**PROJECT INTRODUCTION:**

The main focus of this project:

* To design a data warehouse model for the business model described in this project.
* Identify available open source database to use for the project.
* Create a ETL plan to extract, load and transform this data from the database into data warehouse schema.
* Finally load this sample data into the data warehouse schema and create some visualization dashboard to represent analytics.

**BUSINESS BACKGROUND:**

The main business model I am targeting via this project is the manufacturing. Primarily in this manufacturing business the multiple entities of the business are tied in and interlinked i.e. starting from employee records to purchase orders and online transactions are very important for the success of the business.

Hence this project aims at designing a data warehouse model with the available database of all online transactions. Down the lane this will be used for various analytics activities to significantly improve any business related issues as well as understand market trend worldwide and how the business and transactions fluctuate as a function different entities.

**DATABASE DESCRIPTION:**

In this project I plan to use the online open source AdventureWorks Database available from Microsoft.

Database: <https://github.com/Microsoft/sql-server-samples/releases/tag/adventureworks>

Regarding the Database - all the data generated and stored is in regards to Adventureworks bicycle manufacturing company. Database consists of many tables (<https://technet.microsoft.com/en-us/library/ms124438(v=sql.100).aspx>) relating to Customer vendor and employee details, Product specifics, transactions and ordering specifics, billings details etc.

**ANALYTICAL THEME:**

The main analytical theme of this project is as follows:

* For the defined business model and the specific problem I am addressing - I would like to define the specifics needed in terms of data from the available databases.
* This specific information set will be used to design a data warehouse schema by defining dimensions and fact table.
* Later in Mysql I would like to extract, load and transform the data from the online data sources and make a sample data warehouse schema.
* This sample schema will be used to develop some visual analytics and hence will be presented in the final presentation and report.

**DATABASE FINALIZE:**

In this project I plan to use the online open source Adventure Works Database available from Microsoft.

Database: <https://github.com/Microsoft/sql-server-samples/releases/tag/adventureworks>

Regarding the Database - all the data generated and stored is in regards to Adventure works bicycle manufacturing company. Database consists of many tables (<https://technet.microsoft.com/en-us/library/ms124438(v=sql.100).aspx>) relating to Customer vendor and employee details, Product specifics, transactions and ordering specifics, billings details etc.

**INFORMATION NEEDS:**

* Needed data from the database would be of Customer related tables, Product tables, Sales Tables, Purchase, vendor, and store tables.
* This information will be extracted, filtered and loaded into schema and used for visualization.
* The research and analytics process down the lane can use the data to understand business trends and possibly correlate different factors contributing towards sales.

