Data Warehousing and Business Intelligence

Retail Shop System

Submitted by:
IT18128796 – K.P Aravinda

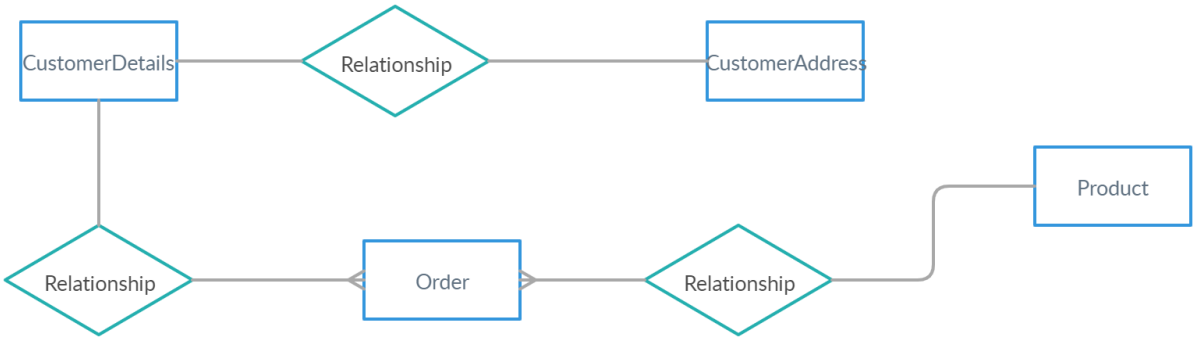
Submitted to:

……..

04/29/2020
Date of submission

Data set selection

My data set is on about retail shop detail data source. I select this data set from the link that provided in the course web. This data set consist on products customer and orders that they received. This data source basically based on the business function of the retail organization. So this consist of product functions, customer functions and order functions. Main reasons for selection of this data set because through this I able to create us slowly changing dimension and SSAS Cubes for future procedure. After selecting this data set I send it to Mr. Aloysius for the confirmation. ER diagram of the data source is below.



Preparation of Data Sources

In my data type is Excel data set type. So I convert two files into text. So I have two type data set. Those are excel and text. Order and area are in excel format and customer and product are in text format. After uploading data into staging data is under below category

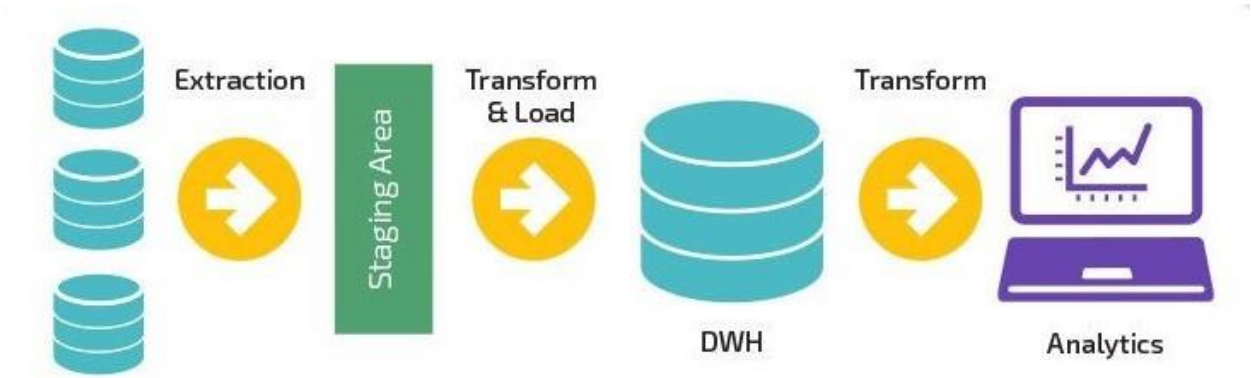
Orders
-order_id(int)
-customer_id(int)
-order_date(datetime)
-product_id(int)
-unit_price(float)
-qty(int)
-amount(float)
-order_line_number(int)

Customer Details
- customer_id(int)
- first_name(varchar)
- last_name(varchar)

Customer Address
-customer_id(int)
-region(varchar)
-city(varchar)

Products
- product_id(int)
- product_name(varchar)
- family(varchar)

Solution Architecture



This system solution architecture I flow like above, first extract data from the source and stage it in the staging data base then it transform which is according to data warehouse requirement then load those data into data ware house.

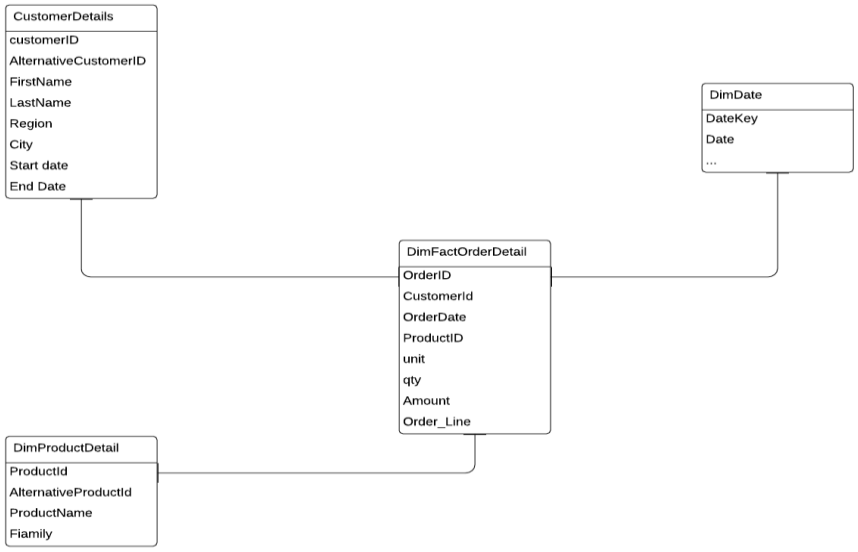
Here I hope to use the top down architected. This architect was introduced by inmon and use data ware house like centralized repository after creation of the data ware house data marts are going to be created. This the approach most of the organization are used. But this architecture is highly cost and time consuming.

