



Data Warehousing and Business Intelligence **(IT3021)**

Airline Passenger Satisfaction **Assignment 1**

Submitted to

Sri Lanka Institute of Information Technology

Bachelor of Science Special Honors Degree in Data Science

Submitted by: Egodawattaarachchige A.B. (IT20141974)

Table of Contents

1. Data Set Selection.....	3
2. Preparation of Data Source.....	5
3. Solution Architecture	6
4. Datawarehouse Design & Development	7
5. ETL Development	8
6. References	21

List of Figures

Figure 1 : ER Diagram 4

Figure 2 : High-level Architecture 6

Figure 3 : Dimensional Model 7

Figure 4 : ETL Process.....8

1.Data selection

Background

This dataset gives us the idea about the satisfaction level of passengers about the services that were provided by an airline company during a flight journey.

Data Set : Airline Passenger Satisfaction

Site : Kaggle

Source Link : <https://www.kaggle.com/binaryjoker/airline-passenger-satisfaction>

This is a dataset of airline passenger satisfaction made of an airline company. It contains the data of passenger's satisfaction level of the services provided by the airline company. The passengers ID, gender, customer type, age, customer class and his satisfaction on the various services are mentioned in the table. The original dataset has been edited and arranged to suit the scenario and the requirements of the assignment. The initial data set contains one csv file recording all the above details.

A high level ER Diagram is shown below to get an overview of the "Airline Passenger Satisfaction" dataset.

ER Diagram

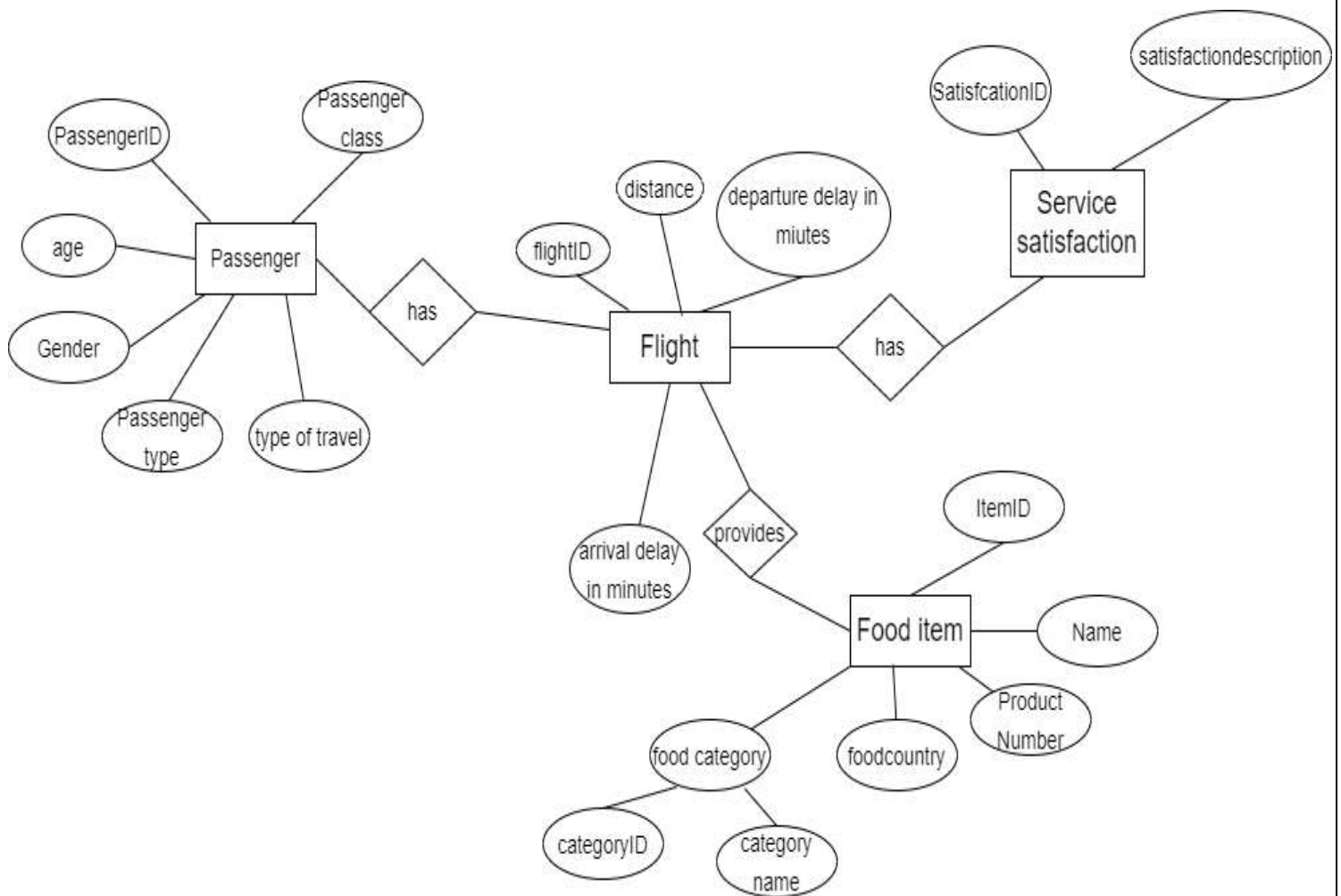


Figure 1: ER diagram

2.Preparation of Data Sources

The original data set contains more than one hundred thousand records with respective variables. Since majority of variables contained null values they were removed and also additional data were added to the dataset in order to enhance the overall quality . Also additional csv were added to the project to get clear conclusions. All data sources were provided in .csv format by the website. In preparation of the data sources a .csv file was imported and separate database tables and text files were created.

The source file format of the original data source was CSV and it was modified to TXT and XLXS

- Excel file = Passengers,flight,FoodItem,FoodCategory,Satisfaction
- Text file = address

The source table contains of below tables with the primary key:

- dbo.flight (pk=FlightID)
- dbo.FoodCategory (pk=FoodCategoryID)
- dbo.FoodItem (pk=FoodID)
- dbo.Passengers (pk=PassengerID)
- dbo.satisfaction (pk=SatisfactionID)

Mostly nvarchar, varchar, int ,datetime are the data type that were used in these tables

3. Solution Architecture

The diagram below presents the overall architecture of the Datawarehouse and Business Intelligence Solution, that has being implemented for Airline Passenger Satisfaction Dataset.

