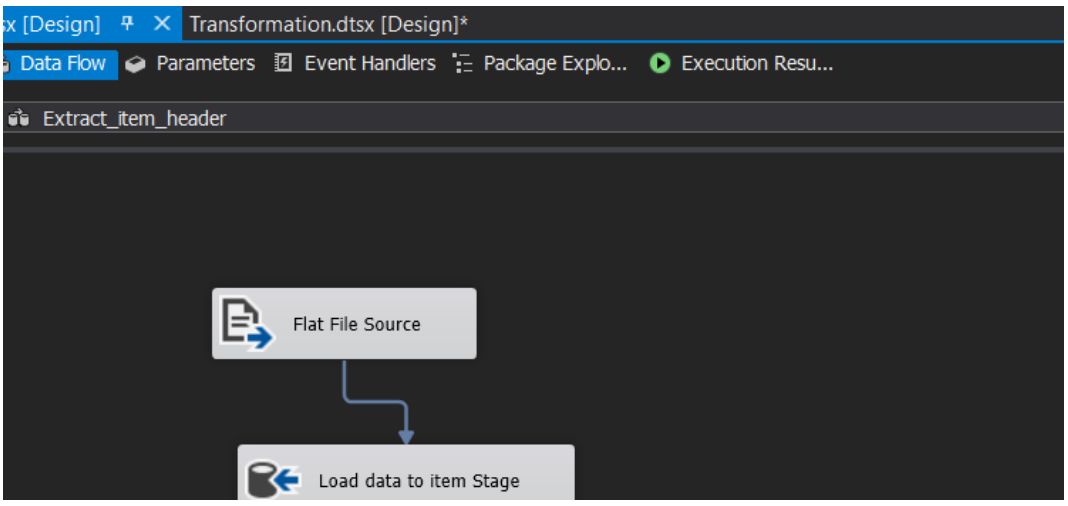
Extract data to dim_vendor

Also here no extended attribute.Then we can directly extract to dim_vendor.