For my approach I use excel for the data cleaning part then for ETL process I have use ssis and data management I have use ssms. For future report generation I have use ssrs.

Data warehouse design & development

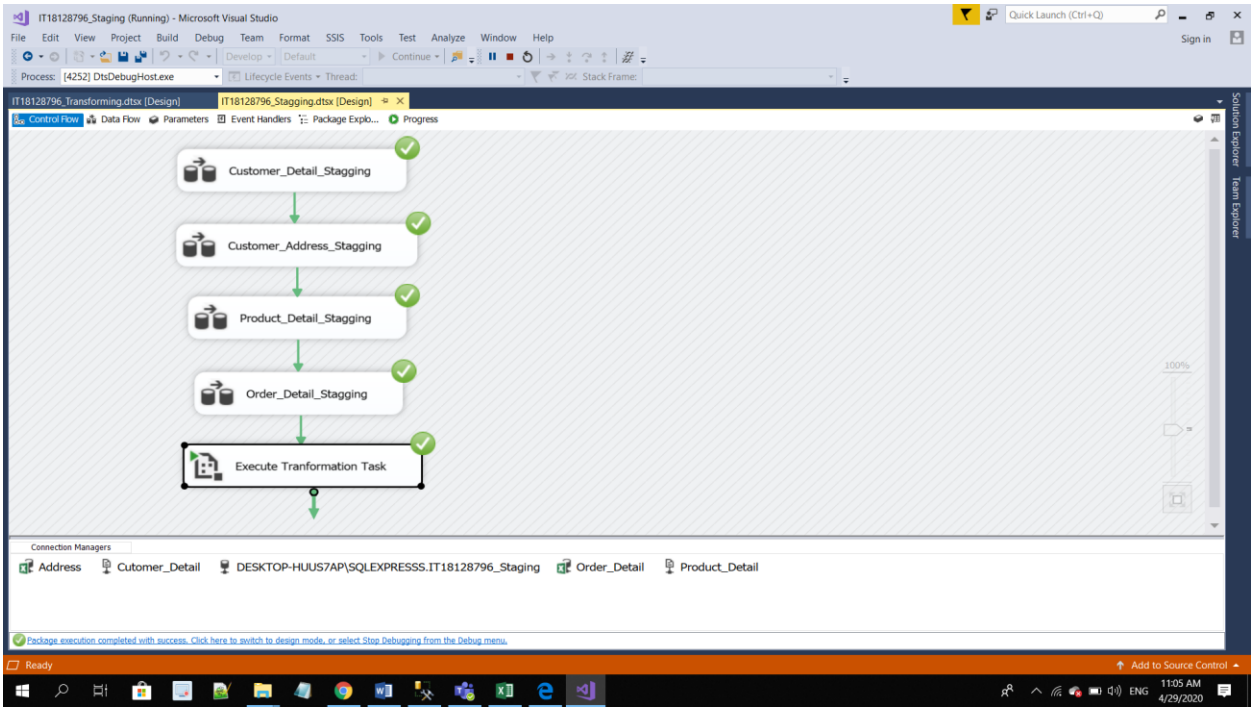
My data set has star schema. A star schema has a single fact table for each dimension and dimension table not have other dim table keys. Typically a de-normalized table solution. This data set consist 3 dim table and one fact table. Here dim customer table has slowly changing dimension. Those are region and city columns.