The architecture comprises of four components, Data Sources, ETL (Extract, Transform, Load), Storage Layer Components and Data Consumption.

- **Data sources** : This comprises of structured data in the format of text and excel and the formats are stored in the local folder.
- **ETL** : This is performed at two occurrences, to begin with occasion when extracting data from the sources and stacking it to the Staging Layer and in moment occurrence when performing extraction, and change on Staging Layer to load data into Datawarehouse Layer.
- **Storage Layer** : This has two main layers as intermediate where staging tables are implemented and in core Datawarehouse and OLAP servers are implemented.
- **Data Consumption** : Analysis, self-service BI, Mining and reporting are done here.

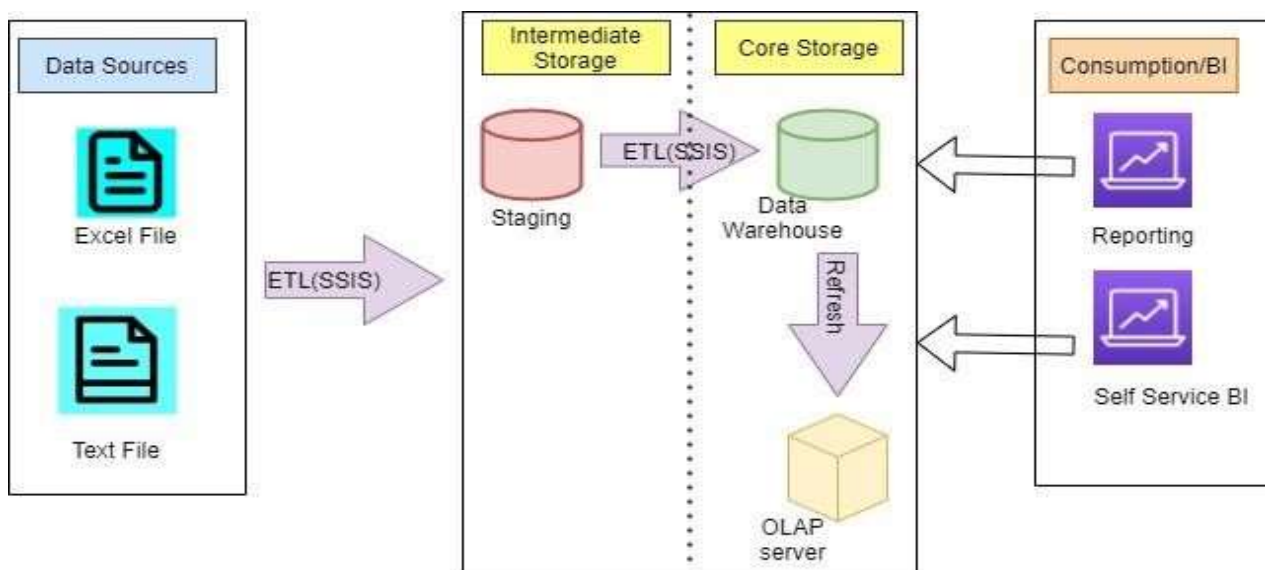


Figure 2 : High-level Architecture

4.Data Warehouse design and development

Dimensional Model

Snowflake schema was selected to design the Data Warehouse for Airline Passenger Satisfaction dataset, after considering the number of dimensional tables and fact tables. There are mainly four dimensional tables and a one fact table. All these dimensional tables are linked with the fact table.

Dimension tables and fact table:

- DimFlight
- DimFoodCategory
- DimFoodItem
- DimSatisfaction
- FactPassengers => slowly changing dimension

Hierarchies

Assumptions : no assumptions made.

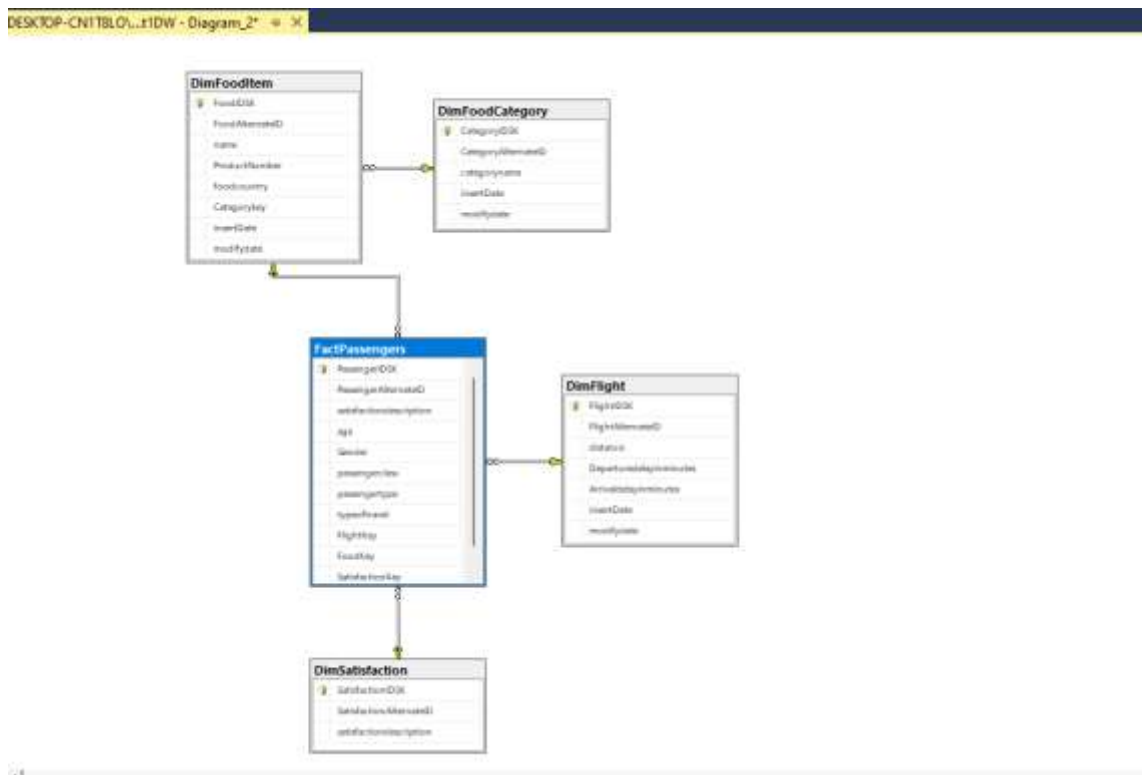


Figure 3 : Dimensional Model (Snowflake Schema)

