# Cover page

**Data Warehousing and Business Intelligence**

**IT3021**

**Assignment 1**

**Delivery Center: Food & Goods orders in Brazil**

**2022**

**IT20220860**

**Karawita K.S.A**

Contents

[Cover page 1](#_Toc105416914)

[Data Set Selection 3](#_Toc105416915)

[ER diagram 4](#_Toc105416916)

[Preparation of Data Sources 5](#_Toc105416917)

[Solution Architecture 6](#_Toc105416918)

[Design and Development 7](#_Toc105416919)

[Dimensional Model 7](#_Toc105416920)

[Hierarchies 8](#_Toc105416921)

[ETL Development Process 9](#_Toc105416922)

[Truncating Staging tables 9](#_Toc105416923)

[Loading to Staging 10](#_Toc105416924)

[Transforming and Loading to Data Warehouse 12](#_Toc105416925)

[Data Profiling 18](#_Toc105416926)

[References 18](#_Toc105416927)

# Data Set Selection

<https://www.kaggle.com/datasets/nosbielcs/brazilian-delivery-center>

The Delivery Center is a platform that integrates retailers and marketplaces, creating a healthy ecosystem for sales of goods and food in Brazilian retail. We currently have a register with more than 14000 items. Thousands of orders and deliveries are processed daily with a network of thousands of merchants and delivery partners spread across all country regions.

All this generates data and more data all the time. In view of this, this network business is increasingly data-driven, that is, using data to make decisions and in a vision of the future, we know that using data intelligently can be our great differential in the market.

Channels: This dataset has information about the sales channels (marketplaces) where our retailers' goods and food are sold.

Deliveries: This dataset has information about deliveries made by our partner delivery partners.

Orders: This dataset has information about sales processed through the Delivery Center platform.

Payments: This dataset has information about payments made to the Delivery Center.

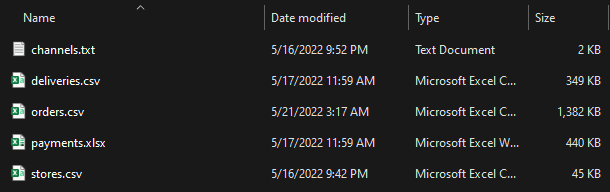
Stores: This dataset has information about the store owners. They use the Delivery Center Platform to sell their items on marketplaces.

# ER diagram

Diagram

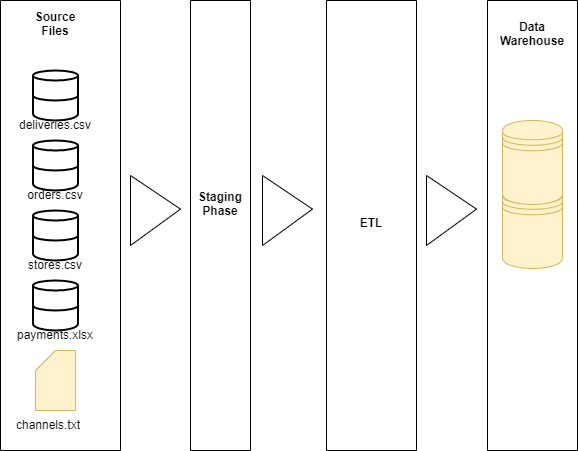
Description automatically generated

# Preparation of Data Sources



The original data set contained more than 321 000 in some files. The database was varied with various types of data source files including csv, excel, txt. I filtered data to contain not more than 15000 records as it was more than enough to develop the solution perfectly.

# Solution Architecture



Data sources containing data of orders, stores, and deliveries are in csv formats. They were imported as flat files to the source DB. Then in the staging phase the excel file and the text file containing payment data and channels data respectively were inserted in to the staging DB. And finally to the Data Warehouse through the ETL process.

# Design and Development

## Dimensional Model

Star schema was selected to design the Data Warehouse for Delivery Center Food & Goods orders in Brazil. There are mainly four, dimensional tables and the fact table. All these dimensional tables are linked with the fact table.