Star Schema

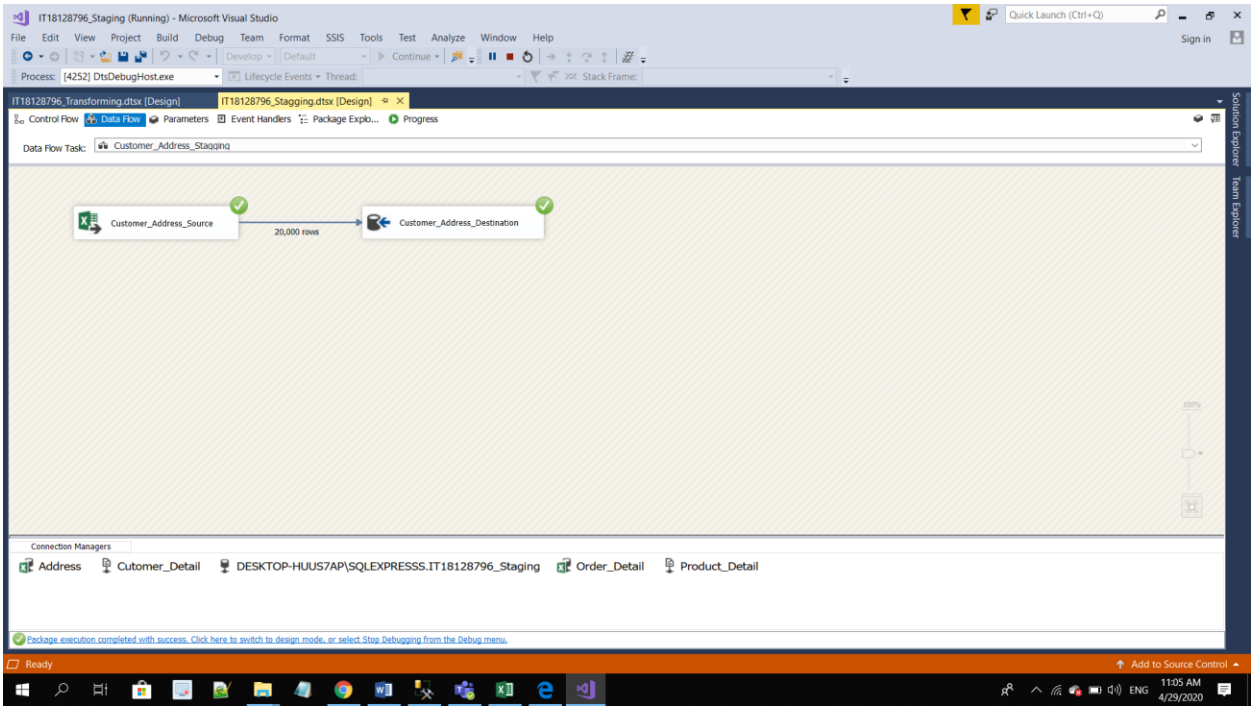


ETL development

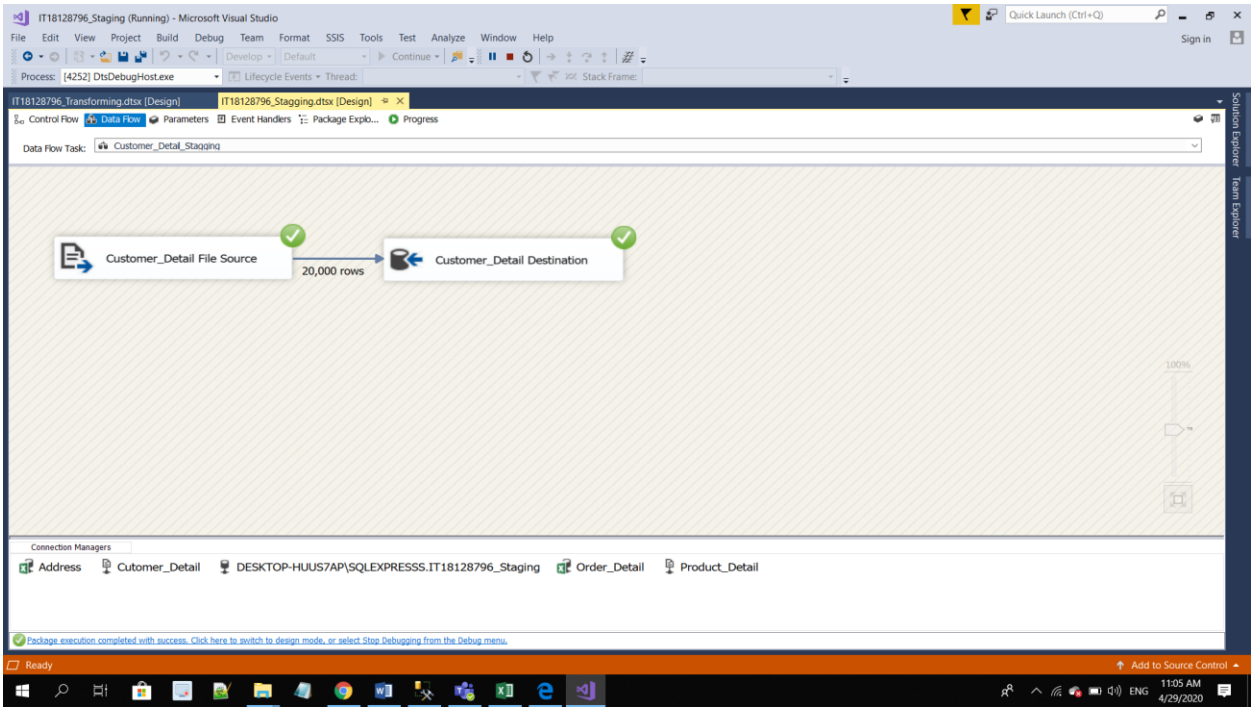
As explain previously staging part is done as below. Customer_Details, Customer_Address,Product_Details and OrderDetail source files are stored in IT18128796_Staging db.



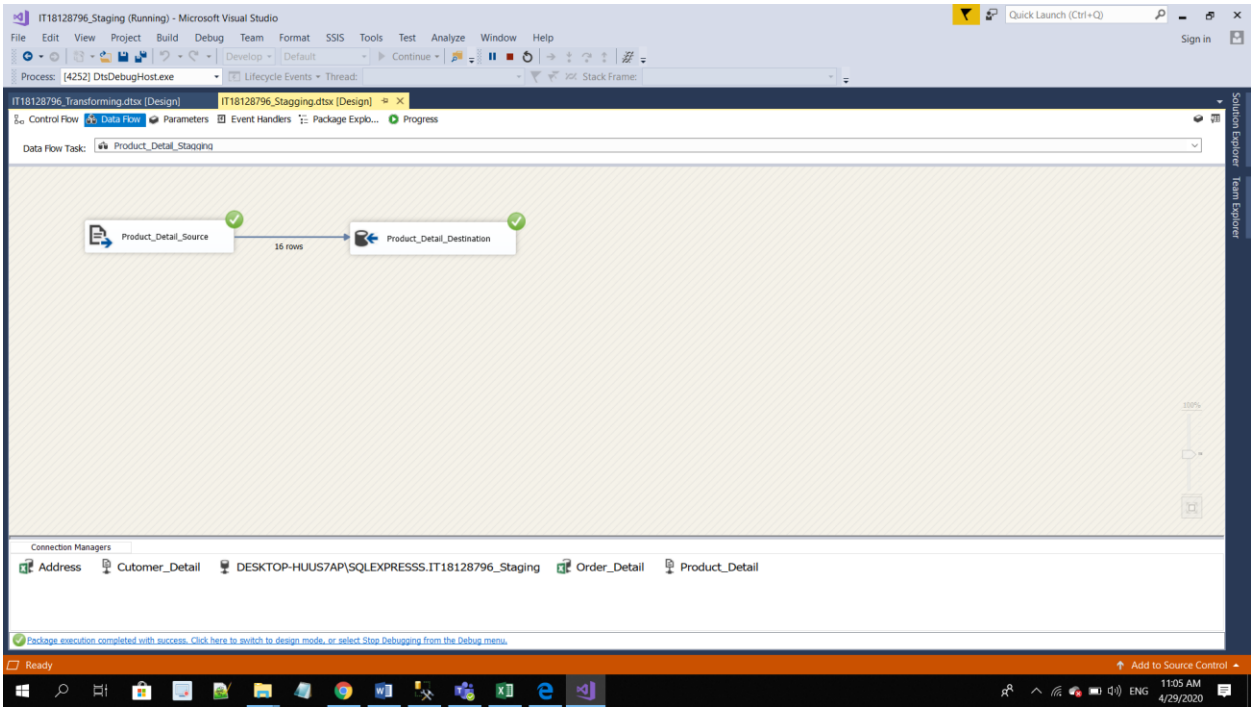
Customer_Address are staging.