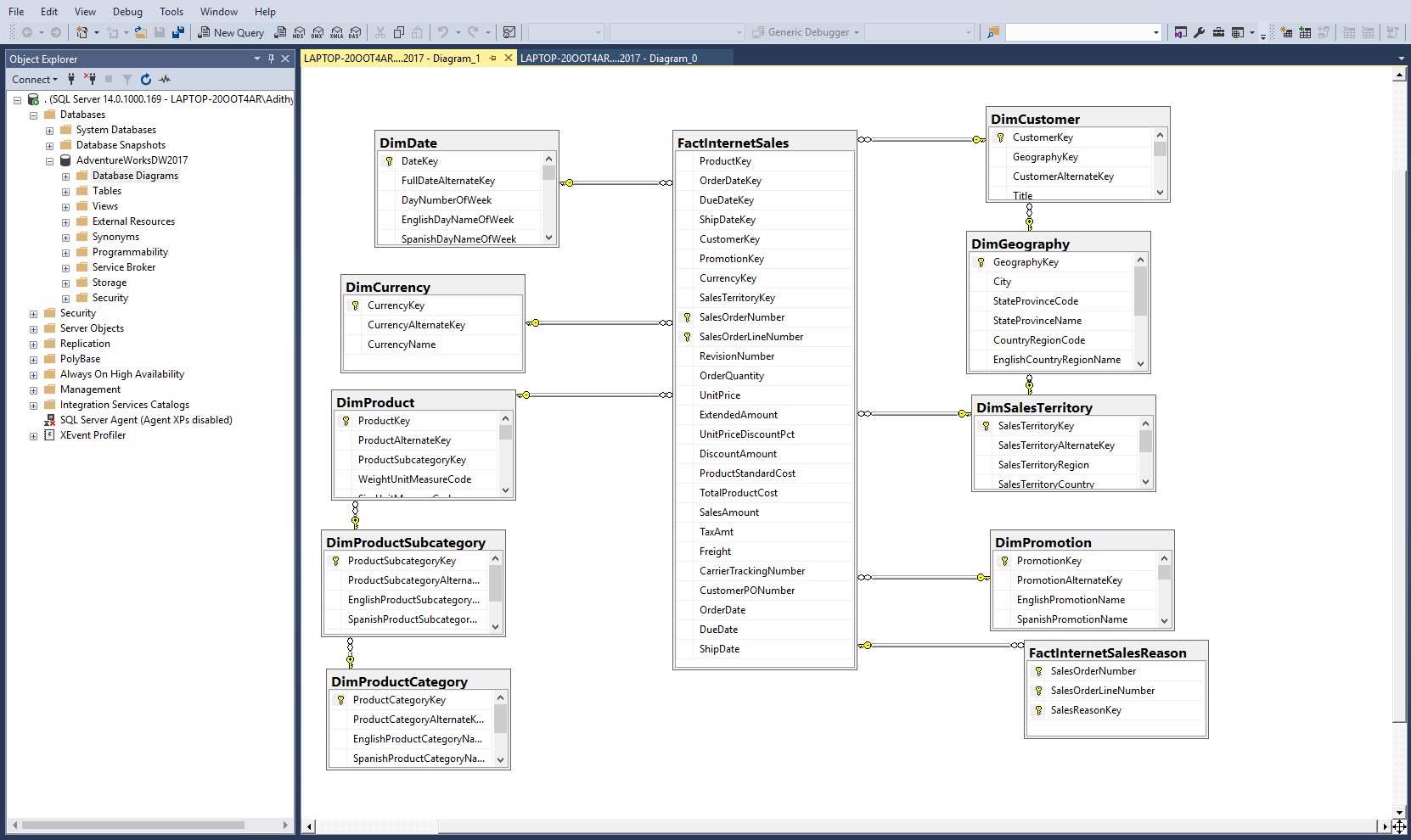
**DIMENSIONAL MODELLING**

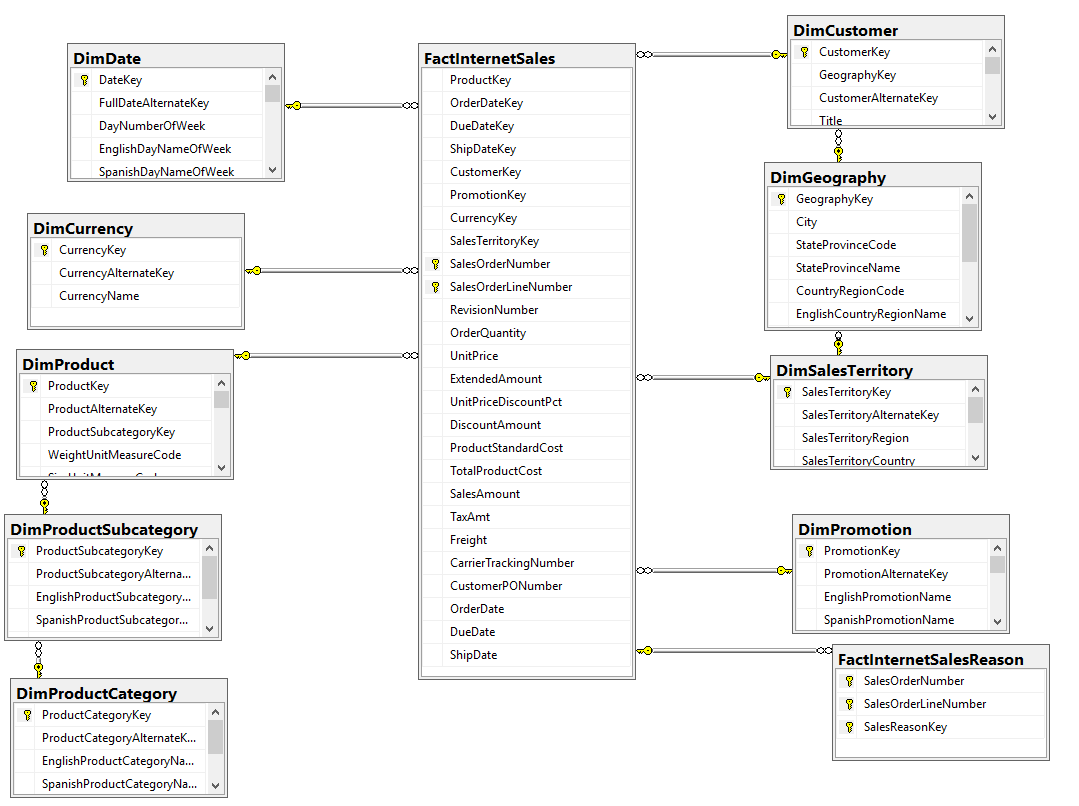
For this Project I chose to work with the AdventureWorks Database and the main theme of this work is to:

* Provide great level of detail and visual insight into the Database to analyze the internet Sales globally considering temporal, spatial and product level details from the database.
* I choose the necessary dimensions and facts from the Database and a start schema representation of the Different dimensions and Facts are presented in this work.
* Dimensional Modelling is all performed in SQL Server Management Studio 2017.