5.ETL Development

ETL process overview

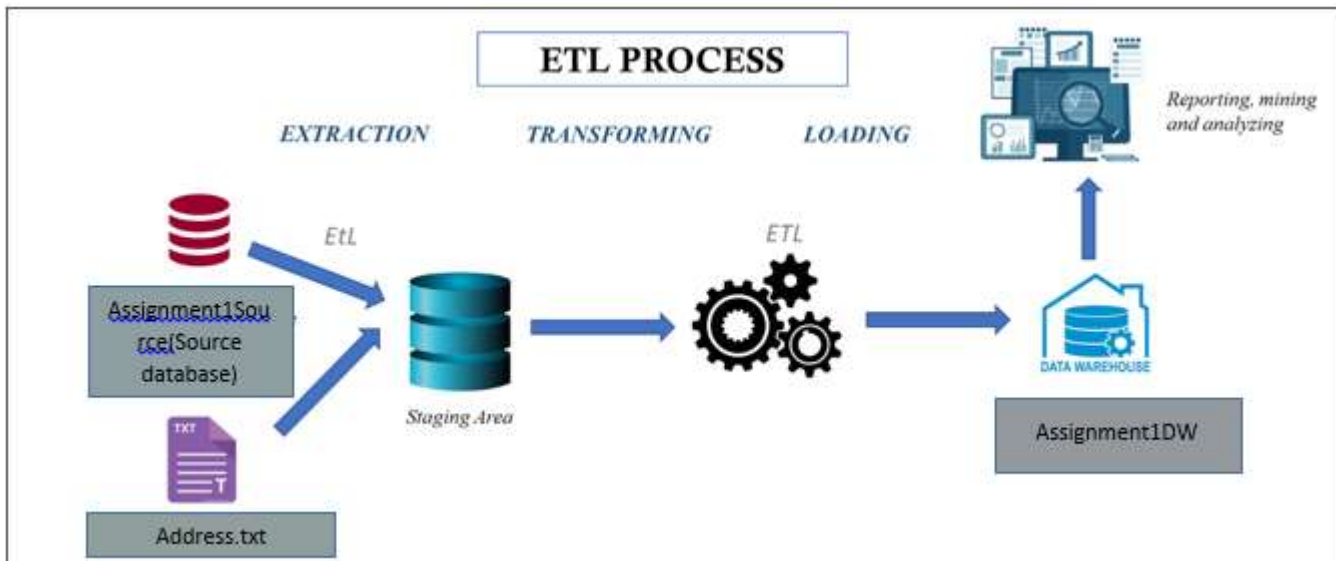


Figure 4 : ETL process

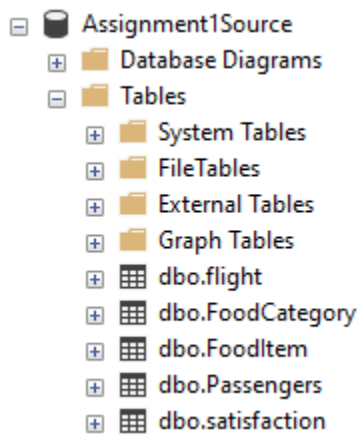
ETL development process

Steps with Screenshots will be displayed in this category.

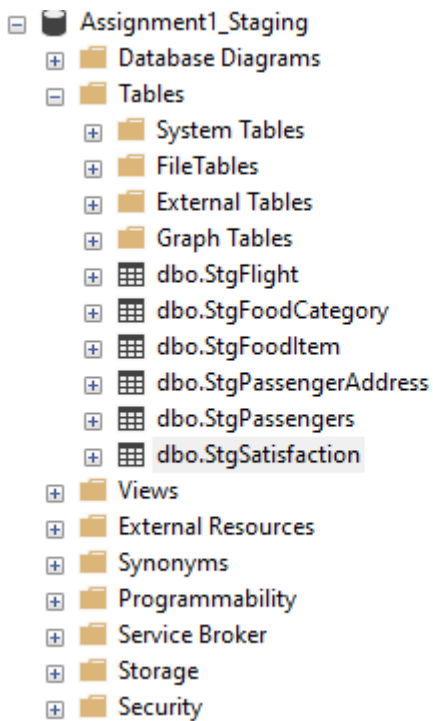
- **Step 01 : Setting up the Environment**

	accumulatd.csv	5/14/2022 3:10 PM	Microsoft Excel C...	222 KB
	address.txt	5/6/2022 11:31 AM	Text Document	565 KB
	Assignment1Source.bak	5/6/2022 9:03 PM	BAK File	4,212 KB
	flight.csv	5/6/2022 5:37 PM	Microsoft Excel C...	1 KB
	FoodCategory.csv	5/6/2022 5:40 PM	Microsoft Excel C...	1 KB
	FoodItem.csv	5/6/2022 5:40 PM	Microsoft Excel C...	1 KB
	Passengers.csv	5/14/2022 9:08 AM	Microsoft Excel C...	844 KB
	satisfaction.csv	5/14/2022 9:18 AM	Microsoft Excel C...	1 KB