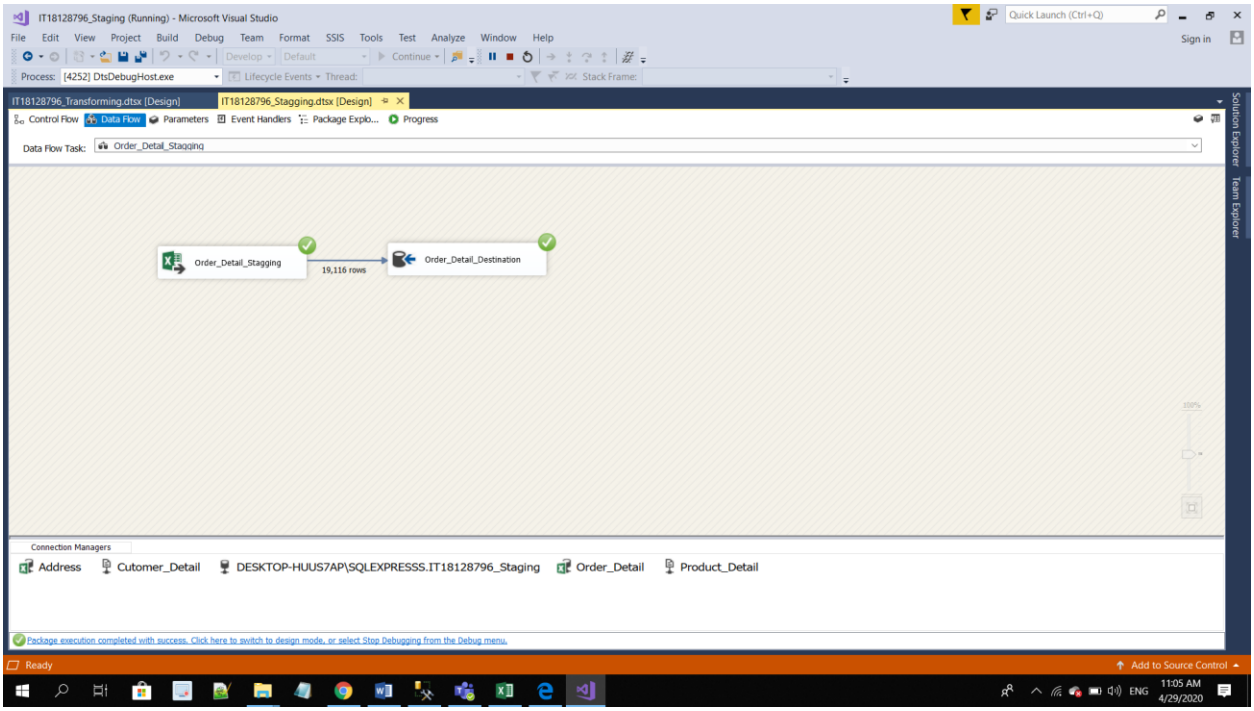
Customer details are staging..



