

Data Warehousing & Business

Intelligence

Shipping Details

Assignment 1

Submitted by:

**K.T.P.M.Kariyawasam IT16063310**

Submitted to:

**MR. Sheron Dinushka**

1.Introduction

My dataset is about Shipping details System. I found this data set from internet then I created a scenario. According to that scenario I modified my my data set. So I used data from a superstore data set(like Customer Address, Date column details.) Some of data I create(IDs, Gender,tittle,..etc).

Content

The data was downloaded from the superstore database is ranging from January 2014 to December 2017.

Data link: https://data.qld.gov.au/dataset?tags=Department+of+Education

Some key Details in Data Set

CustomerId : Customer Id

CustomerAddress: Customer Address

Title: Customer Title

Name: Customer Name

Gender: Customer Gender

CustomerTye: Customer Type

PhoneNo: Customer Phone Number

ShippingID: ShippingID

ShippingDate: Shipping Date

ShippingOption: Shipping Option

ShipFlag: Country of Ship

ProductID: ProductID

ProductCat : Product Category ID

ProductSubCat : Product Sub category ID

ProductName :Name of the product

Condition : Condition Of the Product

UnitPrice : Unit Price Of the product

ProdCatName : Category Name

ProductSubCatName : Sub Category Name

Region : Region of cus adrress

Postal COde

State : State of cus address

City : City Of cus address

Country : Country Of address

2.Preparation of Data Sources

In order to data extraction need to prepare the data sources. From my main data source, I extracted three types of data sources.

1. Database backup (.bak)

2. Text file (.txt)

3. Excel file (.xlsx)

There are eight source tables.

Customer [Excel file (.xlsx)]

CustomerId

Title

Name

Gender

CustomerType

Phone

Customer Address[Text file (.txt)]

Country

City

State

Postal Code

Region

Shipping [Excel file (.xlsx)]

ShippingID

ShippingDate

ShippingOption

ShipName

ShipFlag

Product[Excel file (.xlsx)]

ProductID

ProductName

ProductCat

ProductSubCat

Condition

UnitPrice

Orders[Excel file (.xlsx)]

OrderID

OrderDate

OrderType

Quantity

Discount

ProductCategory[Excel file (.xlsx)]

ProductCategoryID

ProductCategoryName

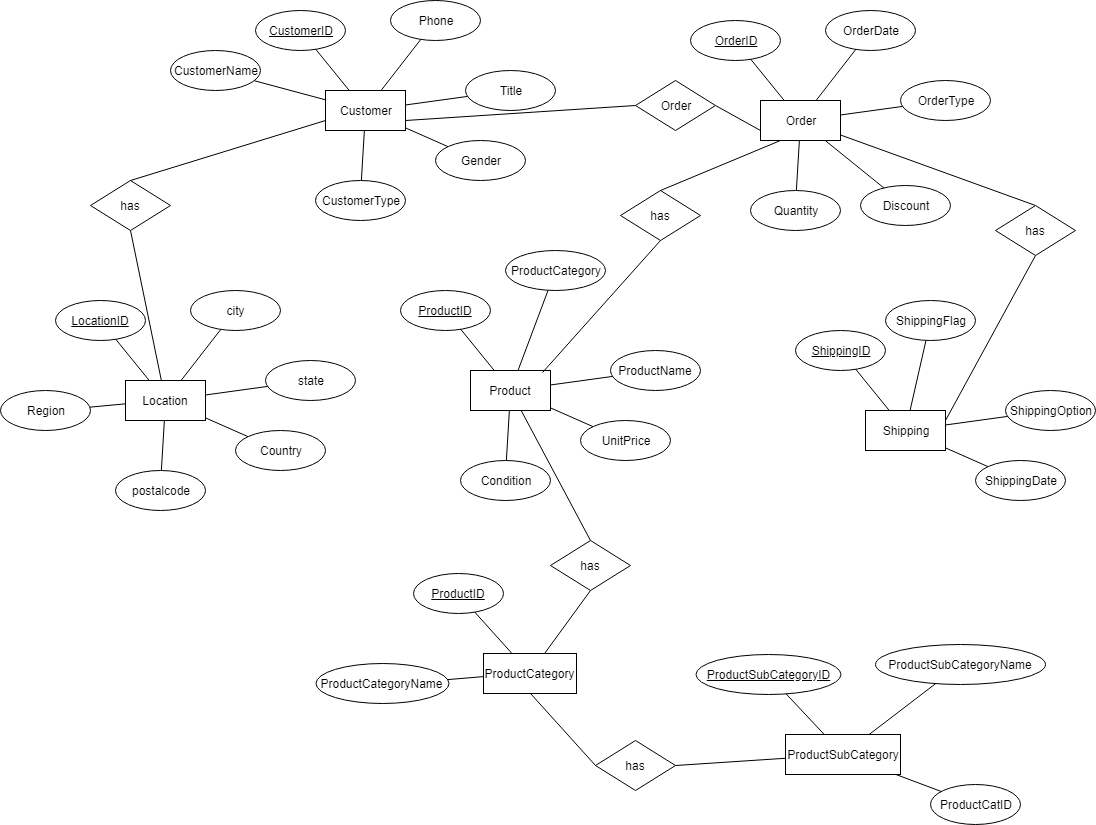
ProductSubCategory[Excel file (.xlsx)]