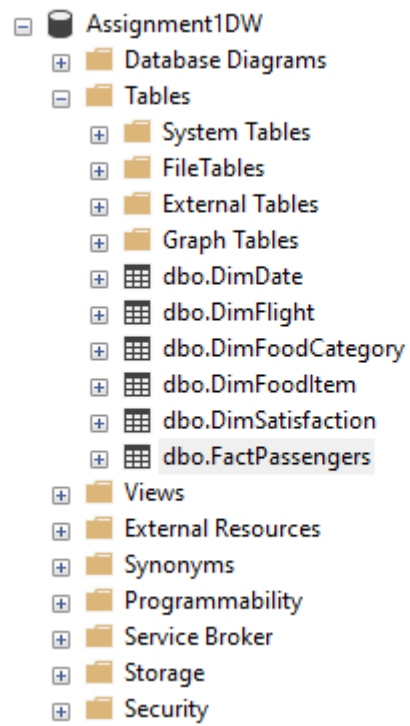
Text and Excel Files



SourceDB in SQL

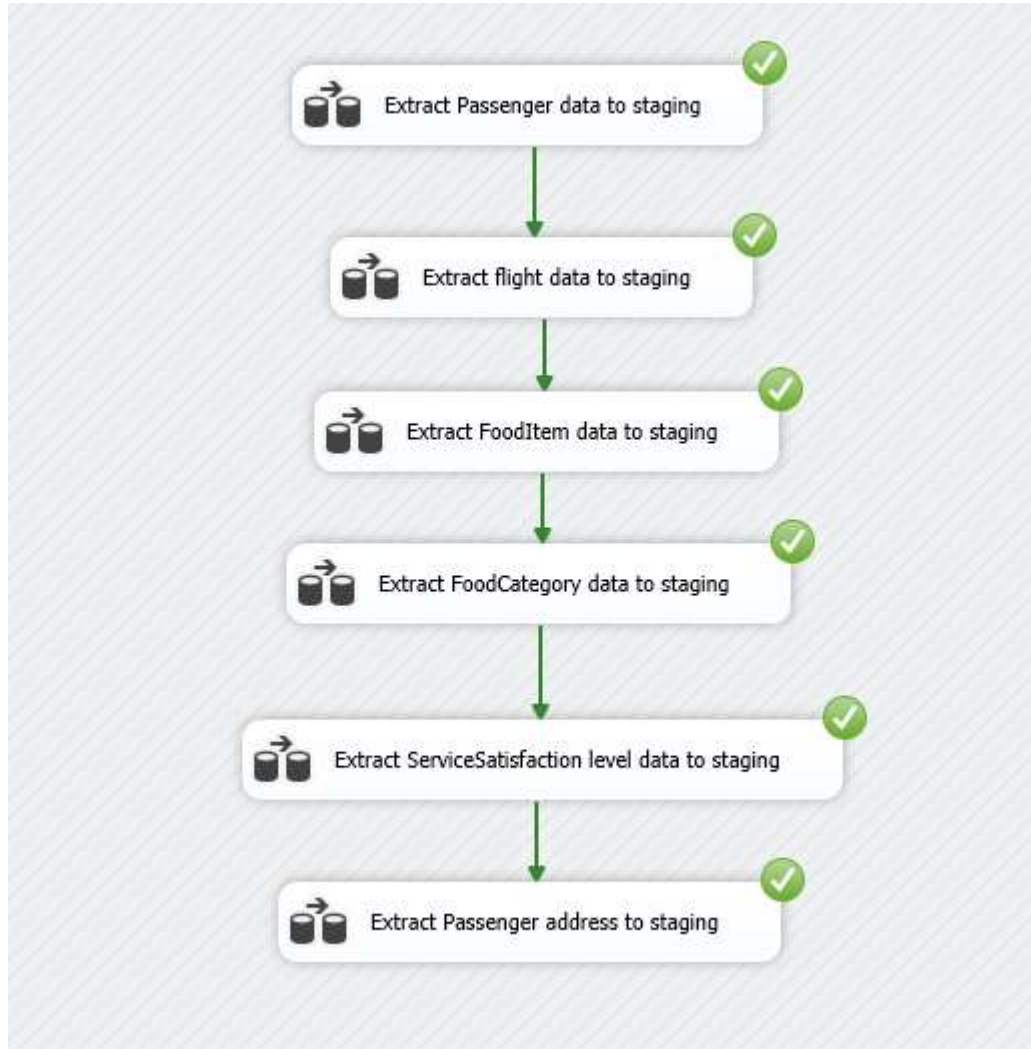


Staging

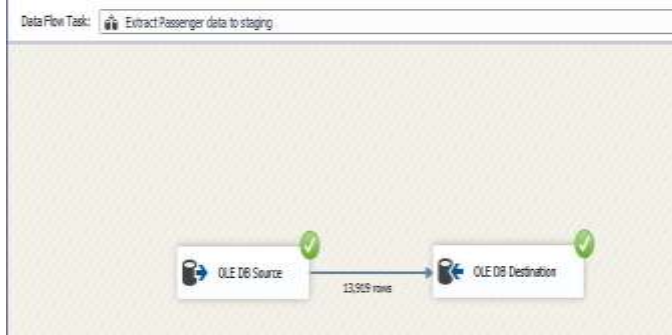


Data Warehouse

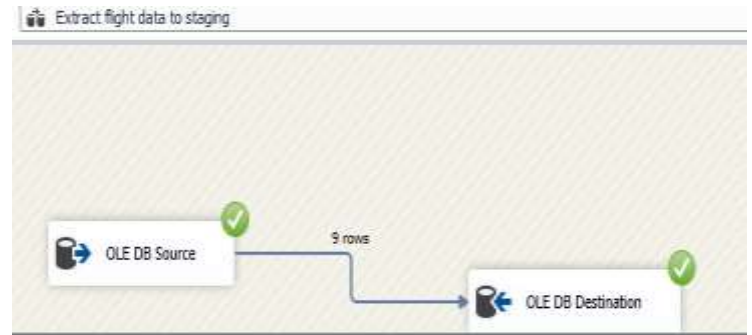
- **Step 02 : Data Extracting from source to staging tables**