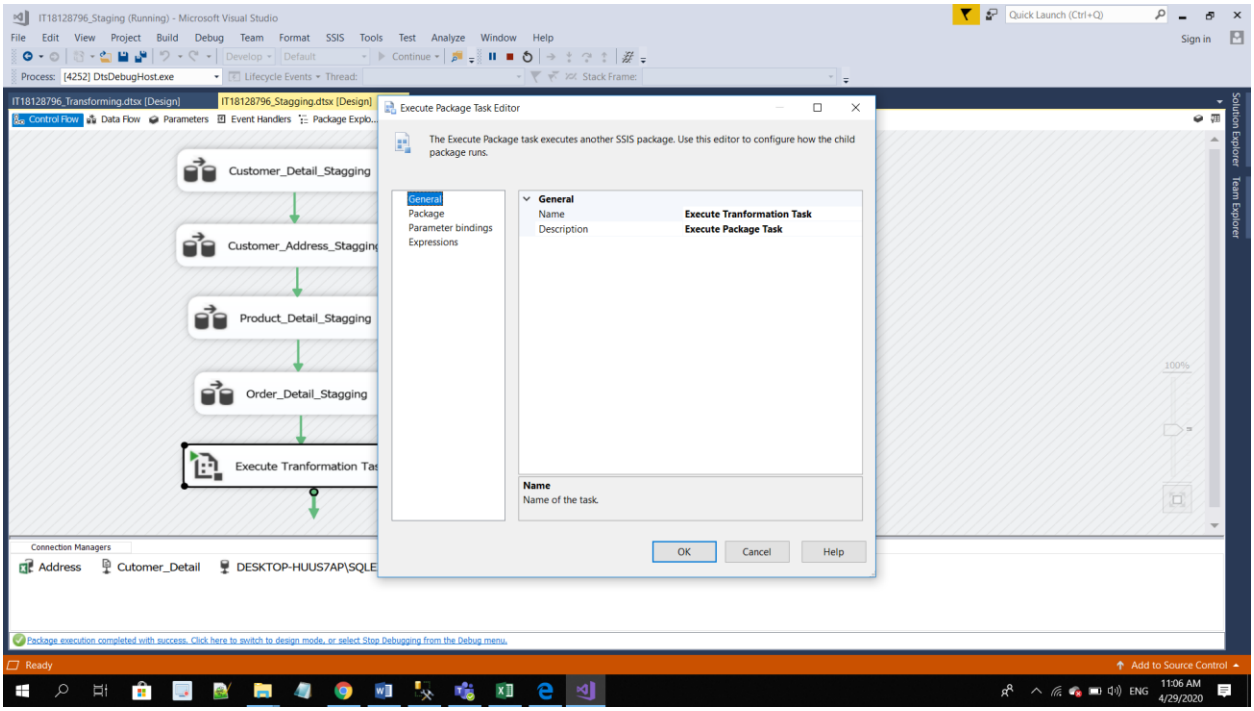
product details are staging



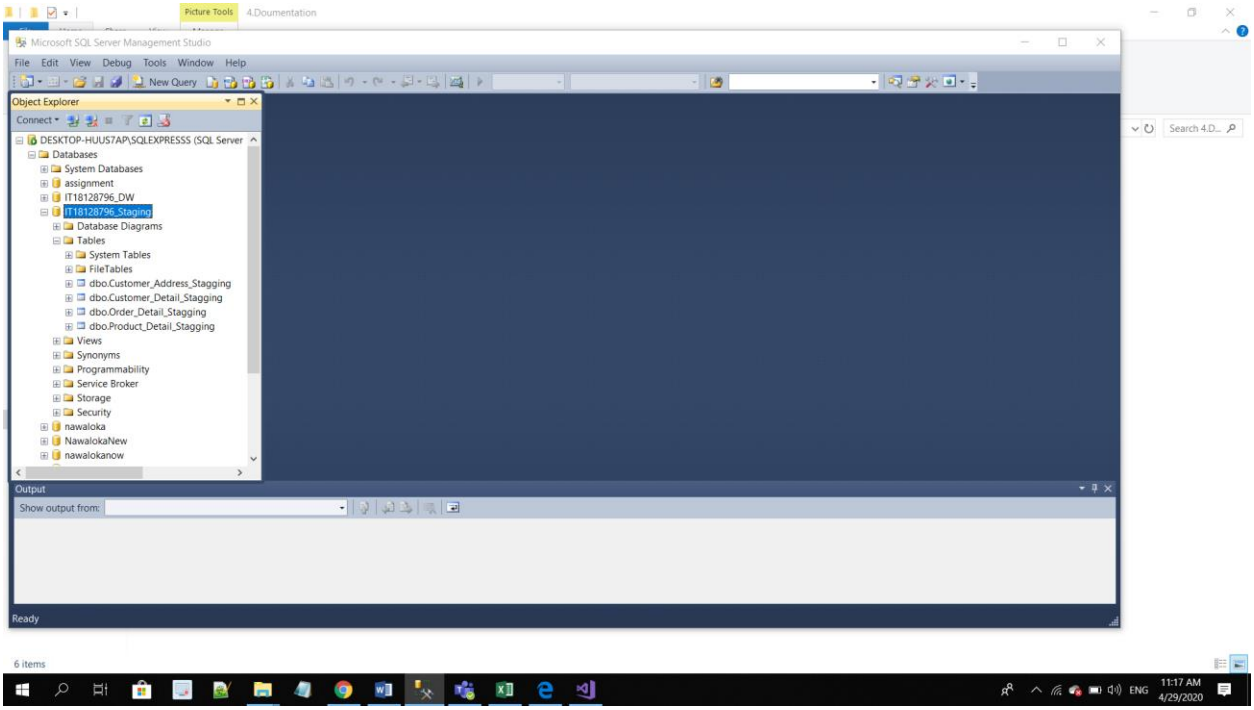
Order Details are staging



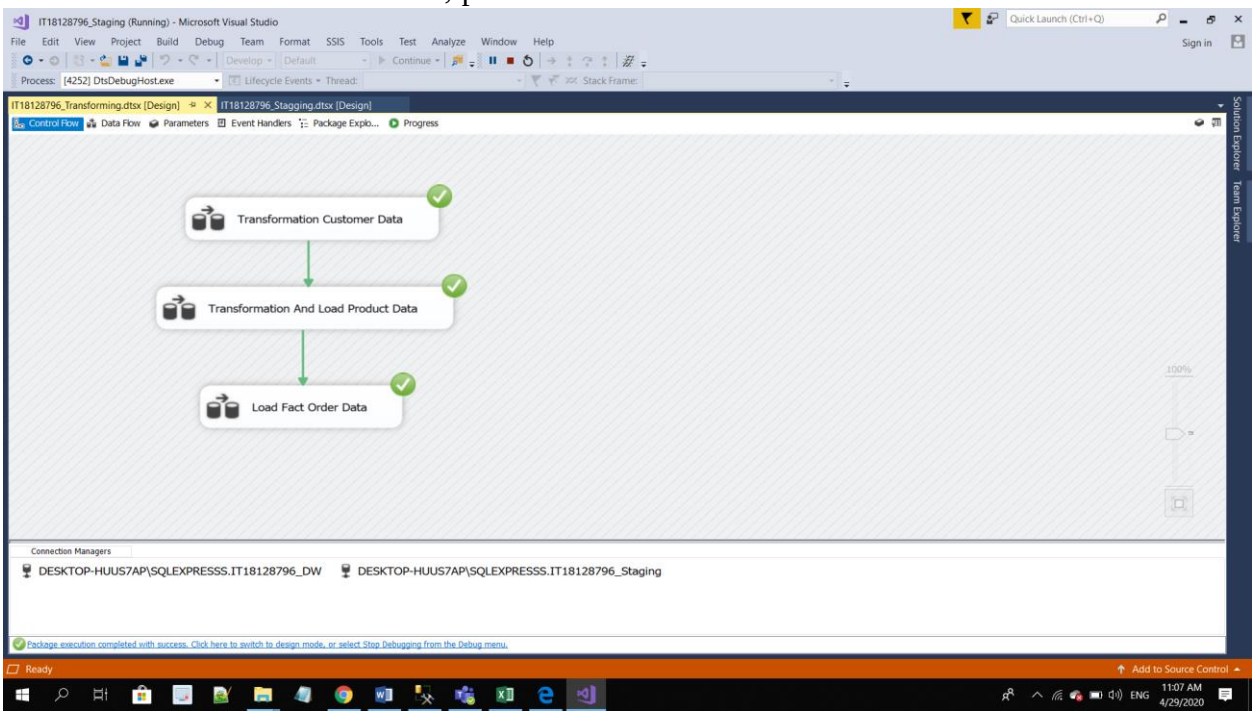