ProductSubCateID

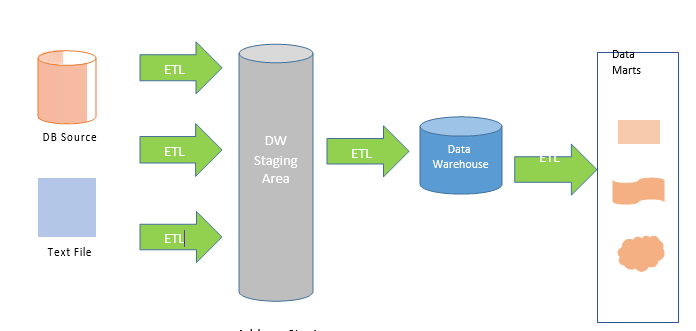
ProductCategoryID

ProductSubCatNAme

ER Diagram



**3.Solution Architecture**



**Architecture Components**

 **Data Sources**

Operational System (Transaction)

External sources

 **Extract, Transform, and Load**

Extract – reading data from source systems

Transform – Combine data from multiple sources, De-duplicating

Load – loading data to destination, Surrogate key assignment, Foreign key constraint checks, Indexing

**• Data Warehouse**

EDW vs Data Mart

Dimensional Modeling - Facts & Dimension

Many Schemas -In here used Star schema

4.Data warehouse Desigh and Development

**5.ETL development**

**ETL (Extract-Transform-Load)**

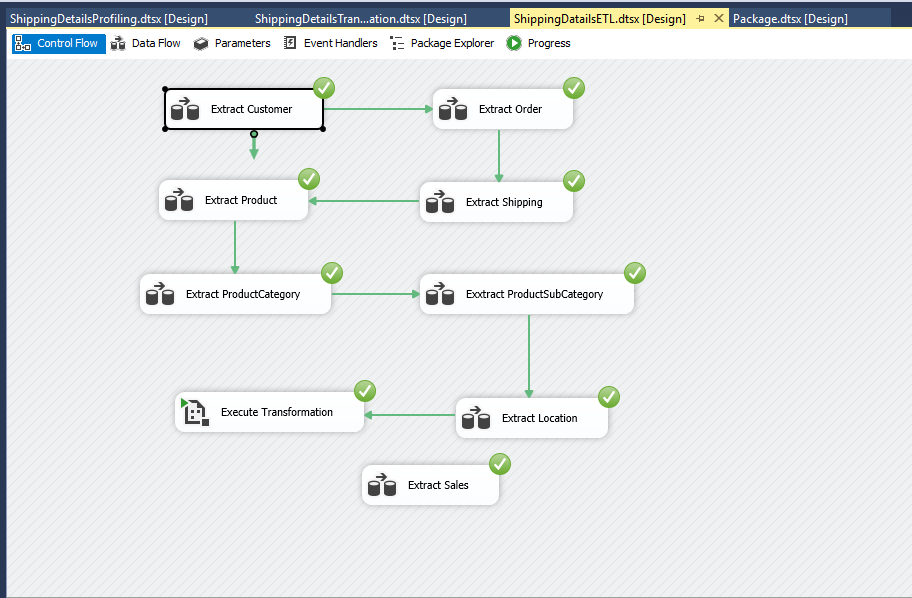
ETL standard for Extract-Transform-Load. ETL covers a process of how the data are loaded from the source system to the data warehouse.