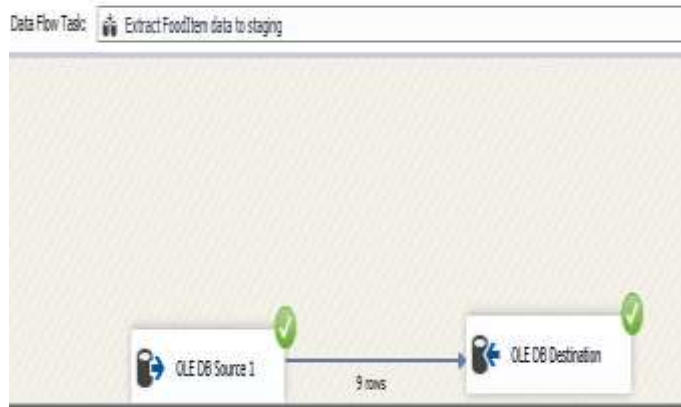
Extracting data



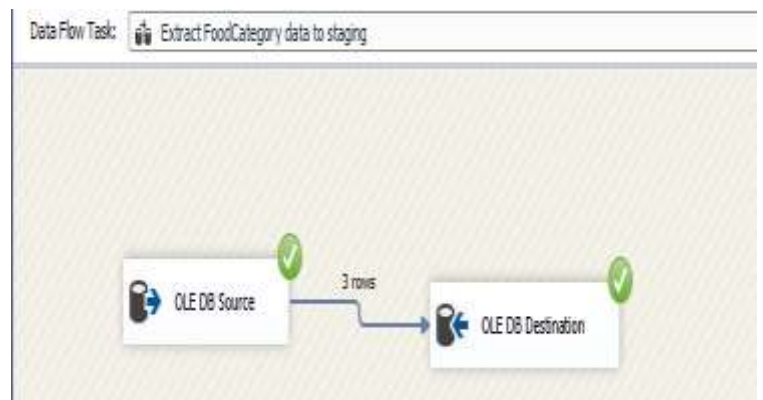
Extracting Passenger data into staging



Extracting flight data into staging



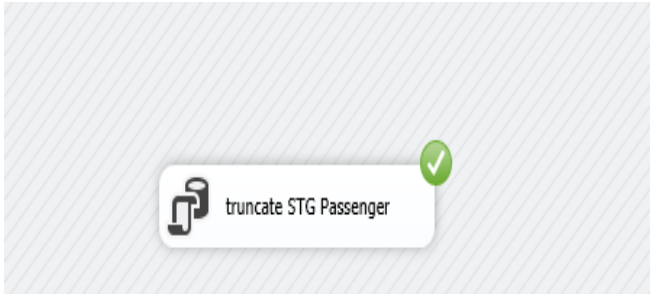
Extracting FoodItem data into staging



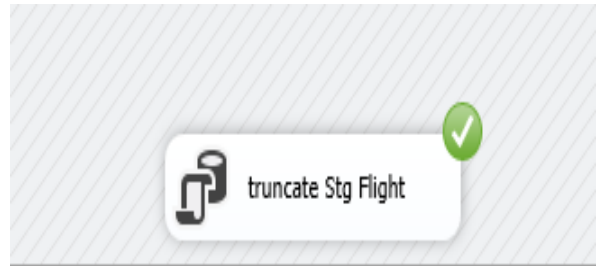
Extracting FoodCategorydata into staging



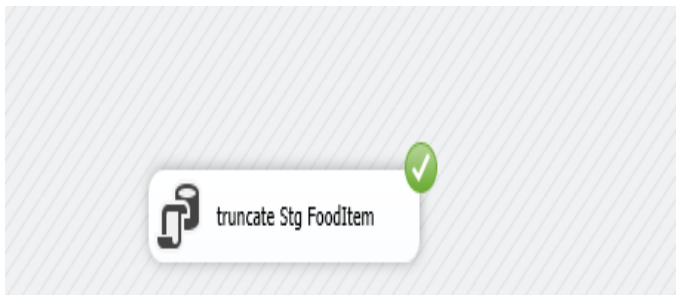
Extracting satisfaction data into staging