Dimension tables and fact table:

• DimChannels

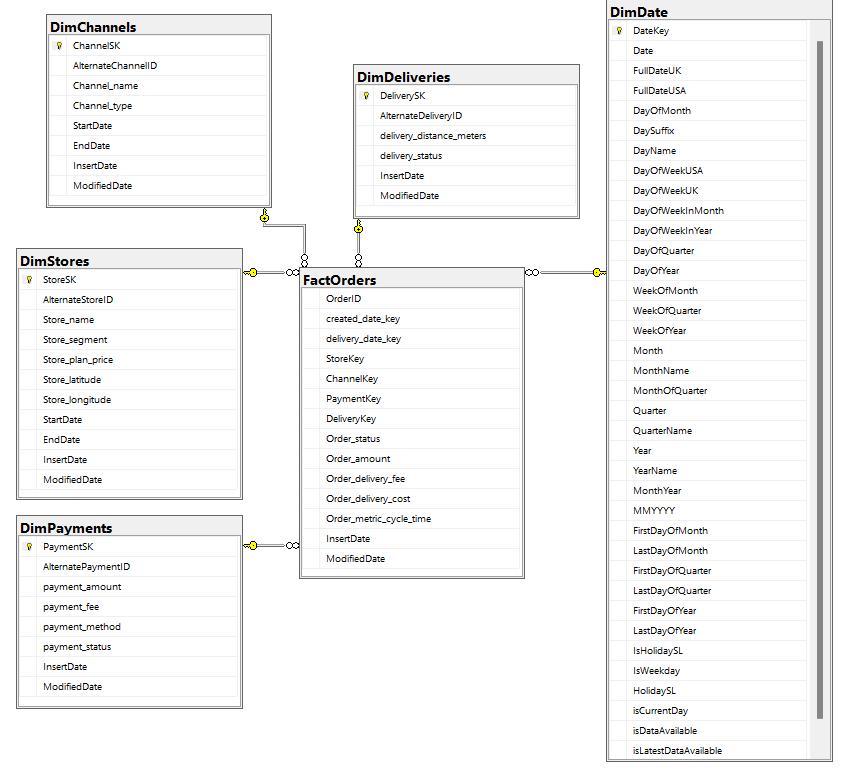
• DimDeliveries

• DimPayments

• DimStores

• FactOrders

## Hierarchies



## ETL Development Process

### Truncating Staging tables

It is required to truncate processes for loading Staging tables with PreExecute Event Hander. That deletes all the existing rows in the target table before loading any new data. This process prevents data duplicating

A screenshot of a computer

Description automatically generated with medium confidenceA screenshot of a computer

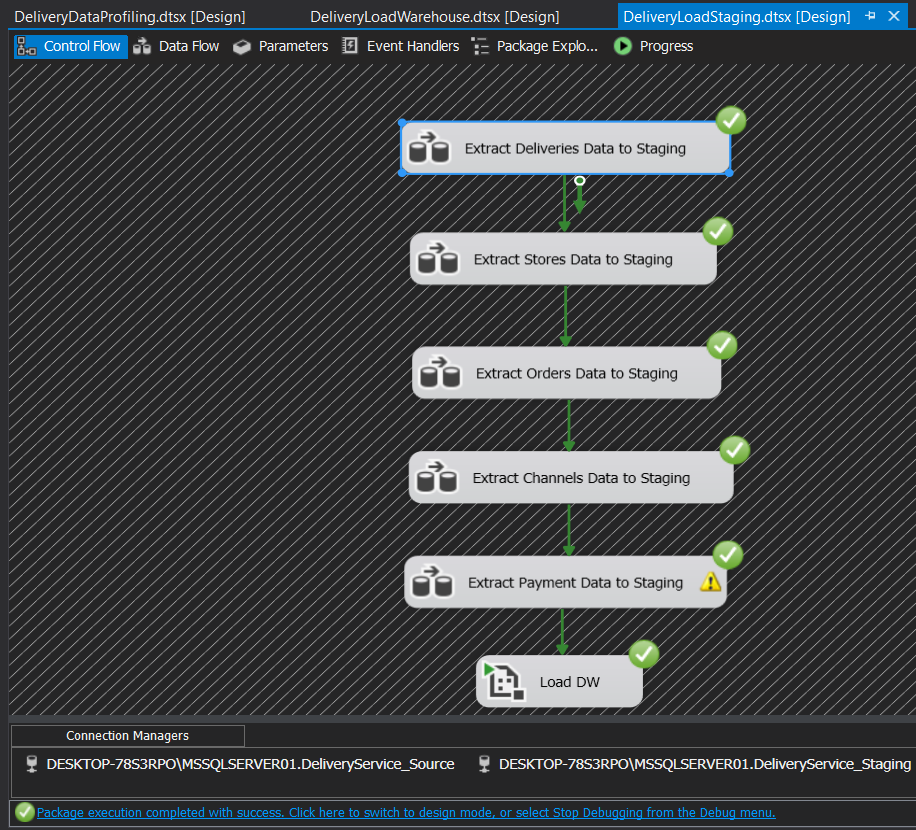
Description automatically generated with medium confidenceGraphical user interface, application