First, the extract function reads data from a specified source database and extracts a desired subset of data. Next, the transform function works with the acquired data - using rules or lookup tables, or creating combinations with other data - to convert it to the desired state. Finally, the load function is used to write the resulting data to a target database.

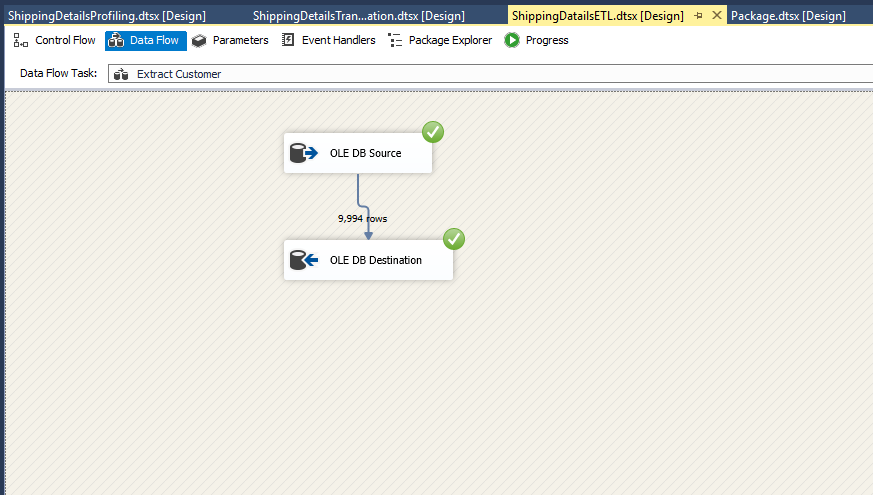
 **Extract**

The main objective of the extract step is to retrieve all the required data from the source system with as little resources as possible.

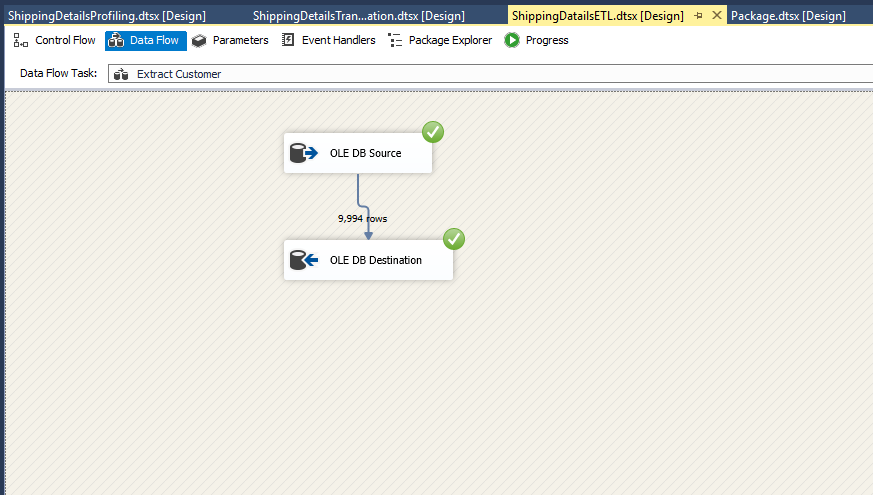
**Data Extraction from Sources to Staging Tables**