Dimensions and Facts chosen from the AdventureWorks 2017 Database for our Data Warehousing Project proposal:

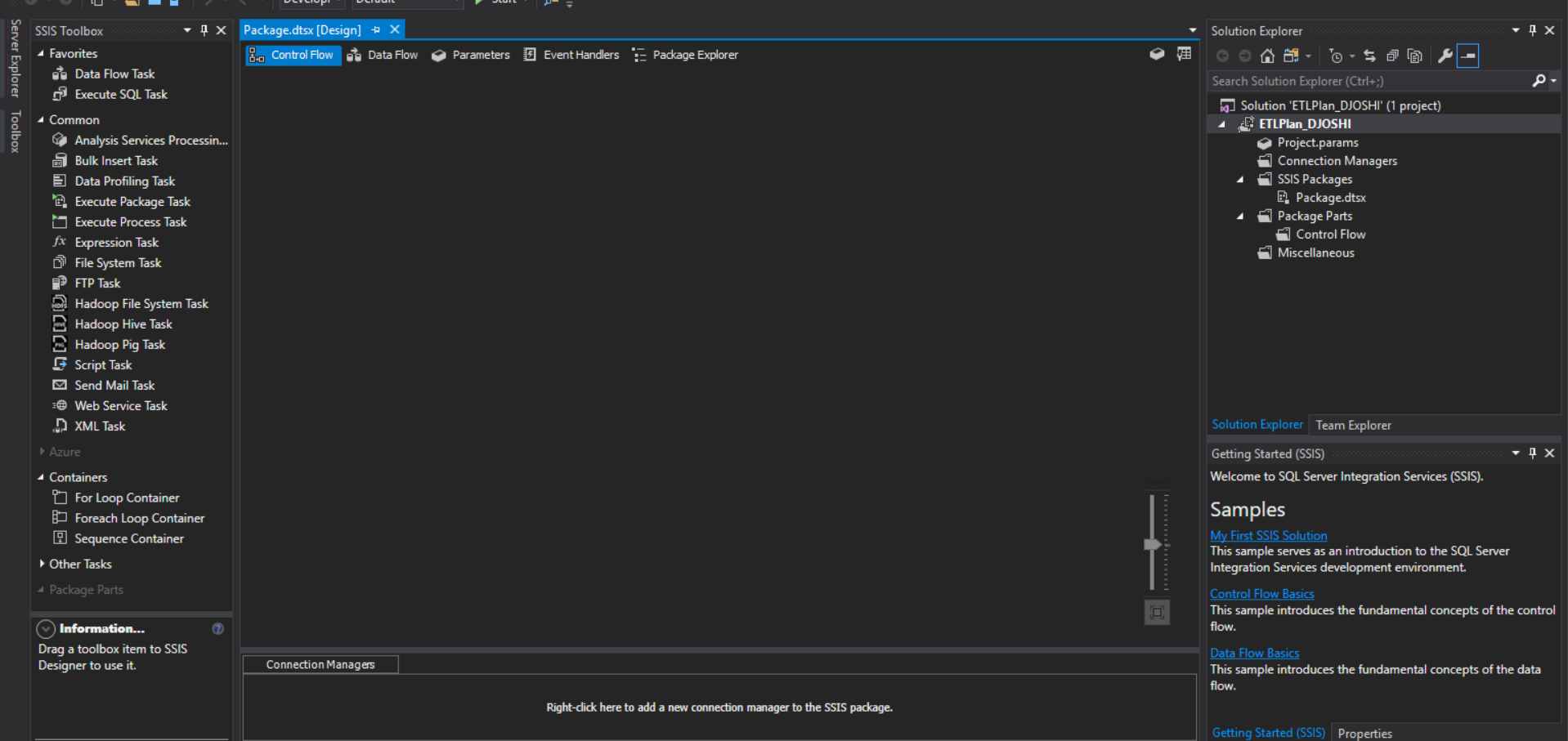
* FactInternetSales
* FactInternetSalesReason
* DimCustomer
* DimGeography
* DimSalesTerritory
* DimProductCategory
* DimProductSubCategory
* DimProduct
* DimCurrency
* DimDate
* DimPromotion