Description automatically generatedA screenshot of a computer

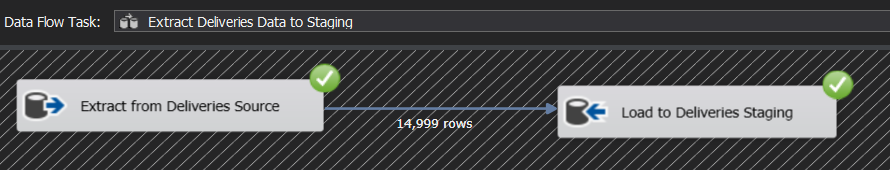
Description automatically generated with medium confidenceBackground pattern

Description automatically generated with medium confidence

### Loading to Staging



#### Extract Deliveries Data to Staging

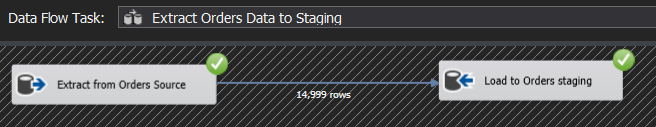


#### Extract Stores Data to Staging

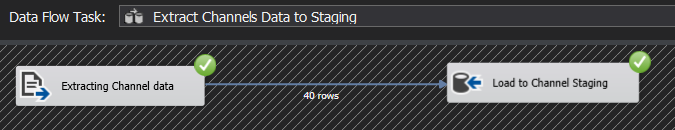
Graphical user interface, application

Description automatically generated

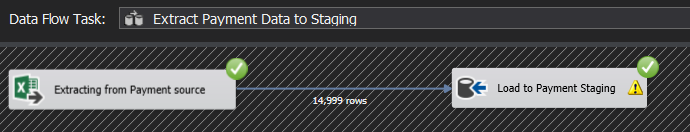
#### Extract Orders Data to Staging



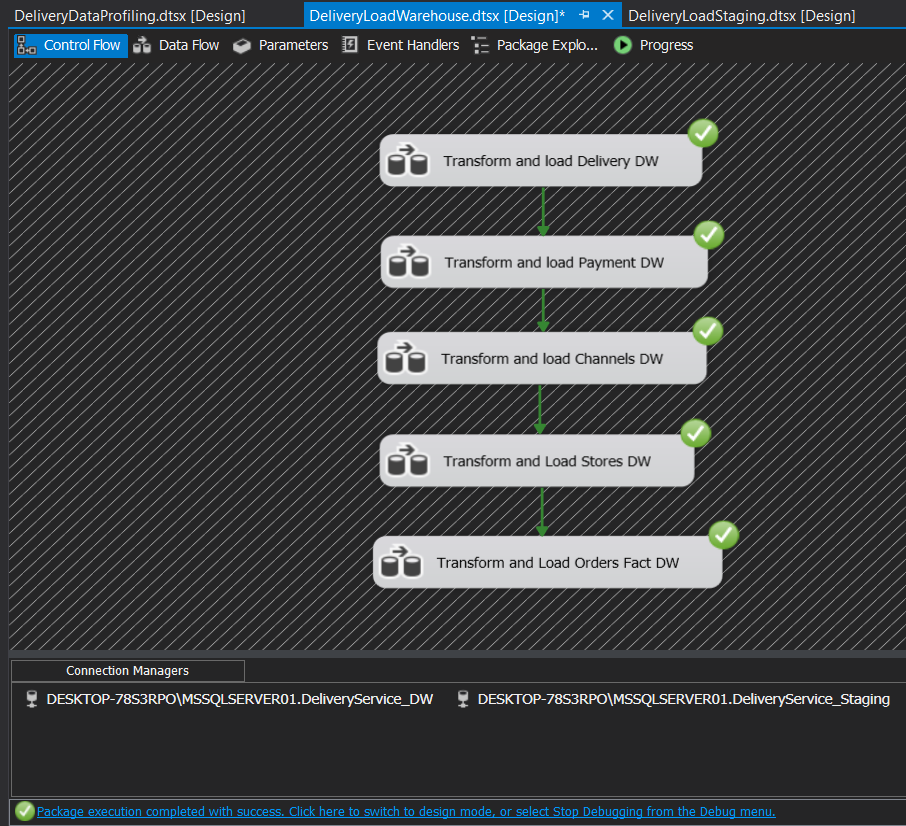
#### Extract Channels Data to Staging



#### Extract Payments Data to Staging



### Transforming and Loading to Data Warehouse



#### Transform and Load Delivery Dimension Table

Graphical user interface, application, website

Description automatically generated

**UpdateDimDeliveries** procedure was used here to update the existing values with the new values entered.

CREATE PROCEDURE dbo.UpdateDimDeliveries

@DeliveryID int,

@delivery\_distance\_meters int,

@delivery\_status nvarchar(50)