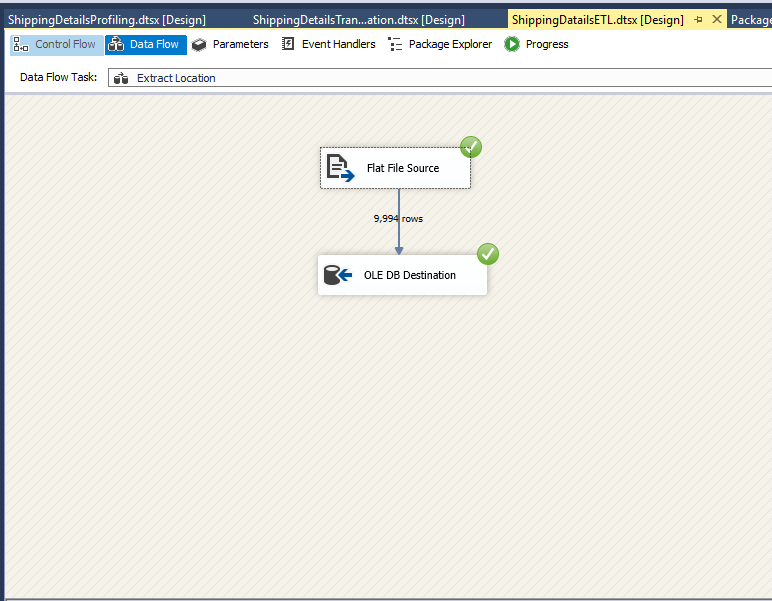
Extract Customer



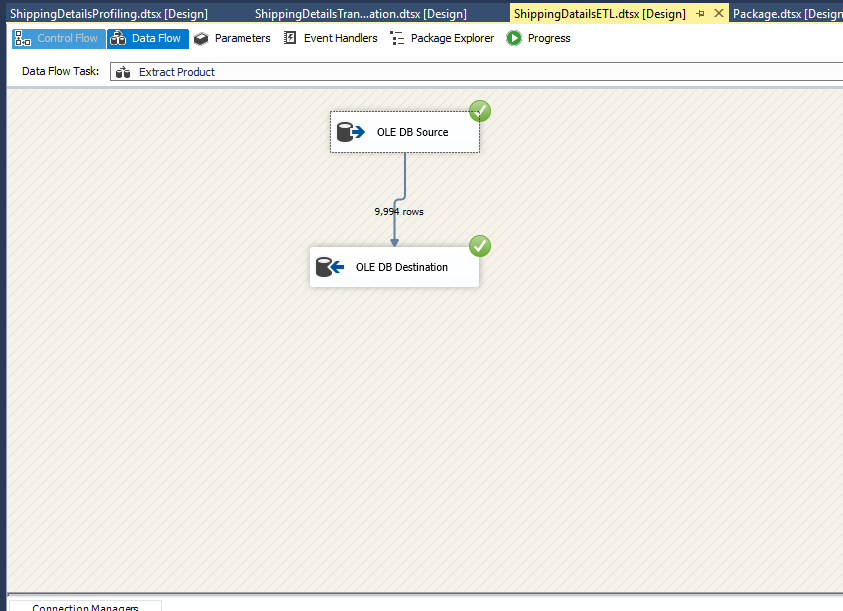
**Extract Order**

****

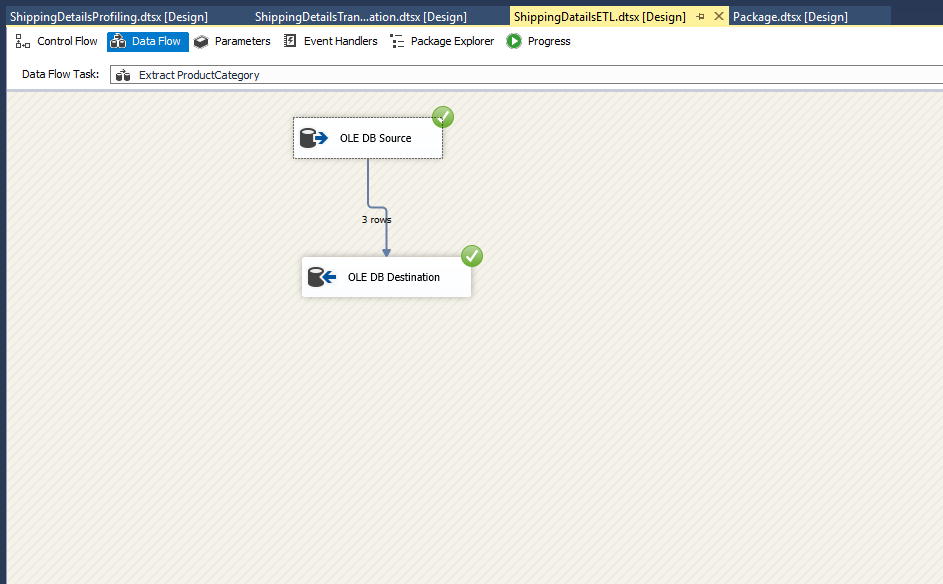