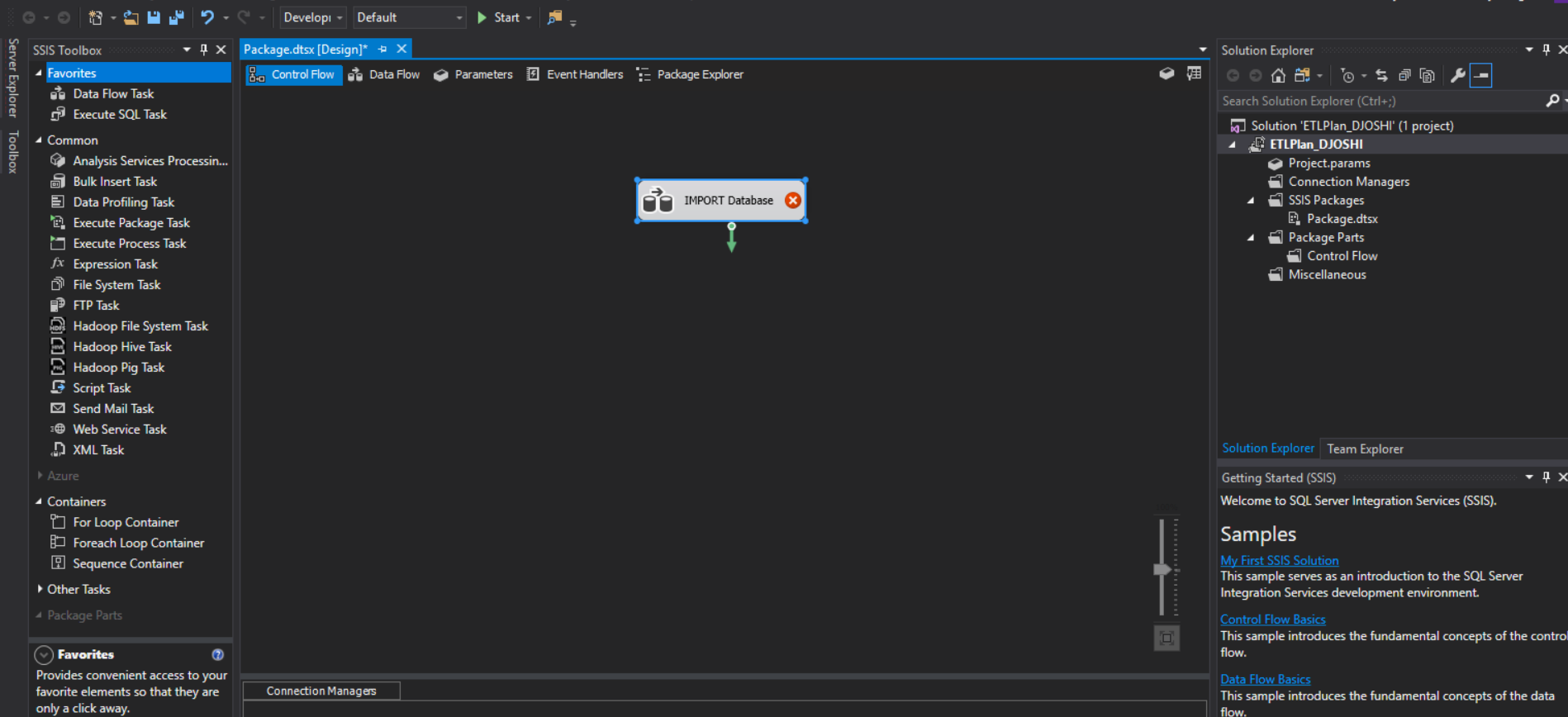


**PROJECT – ETL PLAN:**

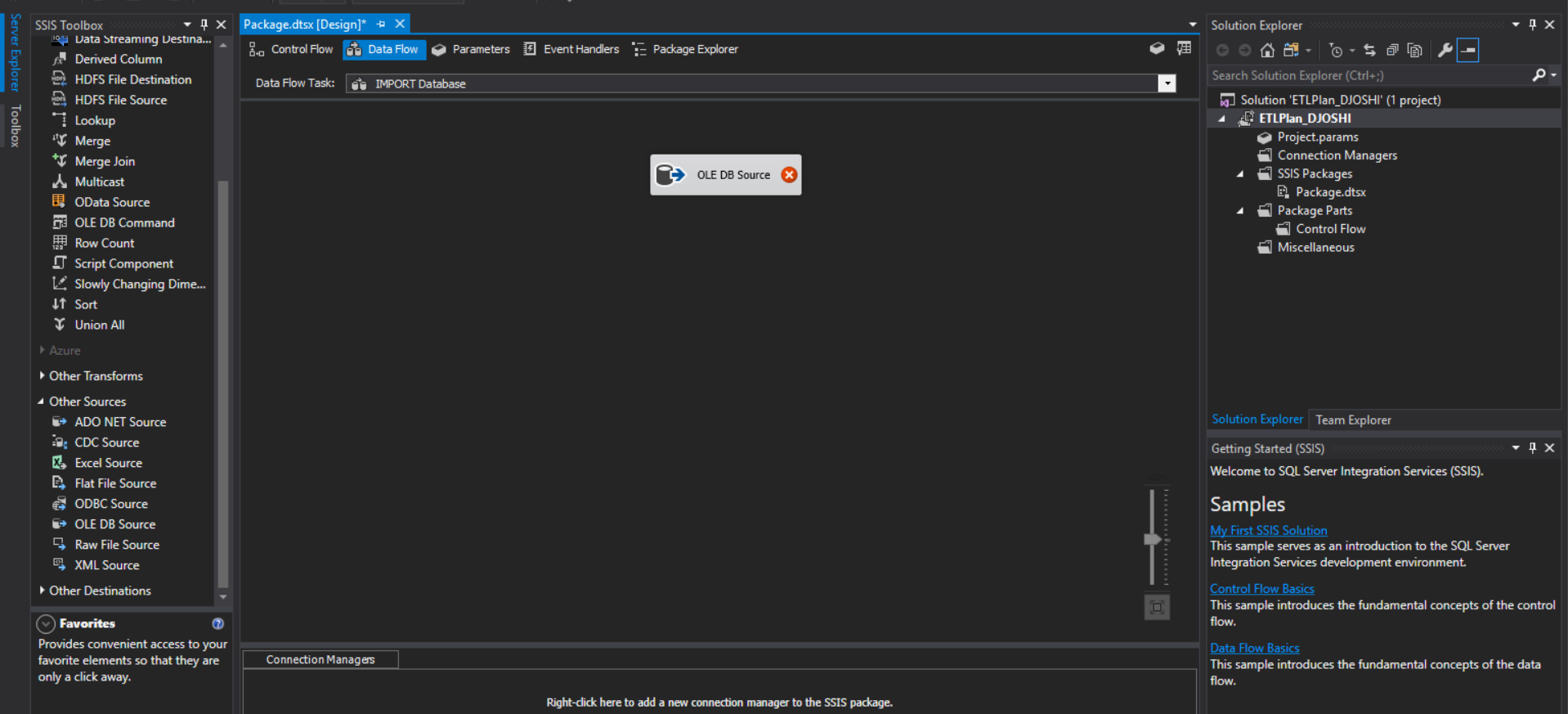
* This part of the project is mainly about creating a ETL plan to extract and load data from the source database into the data warehouse
* The data from the source is imported and final pushes are made to the data warehouse.
* First an Integration Services Project is initiated.