Start Transformation



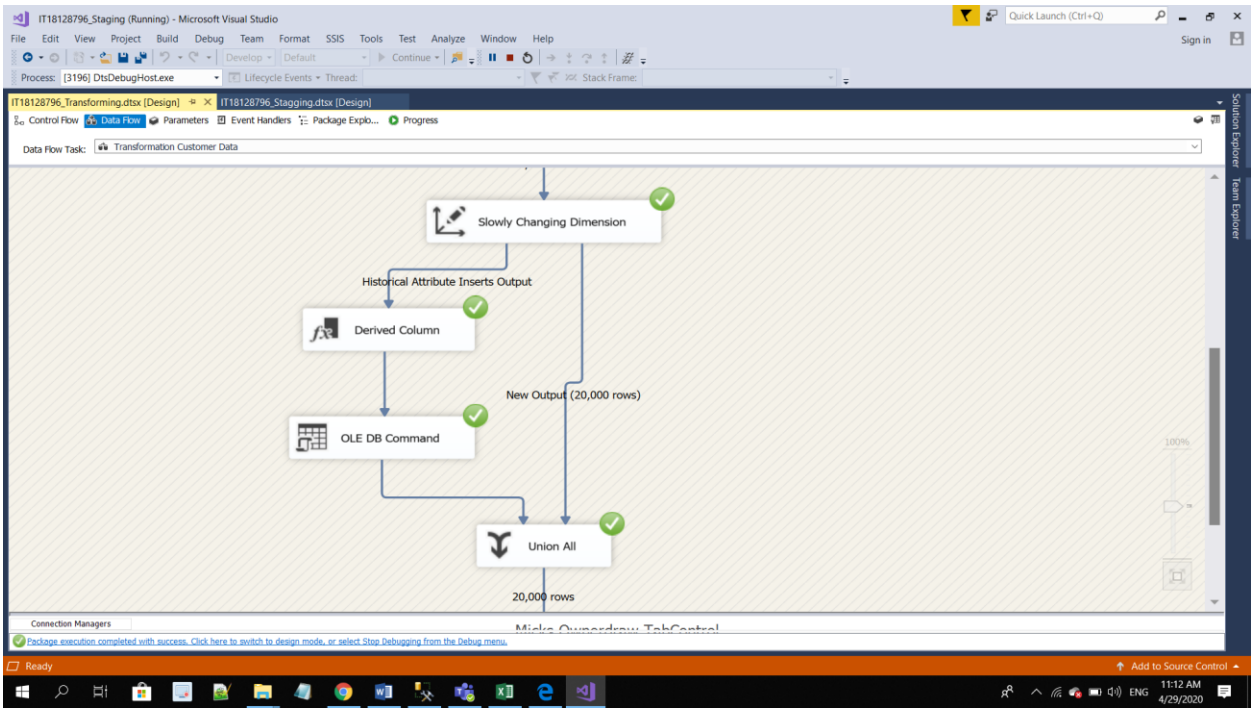
Staging is successful



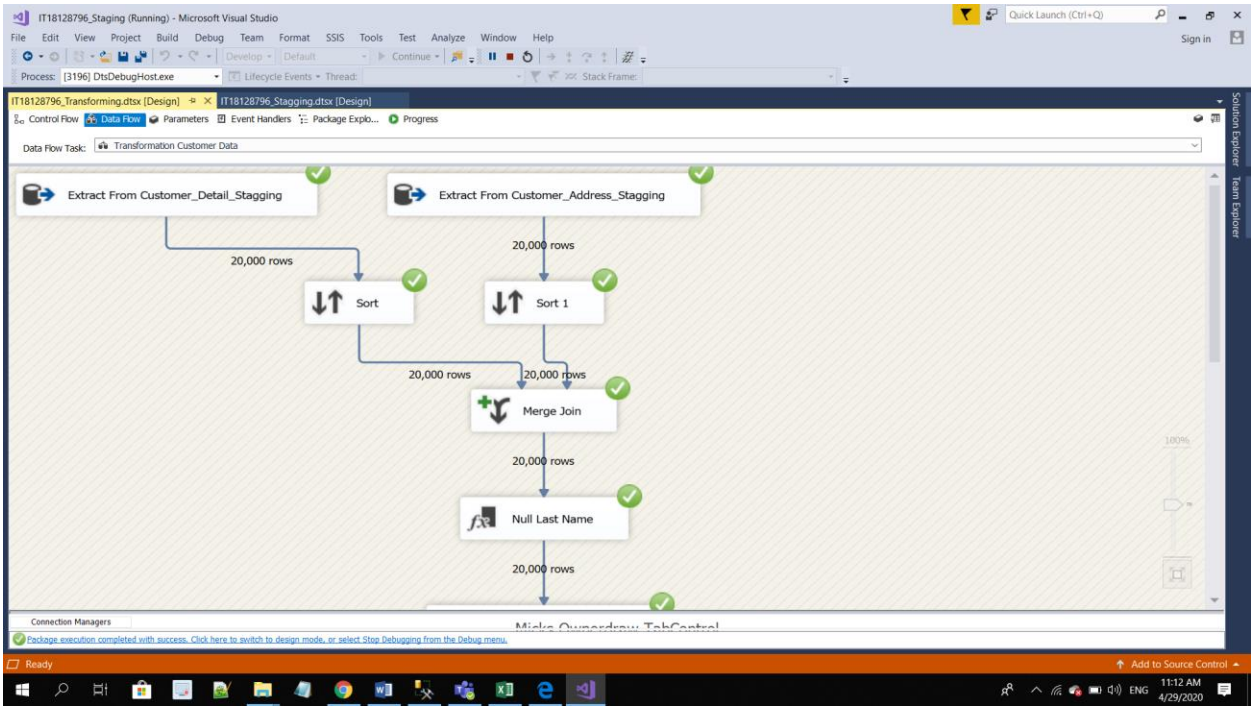
Transformation and load cutomer, product and order