**Extract Location**

****

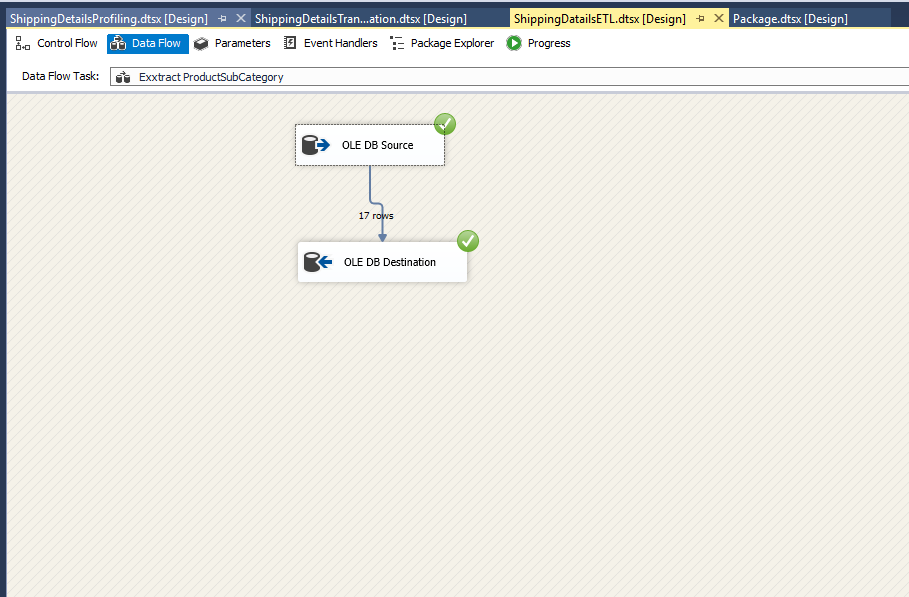
**Extract Product**

****

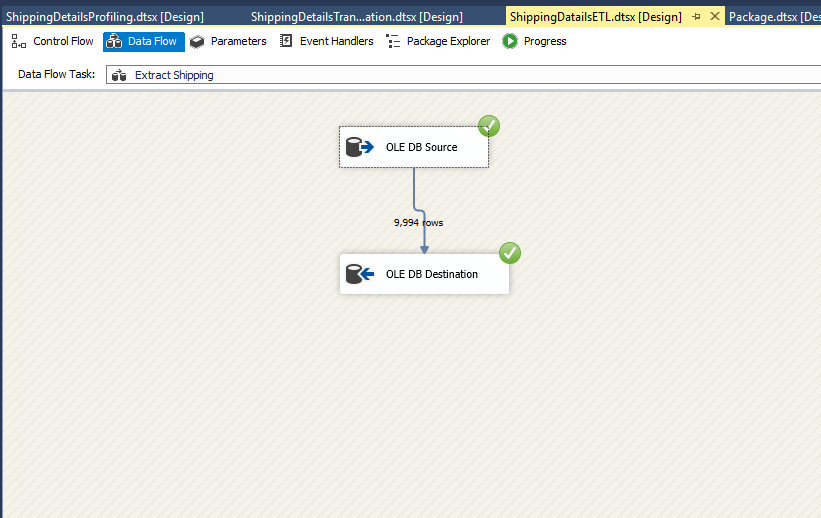