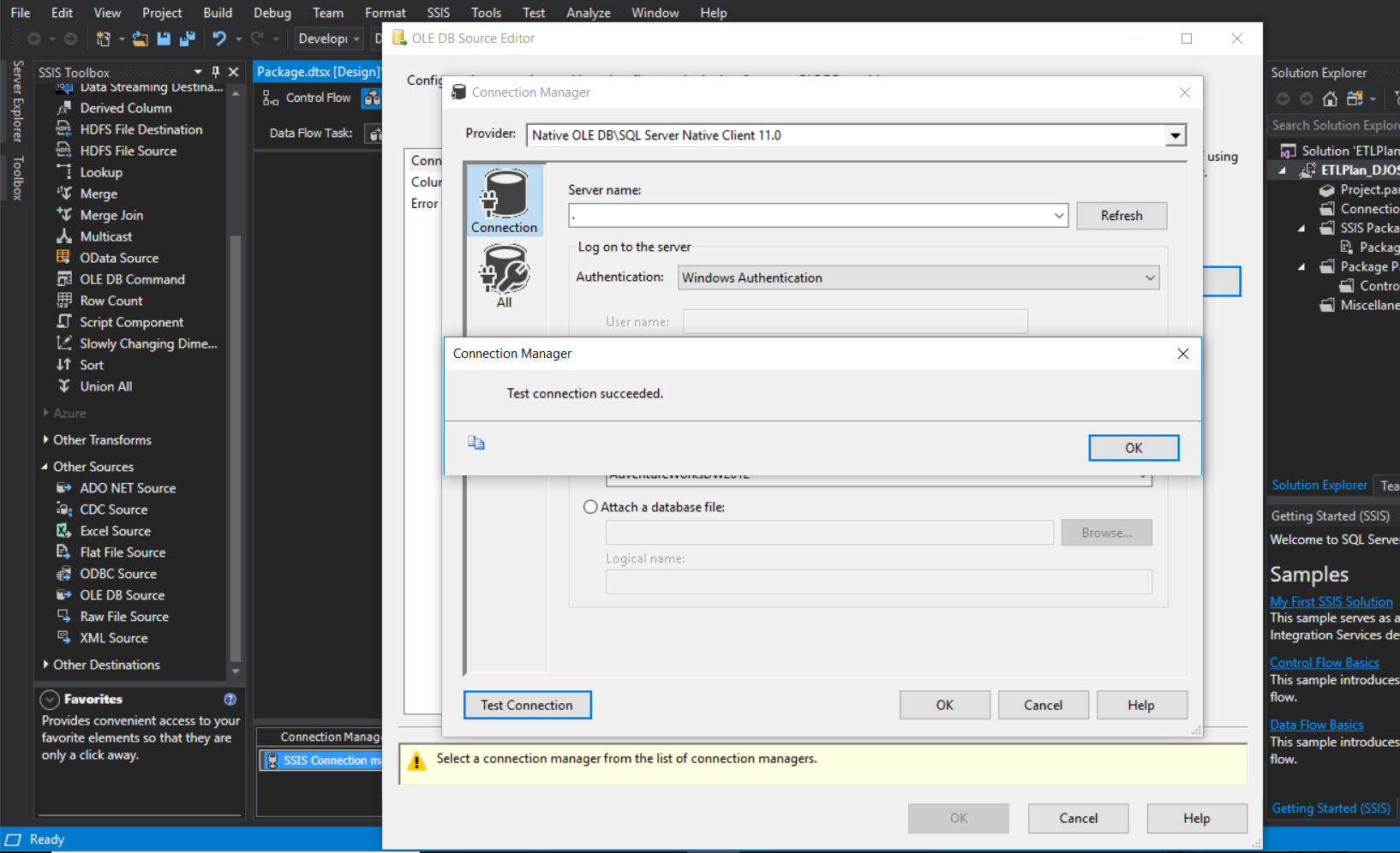
Control flow tasks are created:



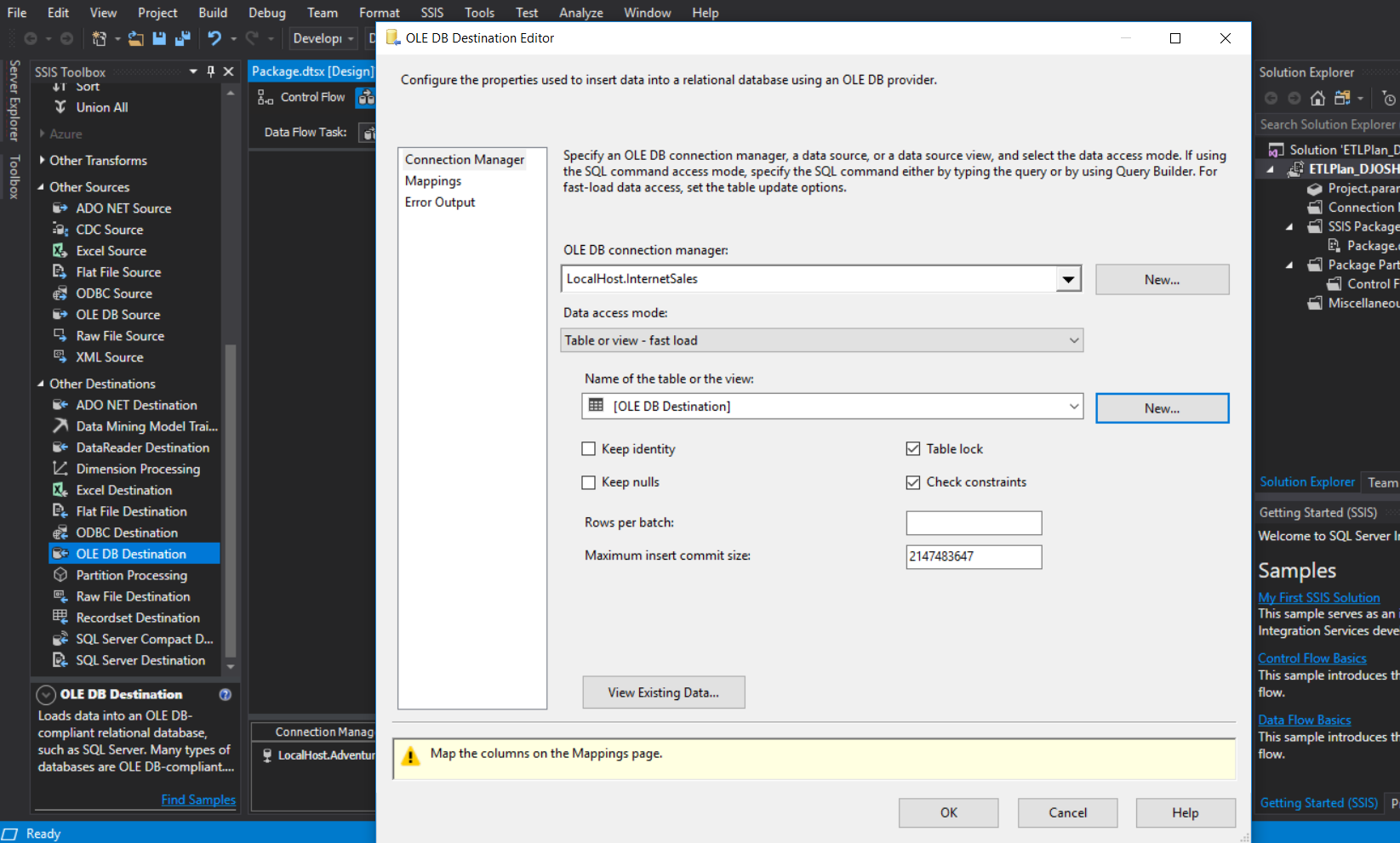
* Creating data flow:



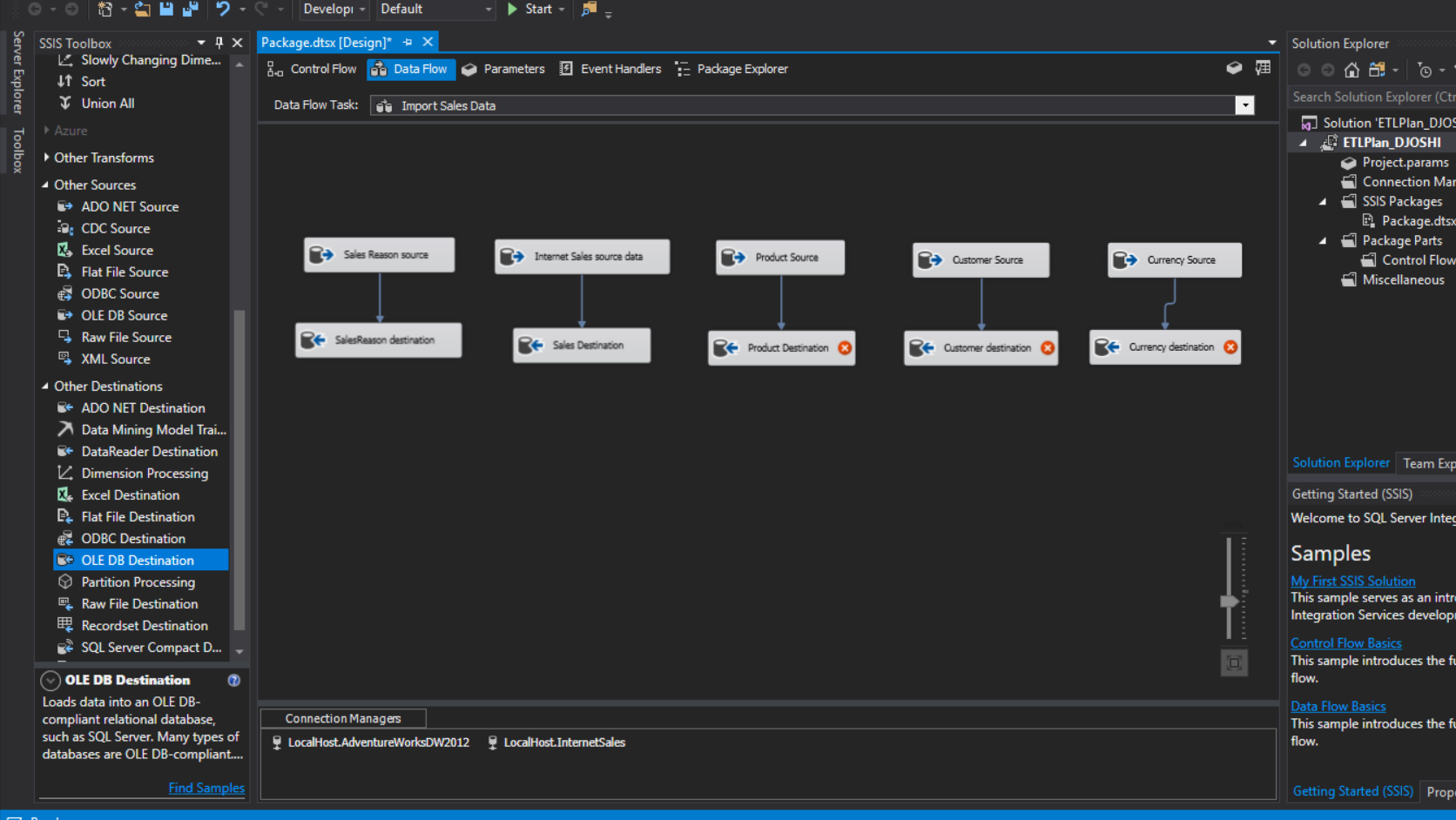
* Connection to the local host and the database source from Adventure works.