AS

BEGIN

if not exists (select DeliverySK

from dbo.DimDeliveries

where AlternateDeliveryID = @DeliveryID)

BEGIN

insert into dbo.DimDeliveries

(AlternateDeliveryID, delivery\_distance\_meters, delivery\_status, InsertDate, ModifiedDate)

values

(@DeliveryID, @delivery\_distance\_meters, @delivery\_status, GETDATE(), GETDATE())

END;

if exists (select DeliverySK

from dbo.DimDeliveries

where AlternateDeliveryID = @DeliveryID)

BEGIN

update dbo.DimDeliveries

set delivery\_distance\_meters = @delivery\_distance\_meters,

delivery\_status = @delivery\_status,

ModifiedDate = GETDATE()

where AlternateDeliveryID = @DeliveryID

END;

END;

#### Transform and Load Payments Dimension Table

Graphical user interface, application

Description automatically generated

**UpdateDimPayments** procedure was used here to update the existing values with the new values entered.

CREATE PROCEDURE dbo.UpdateDimPayments

@PaymentID int,

@payment\_amount float,

@payment\_fee float,

@payment\_method nvarchar(50),

@payment\_status nvarchar(50)

AS

BEGIN

if not exists (select PaymentSK

from dbo.DimPayments

where AlternatePaymentID = @PaymentID)

BEGIN

insert into dbo.DimPayments

(AlternatePaymentID, payment\_amount, payment\_fee, payment\_method, payment\_status, InsertDate, ModifiedDate)

values

(@PaymentID, @payment\_amount, @payment\_fee, @payment\_method, @payment\_status, GETDATE(), GETDATE())

END;

if exists (select PaymentSK

from dbo.DimPayments

where AlternatePaymentID = @PaymentID)

BEGIN

update dbo.DimPayments

set payment\_amount = @payment\_amount,

payment\_fee = @payment\_fee,

payment\_method = @payment\_method,

payment\_status = @payment\_status,

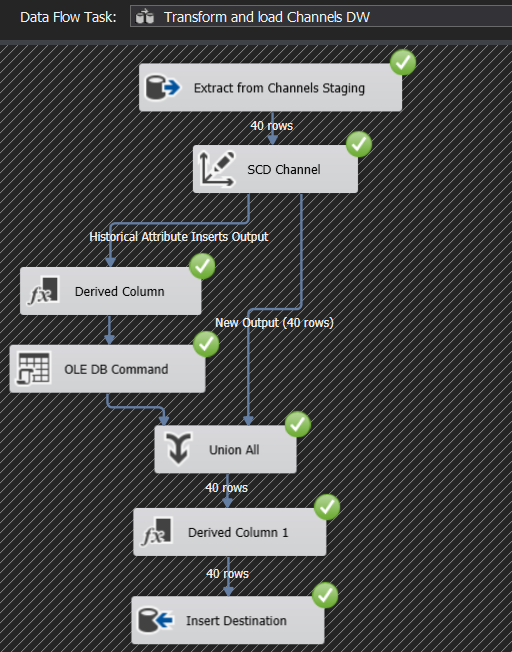
ModifiedDate = GETDATE()

where AlternatePaymentID = @PaymentID

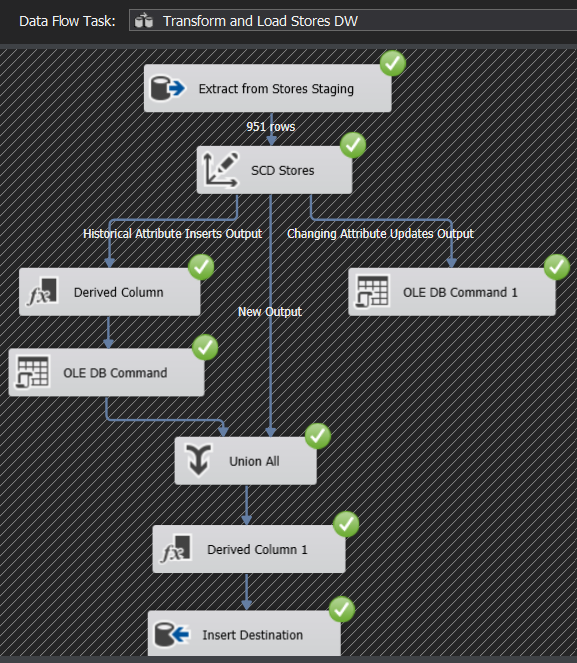
END;