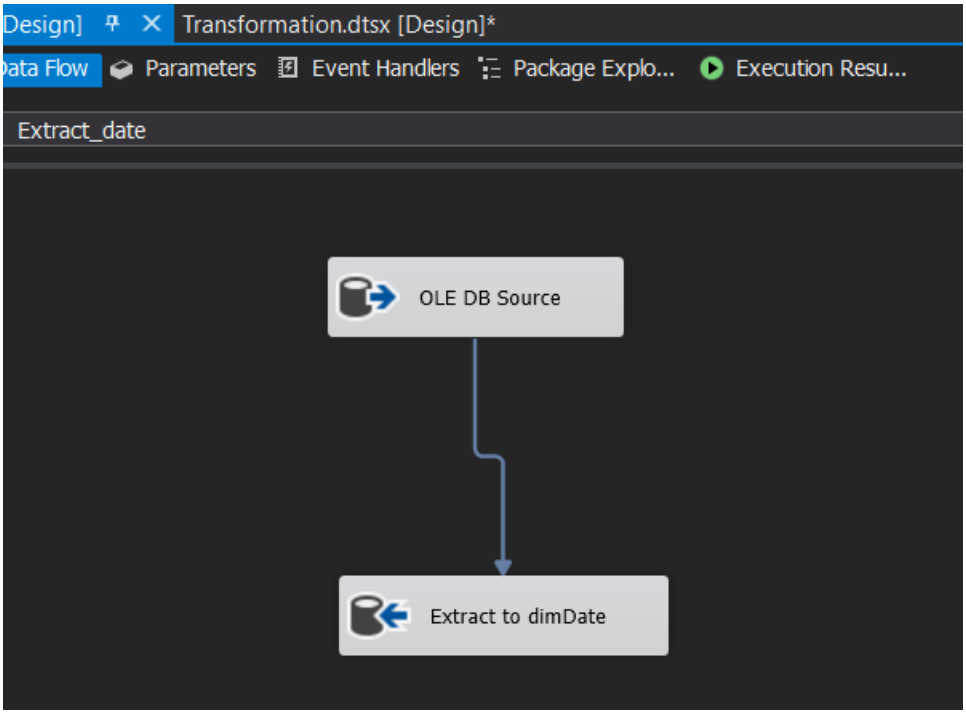


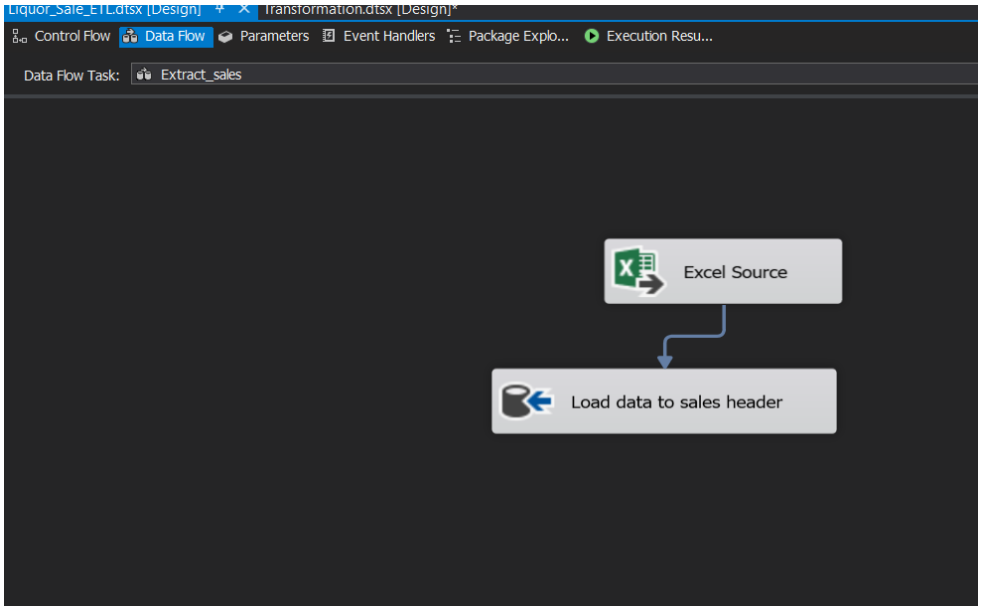
Extract data to Item

This also stage table. According to dimension schema dim_item has two foreign keys. That mean dim_item refer two sub dimensions. So we do not extract data to dim_item directly. Then use stage table.