Truncate Passenger table



Truncate Flight table



Truncate FoodItem table

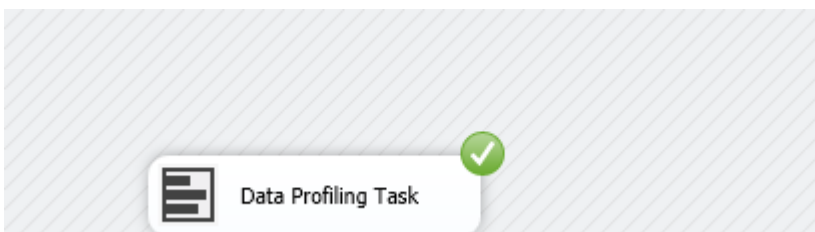


Truncate FoodCategory table

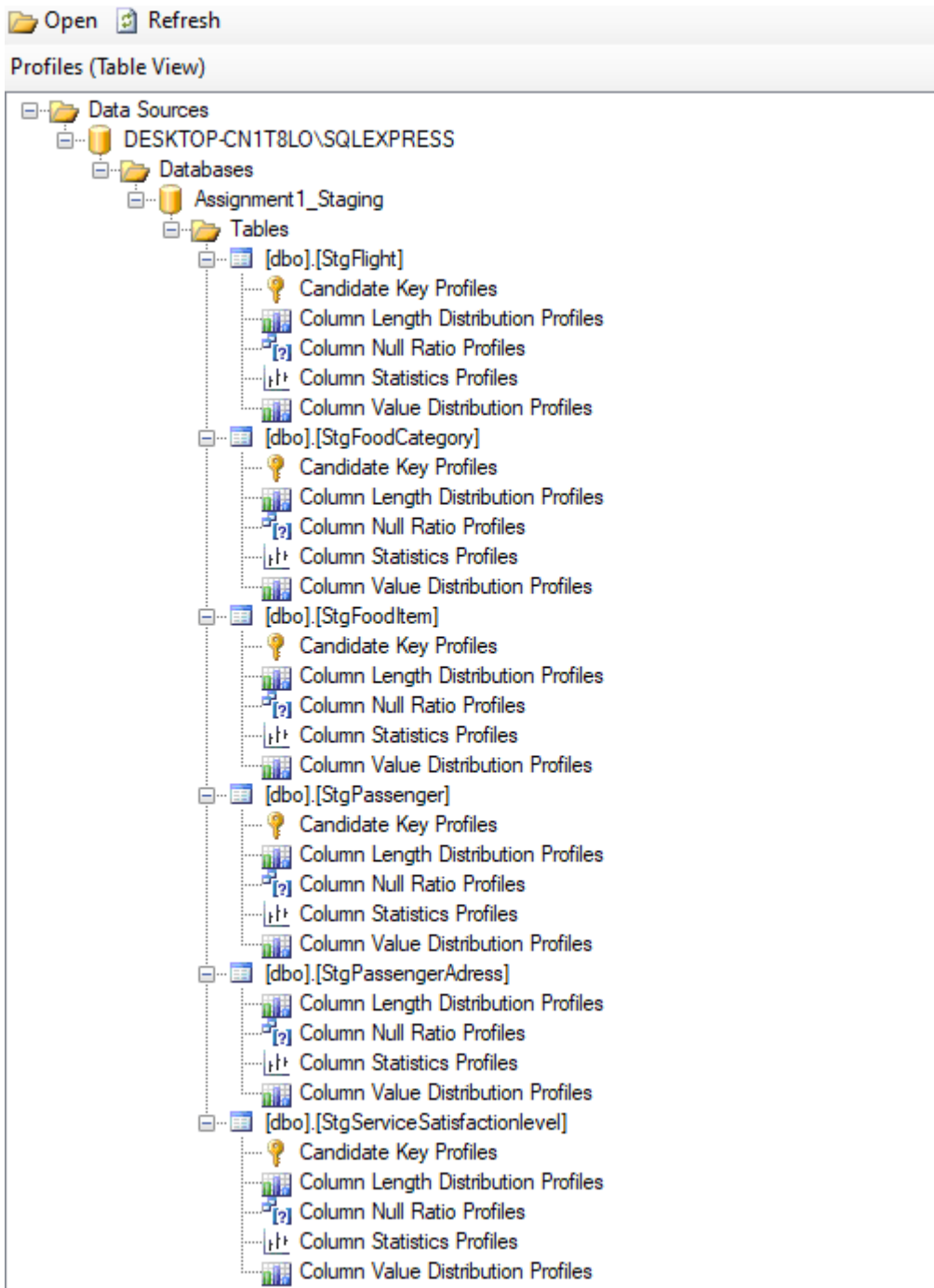


Truncate Passenger Address

- **Step 03 : Data Profiling**



Data Profiling task



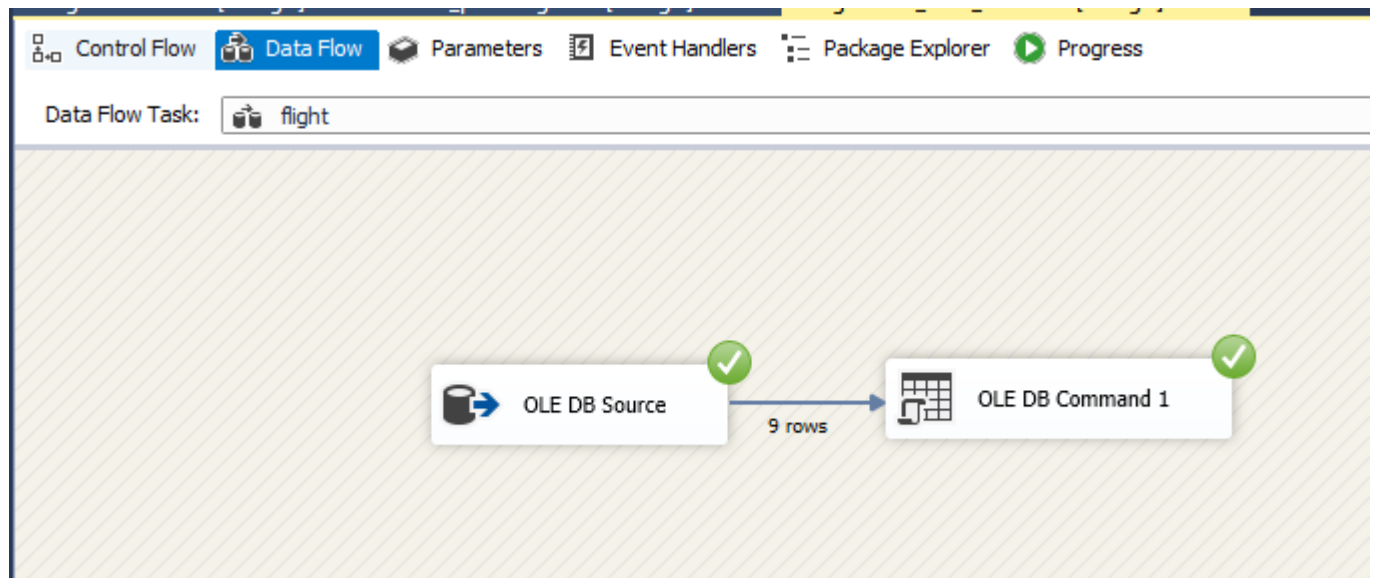
Data profile viewer

- **Step 04 : Data Transformation from staging to warehouse**

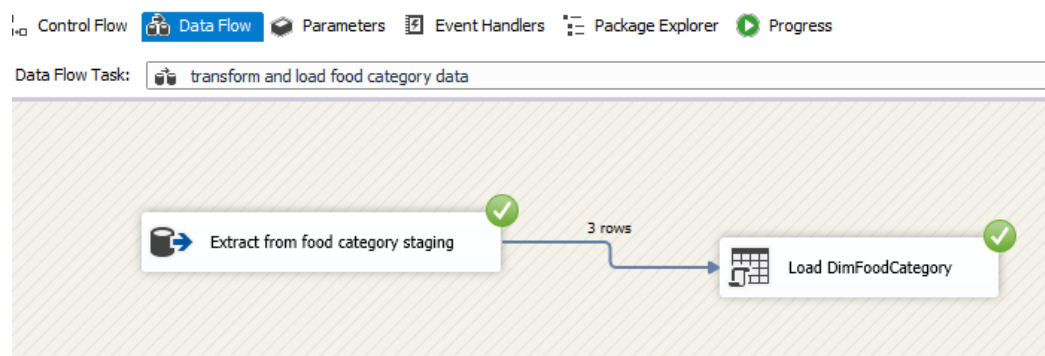
Transforming and loading

Control Flow Data Flow Parameters Event Handlers Package Explorer


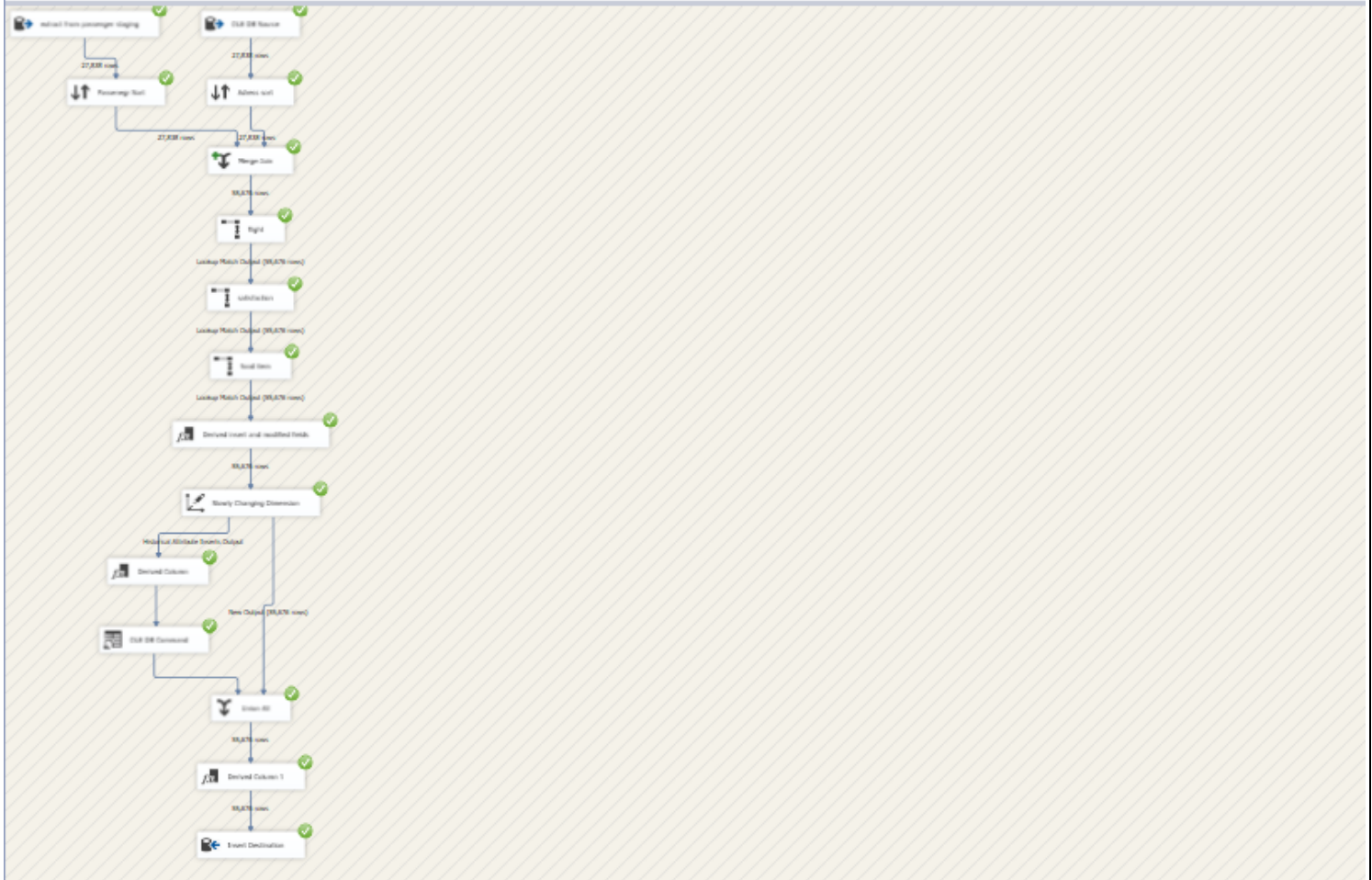




Transforming, loading flights



Transforming, loading Food Category

Data Flow Task:  transform and load passenger fact table

Transforming, loading fact table with a slowly changing dimension

Some of the Stored Procedures,

```
CREATE PROCEDURE [dbo].[UpdateDimFlight]
@FlightID int,
@distance varchar(50),
@departuredelayinminutes int,
@arrivaldelayinminutes int

AS
BEGIN
if not exists (select FlightIDSK
from dbo.DimFlight
where FlightAlternateID = @FlightID)
BEGIN
insert into dbo.DimFlight
(FlightAlternateID,distance , departuredelayinminutes ,arrivaldelayinminutes, InsertDate, ModifyDate)
values
(@FlightID,@distance,@departuredelayinminutes,@arrivaldelayinminutes, GETDATE(), GETDATE())
END;
if exists (select FlightIDSK
from dbo.DimFlight
where FlightAlternateID = @FlightID)
BEGIN
update dbo.DimFlight
set
distance=@distance,
departuredelayinminutes=@departuredelayinminutes,
arrivaldelayinminutes=@arrivaldelayinminutes,