END;

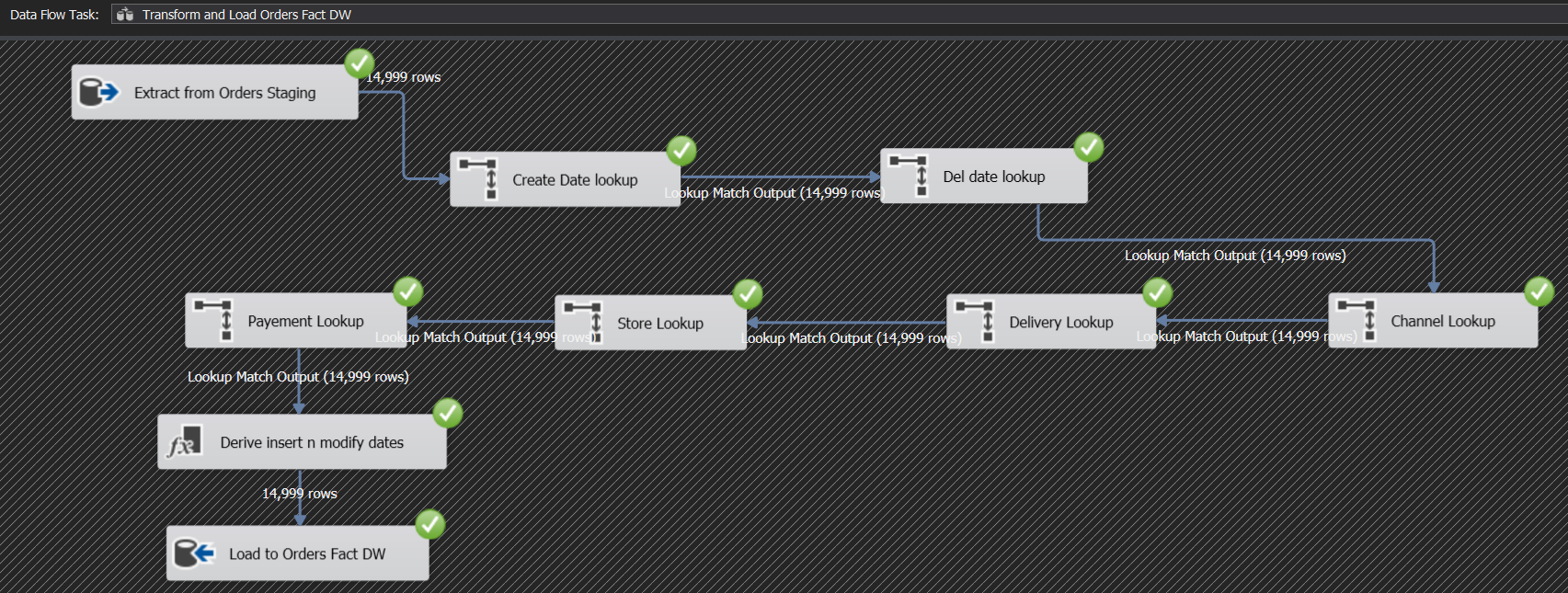
#### Transform and Load Channels Dimension Table