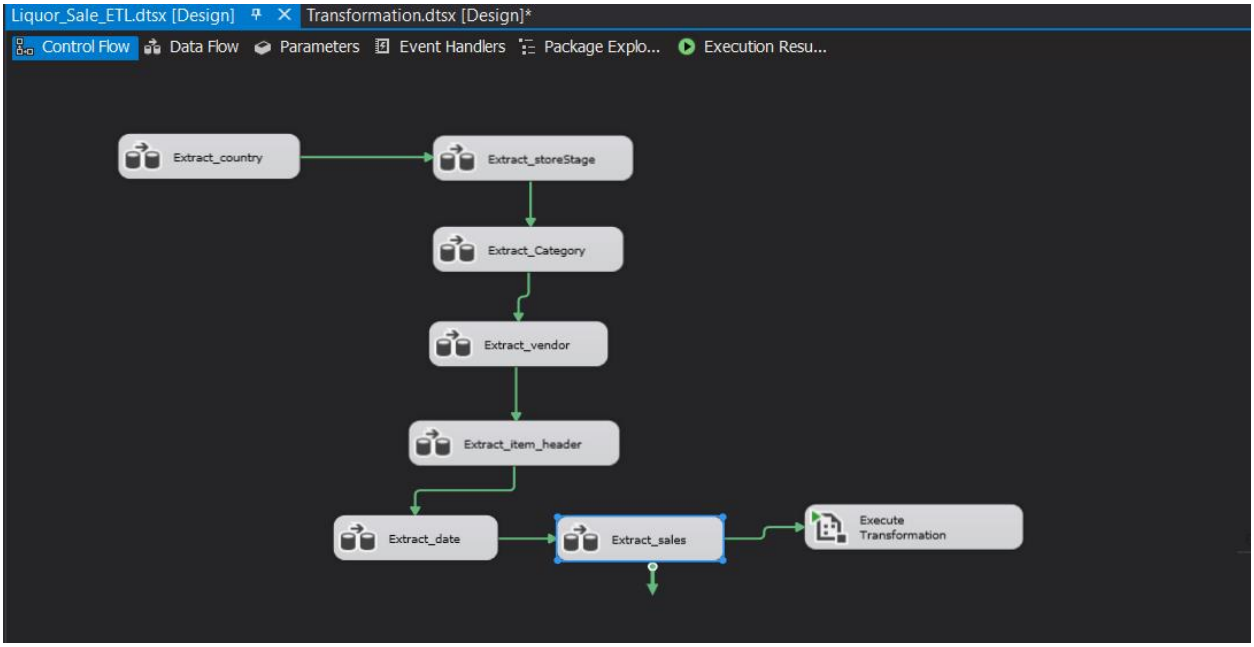


Also extract data to dim_data and sales header,





Finally, Data extraction had been like this,



Next step is Data Transformation,

Step 2: Data Transformation

In this step we should transform data to subjectable manner,

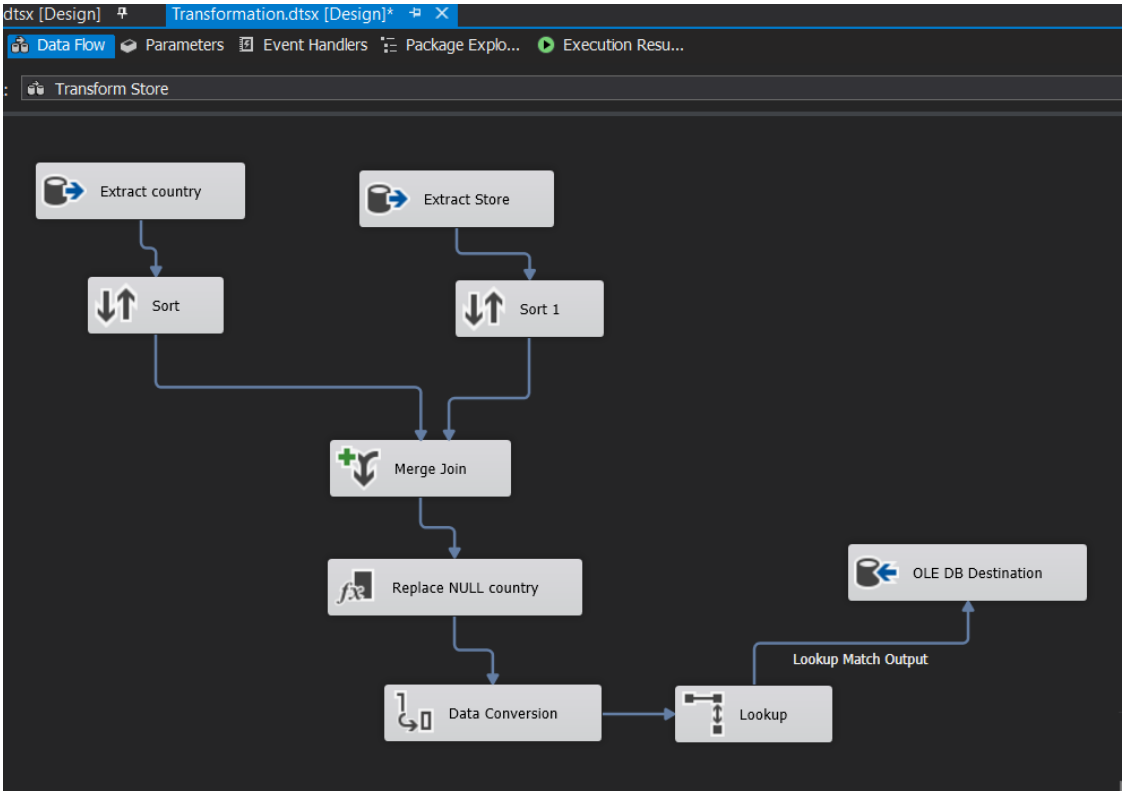
In store dimension, there is hierarchy. Mean the data would populate to dim_store in multiple places. Doing transformations, can get all the data to single location.