**Extract Product Category**

****

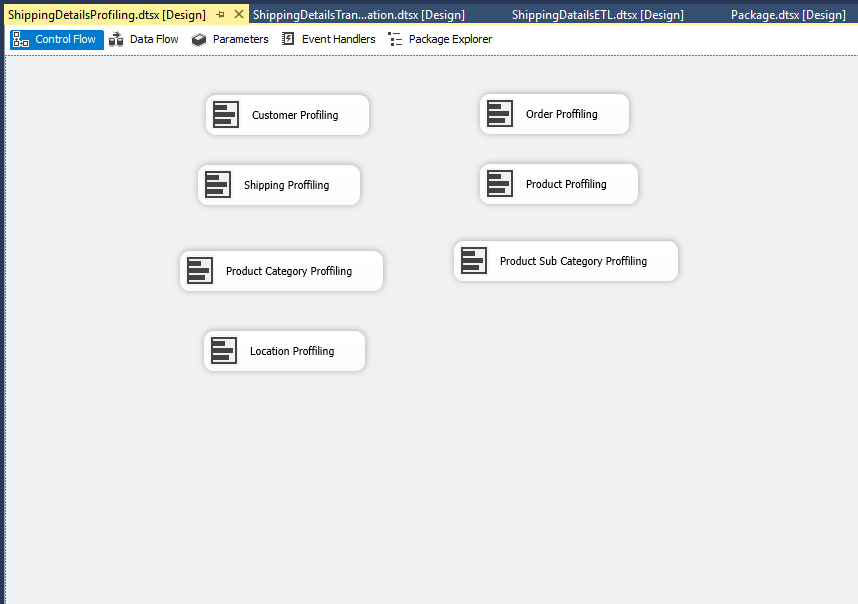
**Extract Product Sub Category**

****

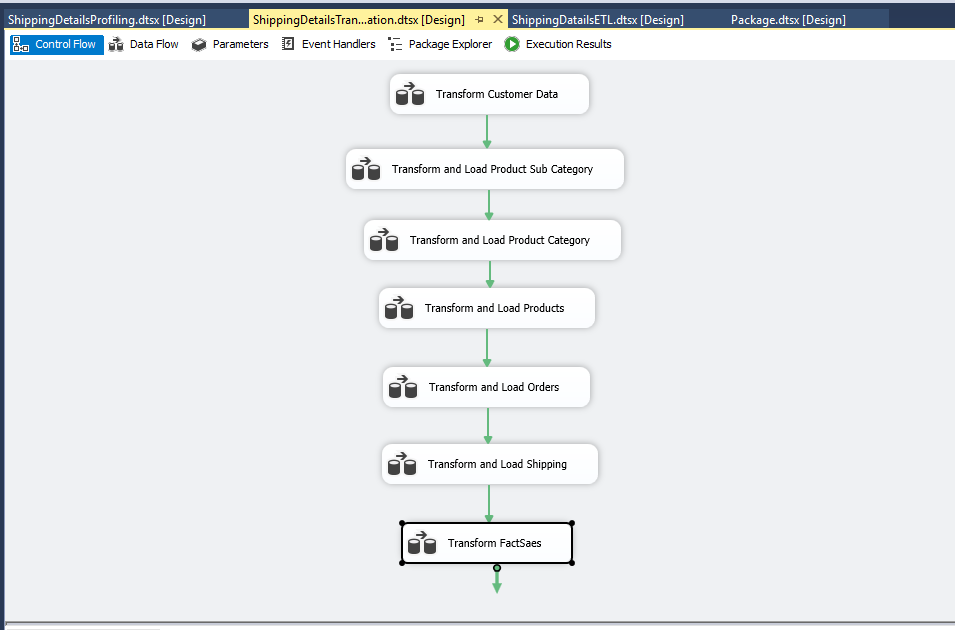
**Extract Shipping**

****