ModifyDate = GETDATE()
where FlightAlternateID = @FlightID
END;
END;
```

Stored procedure for flight

```

CREATE PROCEDURE [dbo].[UpdateDimFoodCategory]
@CategoryID int,
@categoryname varchar(50)

AS BEGIN
if not exists (select CategoryIDSK
from dbo.DimFoodCategory
where CategoryAlternateID = @CategoryID) BEGIN
insert into dbo.DimFoodCategory
(CategoryAlternateID,categoryname, insertDate, modifydate)
values
(@CategoryID, @categoryname, GETDATE(), GETDATE())
END;
if exists (select CategoryIDSK
from dbo.DimFoodCategory
where CategoryAlternateID= @CategoryID) BEGIN
update dbo.DimFoodCategory
set categoryname = @categoryname,
modifydate = GETDATE()
where CategoryAlternateID = @CategoryID END;
END;

```

Stored procedure for Food Category

```

CREATE PROCEDURE [dbo].[UpdateDimFoodItem]
@FoodID int,
@Categorykey int,
@Name varchar(50),
@ProductNumber varchar(50),
@foodcountry varchar(50)
AS
BEGIN
if not exists (select FoodIDSK
from dbo.DimFoodItem
where FoodAlternateID = @FoodID) BEGIN
insert into dbo.DimFoodItem
(FoodAlternateID,CategoryKey,Name, ProductNumber,foodcountry, InsertDate, Modifydate)
values
(@FoodID, @CategoryKey, @Name, @ProductNumber,@foodcountry, GETDATE(), GETDATE())
END;
if exists (select FoodIDSK
from dbo.DimFoodItem
where FoodAlternateID = @FoodID) BEGIN
update dbo.DimFoodItem
set CategoryKey = @CategoryKey, Name = @Name, ProductNumber=@ProductNumber, foodcountry=@foodcountry,
Modifydate = GETDATE()
where FoodAlternateID = @FoodID END;
END