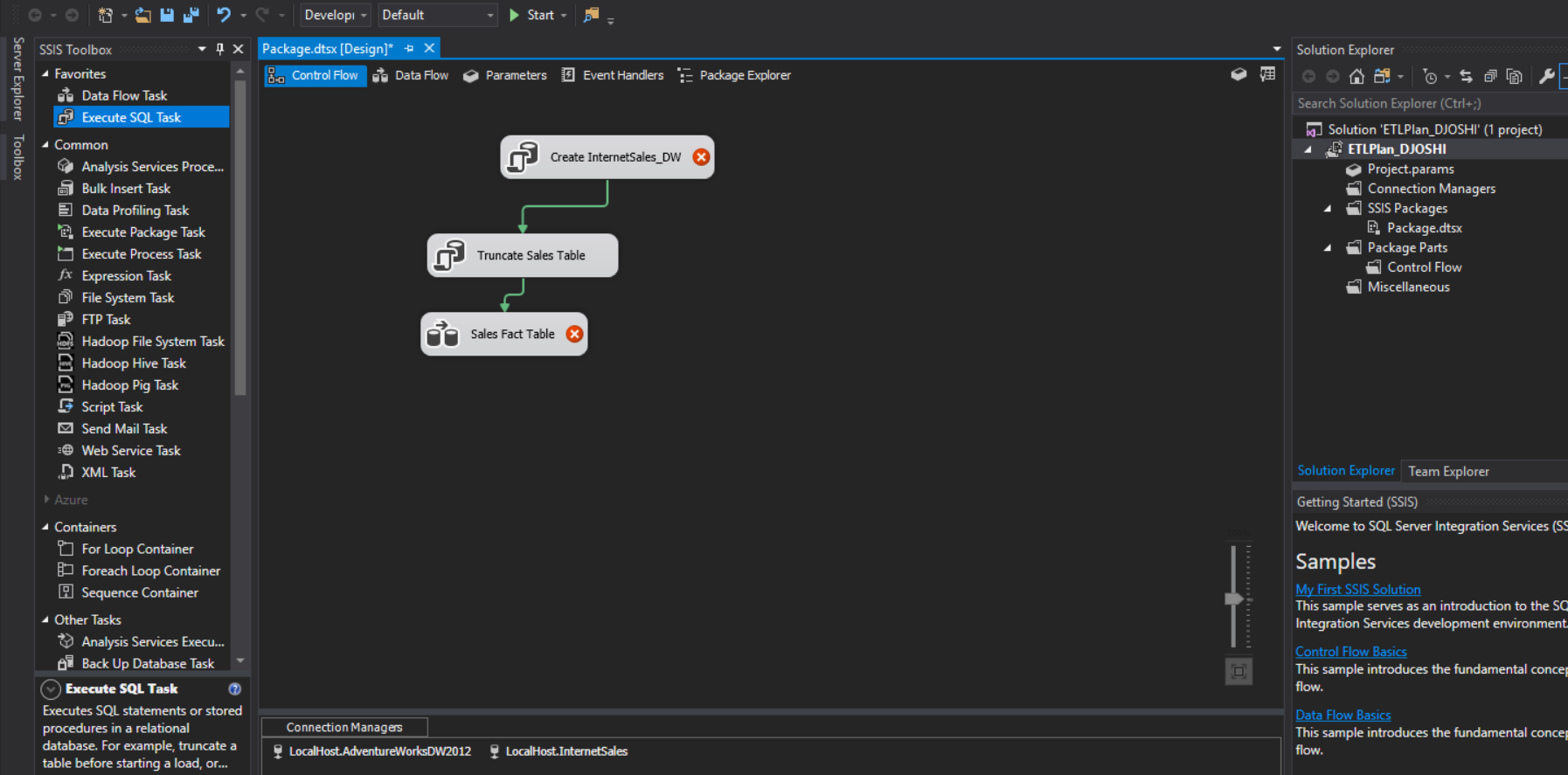
* Destination table and the SQL query along with the table name updated for Internet Sales Data