In here we use Sort , Merge Join and LookUp.

Sort : Sorting data with common key in both table

Merge Join : Using above common key join that two sorted table.

Look Up : It help to manage matching data in tables.



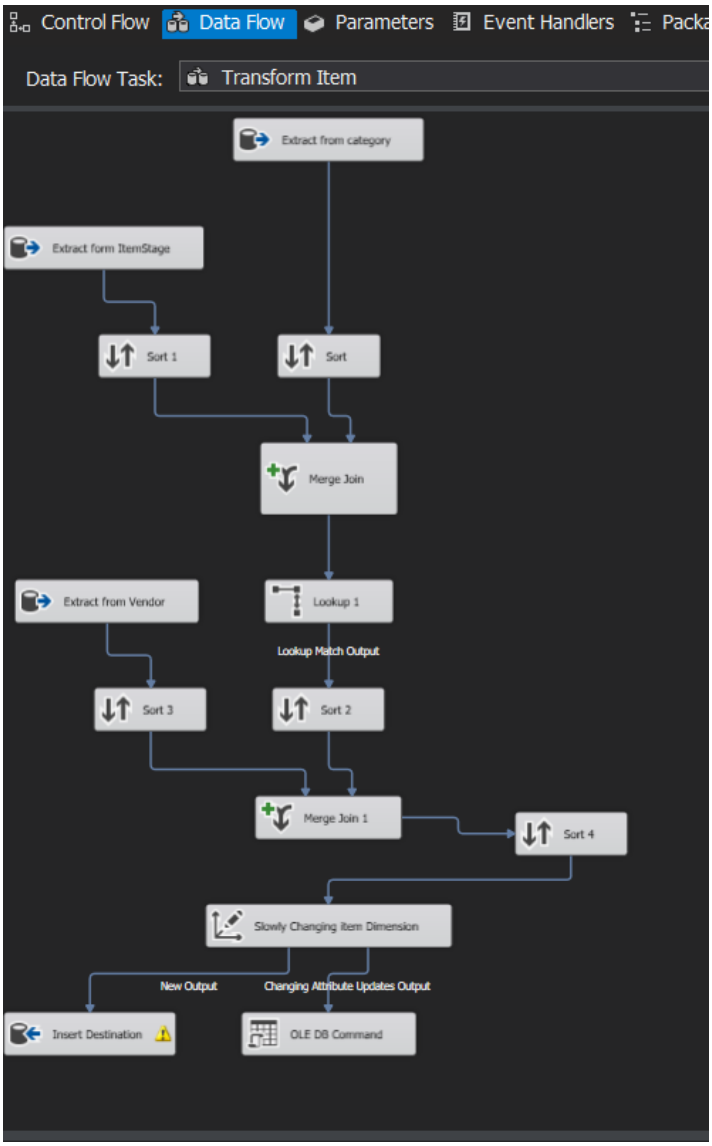
Data set can be have NULL values and missing values.In this transformation step we can avoid that using Derived Column.

Also according to requirement change data types using Data Conversion.

Transformation of dim_item also be like this,

In here rather than dim_store, We use slowly changing dimension,Because there is changle attribute,

State Bottle cost can be change over the time.Then data warehouse should update according to it.



Step 3: Load data to Fact Sales

Before load data to fact sales,

Join above transform data to sales header table. Use the Merge Join, Sort and look up transformation task to fullfil Join. Also In fact table we mention all measures.

In this step derive all computation to relevant column. And convert data type to according to fact table.

Finally using OLE DB Destination, Load data to fact table