#### Transform and Load Stores Dimension Table



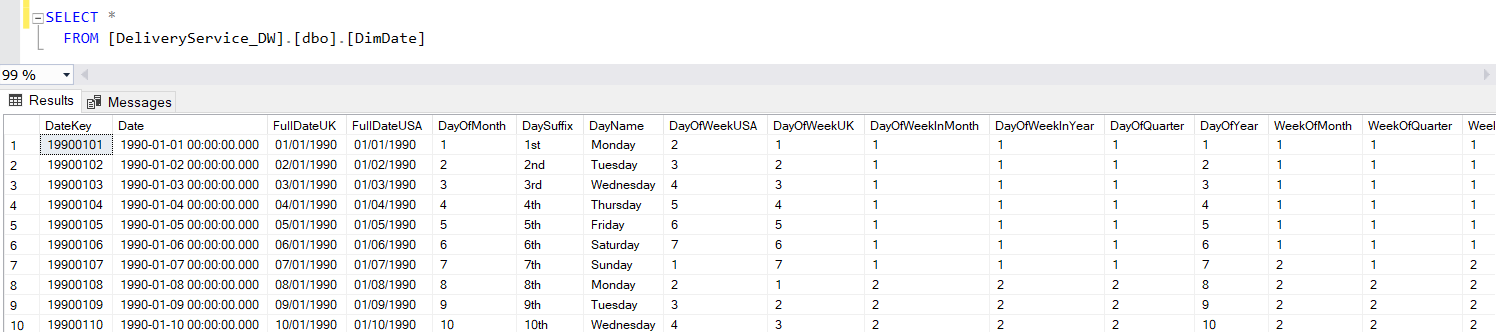
#### Transform and Load Orders Fact Table



#### Load DimDate Dimension Table

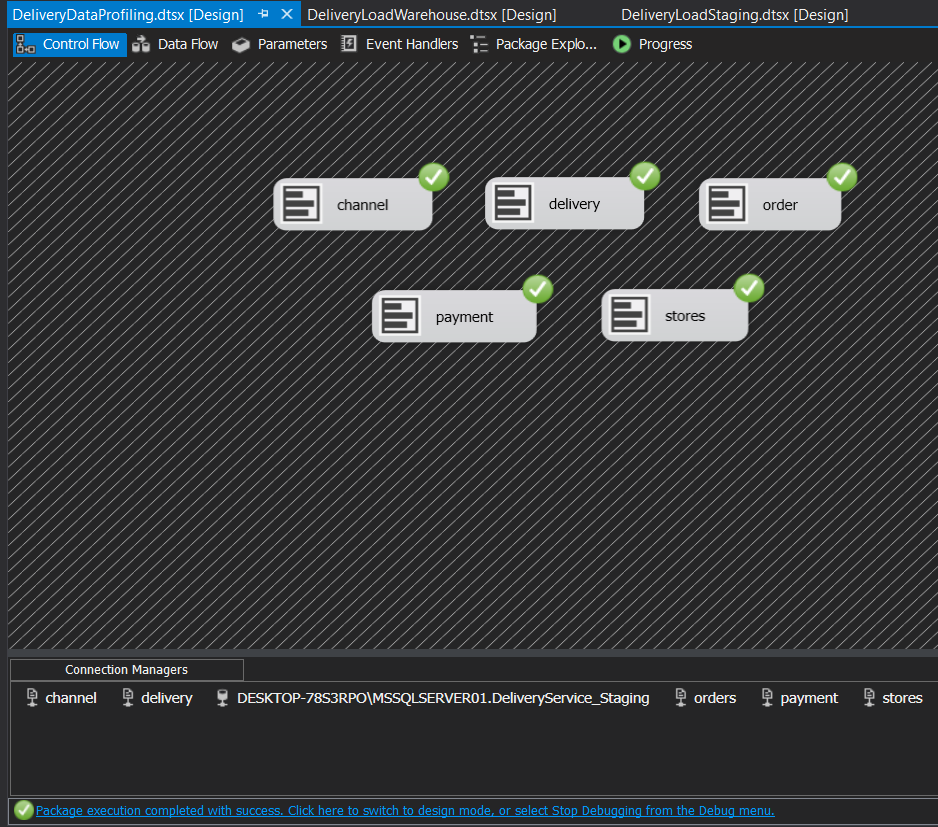
In the process I have used the DimDate Table that have been provided to get reference to 2 columns in the fact Table.

* created\_date\_key
* delivery\_date\_key



### Data Profiling

ETL data profiling is a detailed analysis of source data. It attempts to identify the structure, quality, and content of the source data, as well as its relationships with other data. It has been used during the extraction, transformation, and loading (ETL) process.



# References

<https://docs.microsoft.com/en-us/sql/sql-server/end-of-support/sql-server-end-of-support-overview?view=sql-server-ver16>

<https://www.w3schools.com/sql/>

<https://stackoverflow.com/>

<https://www.toolbox.com/tech/data-management/question/is-it-necessary-to-truncate-table-before-running-loading-mapping-workflow-021012/>

Thank You