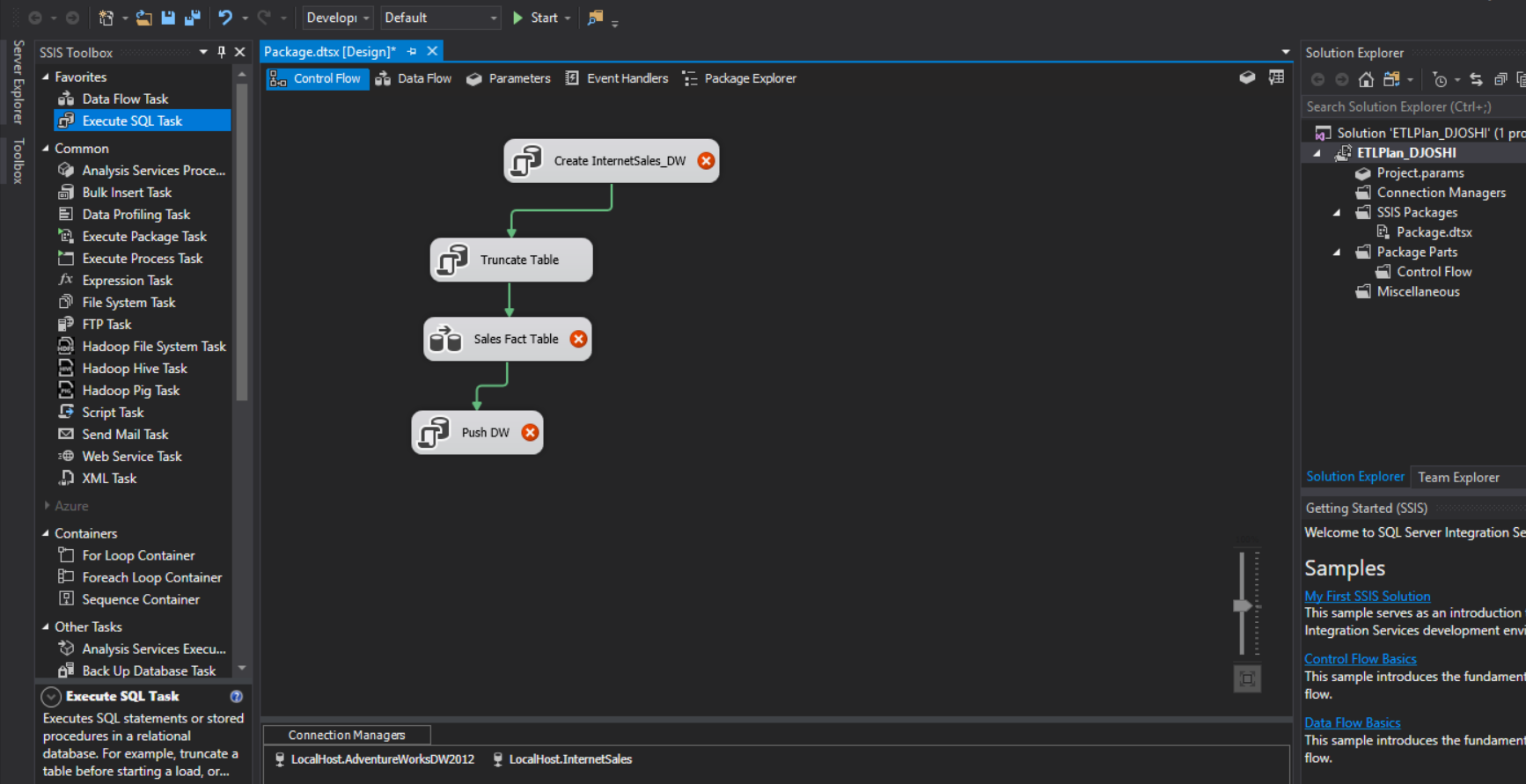
Other Dimension Tables created with the flow and appropriate truncation:



* Execute SQL queries added to control flow/



* Execute SQL query 🡪 create database if it doesn’t exist query added.



Execute SQL query with push all the updates and changes to the data warehouse.

* The process was all executed using the SSIS tools with visual studio 2017.

Conclusion:

* All the objectives of the data warehouse modelling project were met.
* Successfully implemented ETL plan by extracting the required tables from the database and executing transform and loading steps to the new data warehouse
* The hence transformed data can be easily visualized for further analytics in any database visualization tools or even Excel integrated with SQL.