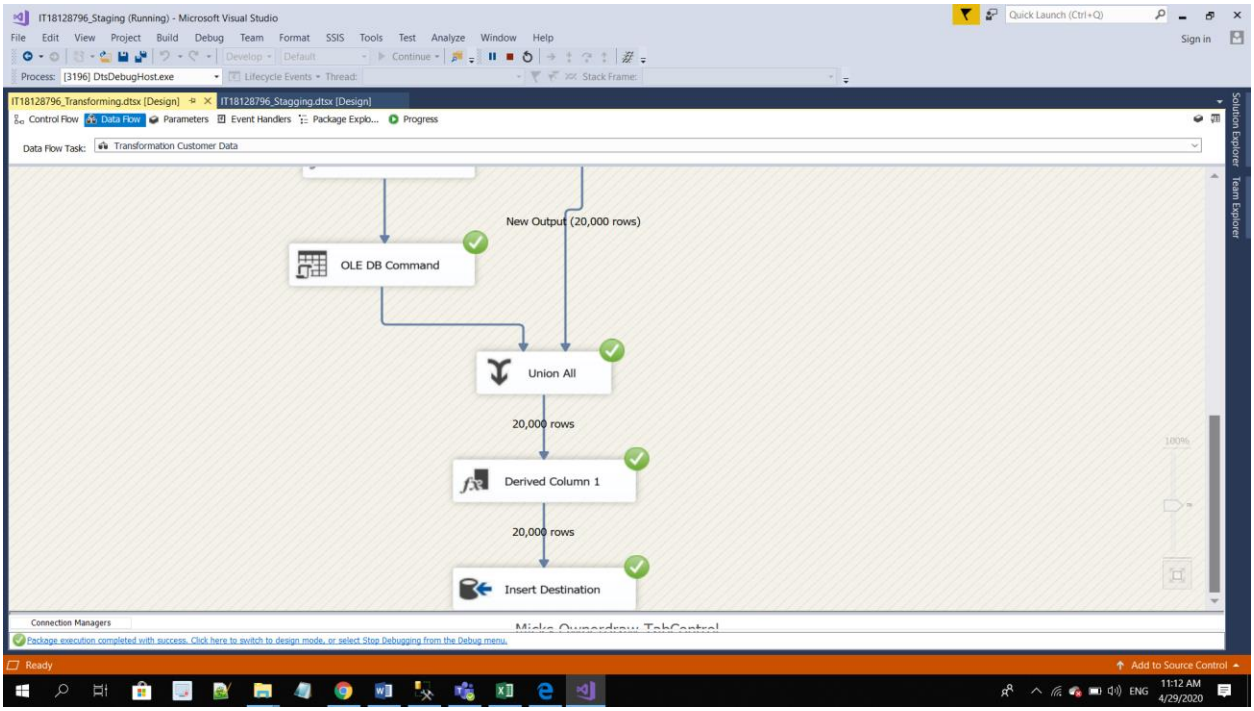
This show the configuration of the slowly changing dimension.



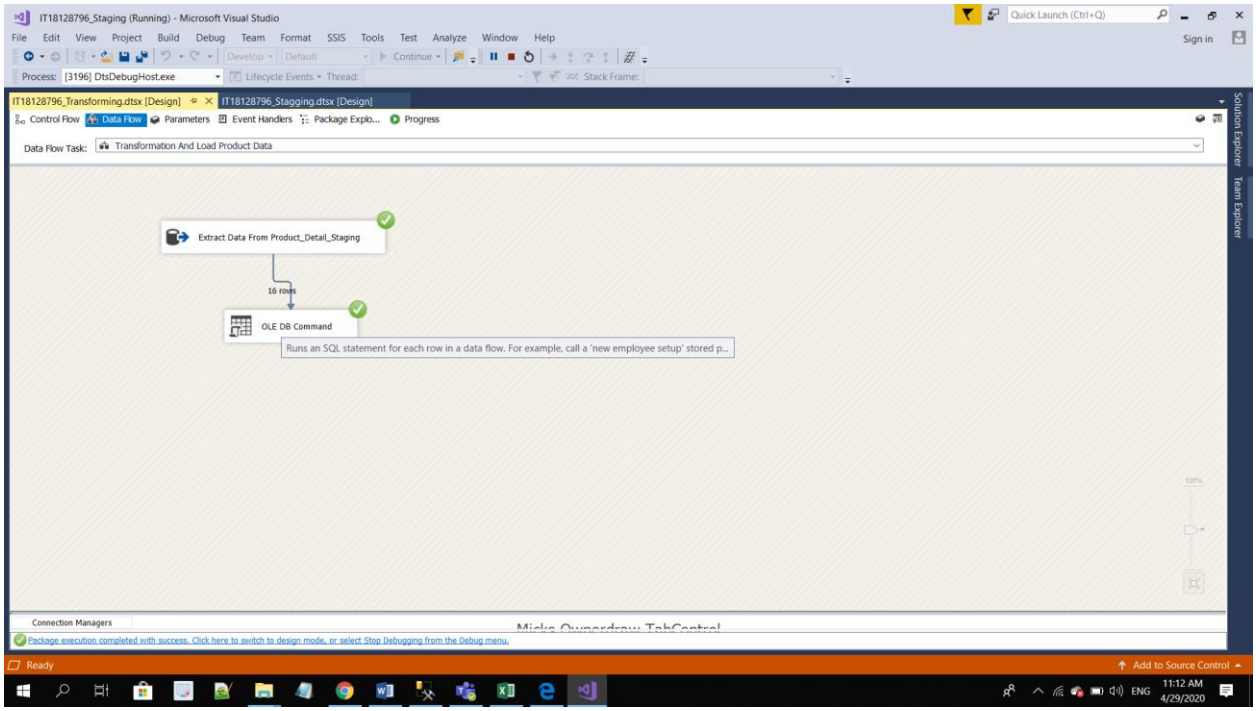
This show if last name is empty make it null.