```

Stored Procedure for FoodItem

Some of the tables with values

SQLQuery7.sql - D:\nment1DW (sa (62)) * X SQLQuery8.sql - D:\nment1DW (sa (66)) * SQLQuery3.sql - D:\nment1DW (sa (66)) * SQLQuery4.sql - D:\nment1DW (sa (65)) *

```

SELECT *
FROM [Assignment1DW].[dbo].[FactPassengers]
  
```

100 %

Results Messages

	PassengerDSK	PassengerAlternateID	age	Gender	passengerclass	passengerstype	typeoftravel	FlightKey	FoodKey	SatisfactionKey	passengerAddress	startdate
1	1	500	24	Female	Business	Loyal Customer	Business travel	1	2	5	20 Opel St	2022-05-14 15:34:42.000
2	2	501	43	Male	Business	Loyal Customer	Business travel	1	2	5	15 Osborne Gr	2022-05-14 15:34:42.000
3	3	502	7	Male	Business	Loyal Customer	Personal Travel	1	2	5	4 Tyler St	2022-05-14 15:34:42.000
4	4	503	65	Female	Business	Loyal Customer	Personal Travel	1	2	5	3 Botha Av	2022-05-14 15:34:42.000
5	5	504	8	Female	Eco	Loyal Customer	Personal Travel	1	2	5	2 Colthur St	2022-05-14 15:34:42.000
6	6	505	50	Male	Eco	Loyal Customer	Personal Travel	1	2	5	43 Dawson St	2022-05-14 15:34:42.000
7	7	506	25	Male	Eco Plus	Loyal Customer	Personal Travel	1	2	5	1/60 Dundee St	2022-05-14 15:34:42.000
8	8	507	24	Female	Eco	Loyal Customer	Personal Travel	1	2	5	1/23 Glasgow Av	2022-05-14 15:34:42.000
9	9	508	55	Female	Eco	Loyal Customer	Personal Travel	1	2	5	2/42 Hebbury St	2022-05-14 15:34:42.000
10	10	509	21	Female	Eco	Disloyal Customer	Business travel	1	2	5	802 High St	2022-05-14 15:34:42.000

Query executed successfully. DESKTOP-CN1TRLO\SQLEXPRESS... sa (62) Assignment1DW 00:00:00 13,919 rows

Fact table with values

SQLQuery3.sql - D:\nment1DW (sa (66)) * X SQLQuery4.sql - D:\nment1DW (sa (65)) * SQLQuery3.sql - D:\nment1DW (sa (58)) * SQLQuery2.sql - D:\nment1DW (sa (54)) *

```

/***** Script for SelectTopNRows command from SSPS *****/
SELECT *
FROM [Assignment1DW].[dbo].[DimFoodItems]
  
```

100 %

Results Messages

	FoodDSK	FoodAlternateID	name	ProductNumber	foodcountry	Categorykey	insertDate	modifydate
1	1	1	sa rice with vegetables	AR-5361	Thailand	1	2022-05-10 21:59:27.310	2022-05-13 20:27:39.653
2	2	3	steamed rice served with salad	DR-5378	Japanese	1	2022-05-10 21:59:27.320	2022-05-13 20:27:39.653
3	3	5	St Lankan fish curry	SDE-5309	St Lankan	3	2022-05-10 21:59:27.320	2022-05-13 20:27:39.653
4	4	6	Gluten Free Meal	AR-484	St Lankan	2	2022-05-10 21:59:27.320	2022-05-13 20:27:39.653
5	5	21	yfroe-spice chicken curry with pasta salad	DR-5385	China	1	2022-05-10 21:59:27.320	2022-05-13 20:27:39.653
6	6	32	Pan seared salmon steak	SDE-5386	Thailand	3	2022-05-10 21:59:27.320	2022-05-13 20:27:39.653
7	7	40	Low Fat Meal	DR-387	China	2	2022-05-10 21:59:27.320	2022-05-13 20:27:39.653
8	8	45	St Lankan special	SDE-5388	St Lankan	3	2022-05-10 21:59:27.320	2022-05-13 20:27:39.657
9	9	66	Kosher meal	AR-5389	Thailand	2	2022-05-10 21:59:27.320	2022-05-13 20:27:39.657

Query executed successfully. DESKTOP-CN1TRLO\SQLEXPRESS... sa (66) Assignment1DW 00:00:00 9 rows

DimFoodItems with values

SQLQuery4.sql - D:\nment1DW (sa (63)) * SQLQuery3.sql - D:\nment1DW (sa (58)) * SQLQuery2.sql - D:\nment1DW (sa (54)) * SQLQuery1.sql - D:\nment1DW (sa (52))

```

/***** Script for selectTopNRows command from SSRS *****/
SELECT
FROM [Assignment1DW].[dbo].[DimFlight]

```

100 %

Results Messages

	FlightIDSK	RightAlternateID	distance	Departuredelayinminutes	Arrivaldelayinminutes	insertDate	modifydate
1	1	123	460	25	18	2022-05-13 20:02:34.158	2022-05-13 20:27:39.747
2	2	124	235	1	6	2022-05-13 20:02:34.167	2022-05-13 20:27:39.747
3	3	135	1142	0	0	2022-05-13 20:02:34.167	2022-05-13 20:27:39.747
4	4	201	562	11	9	2022-05-13 20:02:34.170	2022-05-13 20:27:39.747
5	5	233	214	0	0	2022-05-13 20:02:34.170	2022-05-13 20:27:39.747
6	6	245	1180	0	0	2022-05-13 20:02:34.170	2022-05-13 20:27:39.747
7	7	345	1275	9	23	2022-05-13 20:02:34.173	2022-05-13 20:27:39.747
8	8	380	2035	4	0	2022-05-13 20:02:34.173	2022-05-13 20:27:39.747
9	9	399	853	0	0	2022-05-13 20:02:34.177	2022-05-13 20:27:39.747

Query executed successfully. DESKTOP-CN1T8LO\SQLEXPRESS - sa (63) Assignment1DW 00:00:00 9 rows

DimFlight with values

6.References

- Kaggle : <https://www.kaggle.com/datasets>