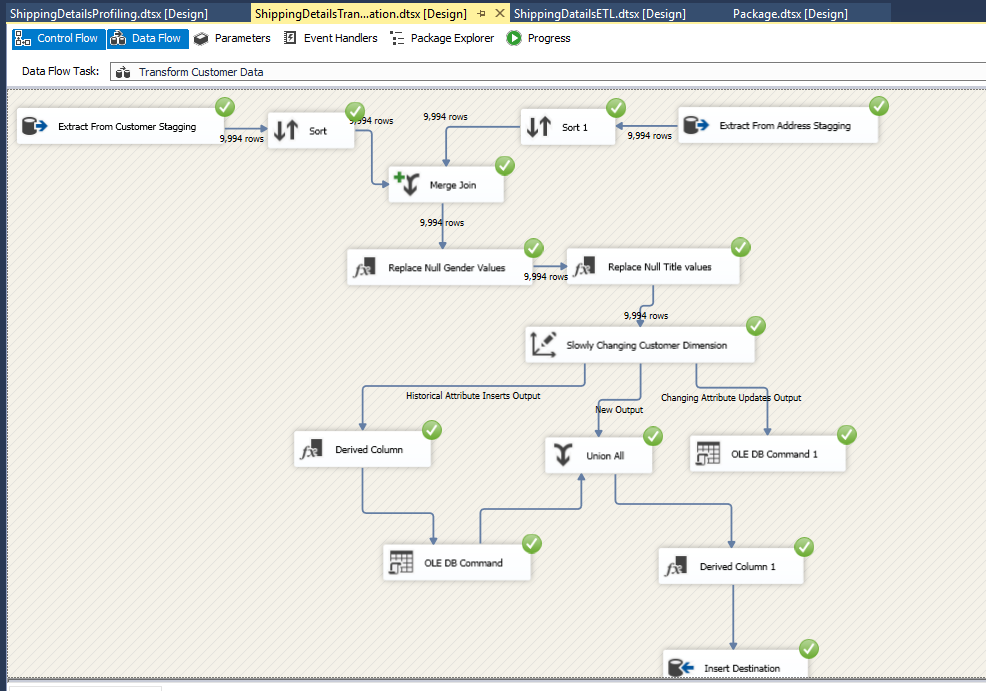
**Data Profiling**

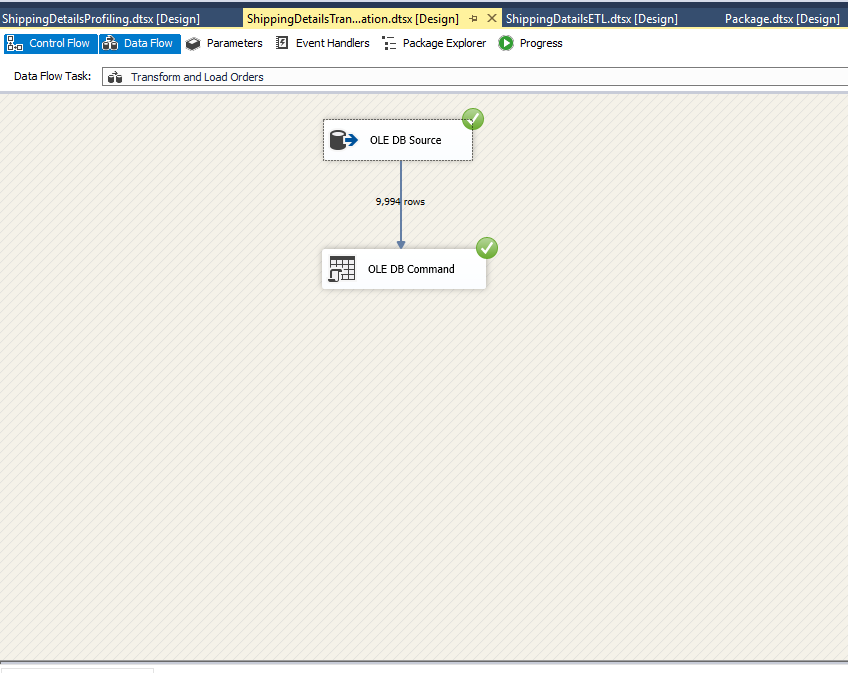


Data Transformation Control Flow

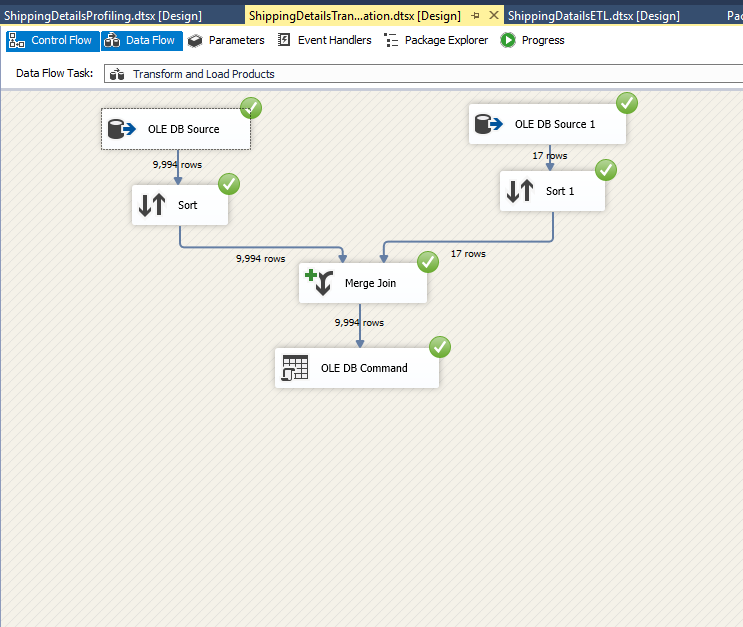