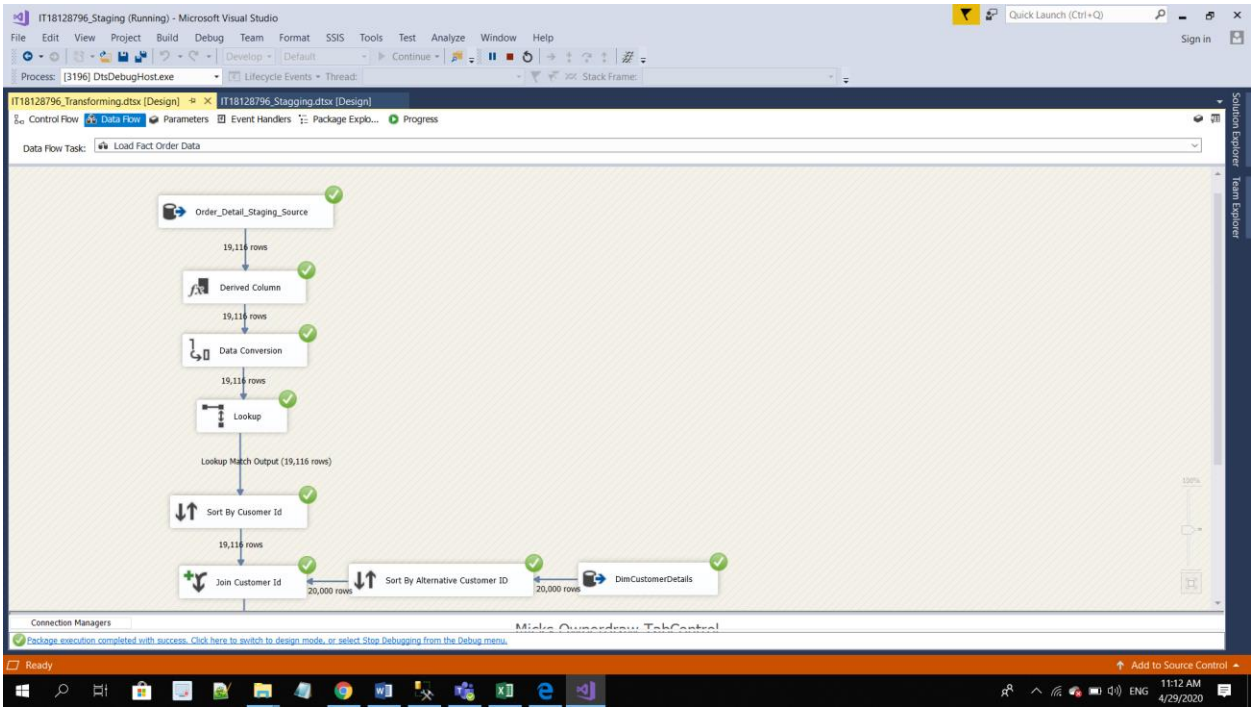
Finally load to data ware house customer dim table



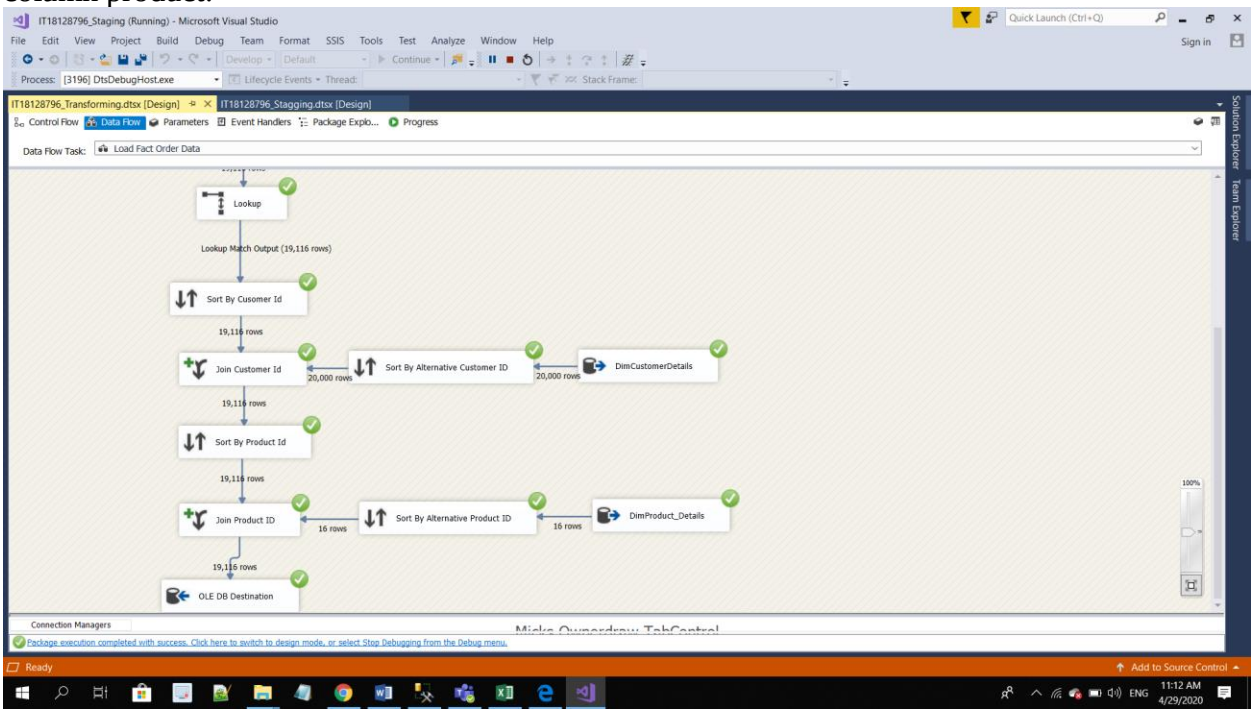
Product details are directly upload to the data ware house dim product table



Here order date column match with the dim date whether it exist. Match dim customer with order column cutomer.



Here order date column match with the dim date whether it exist. Match dim product with order column product.



This display data warehouse load table