Transform Customer Data

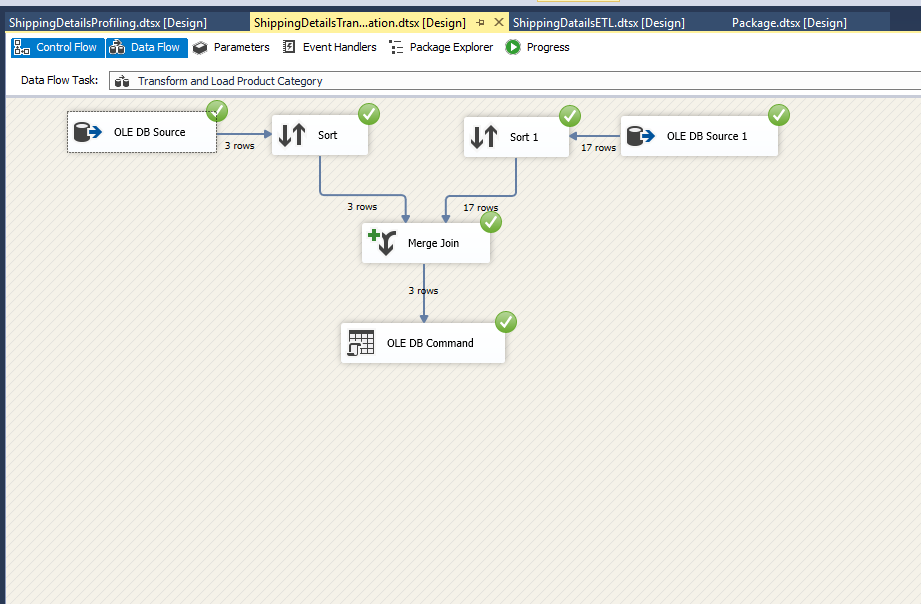


Transform Order

Transform Product



Transform Category



Transform Fact

