**Data Warehousing & Business Intelligence**

**Week-2 | Assignment 2 Documentation**

**Vedant Desai (001830150)**

Creation Date: 01/24/2019 Technical Design Document

**Table of Contents**

|  |  |
| --- | --- |
| 1. | Document Control |
| 1.1 | Version History and Approvers |
| 2. | Introductions |
| 2.1 | Assignment Overview |
| 3. | Architecture Overview |
| 3.1 | Databases |
| 3.2 | Extract Transform and Load (ETL) |
| 4. | Monitoring and Alerting |

1. DOCUMENT CONTROL
   1. Version History and Approvers

|  |  |  |  |
| --- | --- | --- | --- |
| Version No. | Date | Author | Approvers |
| 1 | 01/24/2019 | Vedant | Ayush, Professor Vincent |

1. INTRODUCTIONS
   1. Assignment Overview

This assignment requires us to perform initial ETL by exporting data from existing table to CSV. Secondly, we need to create a staging table of original table and load that CSV file again in that staging table. For this we need to use SSIS packages and automate the flow. y

And Lastly, we need to upload it to GITHUB repository.

1. ARCHITECTURE OVERVIEW

For initial export of data from database to CSV we need to work on AdventureWorks database and HumanResources.Employee table.

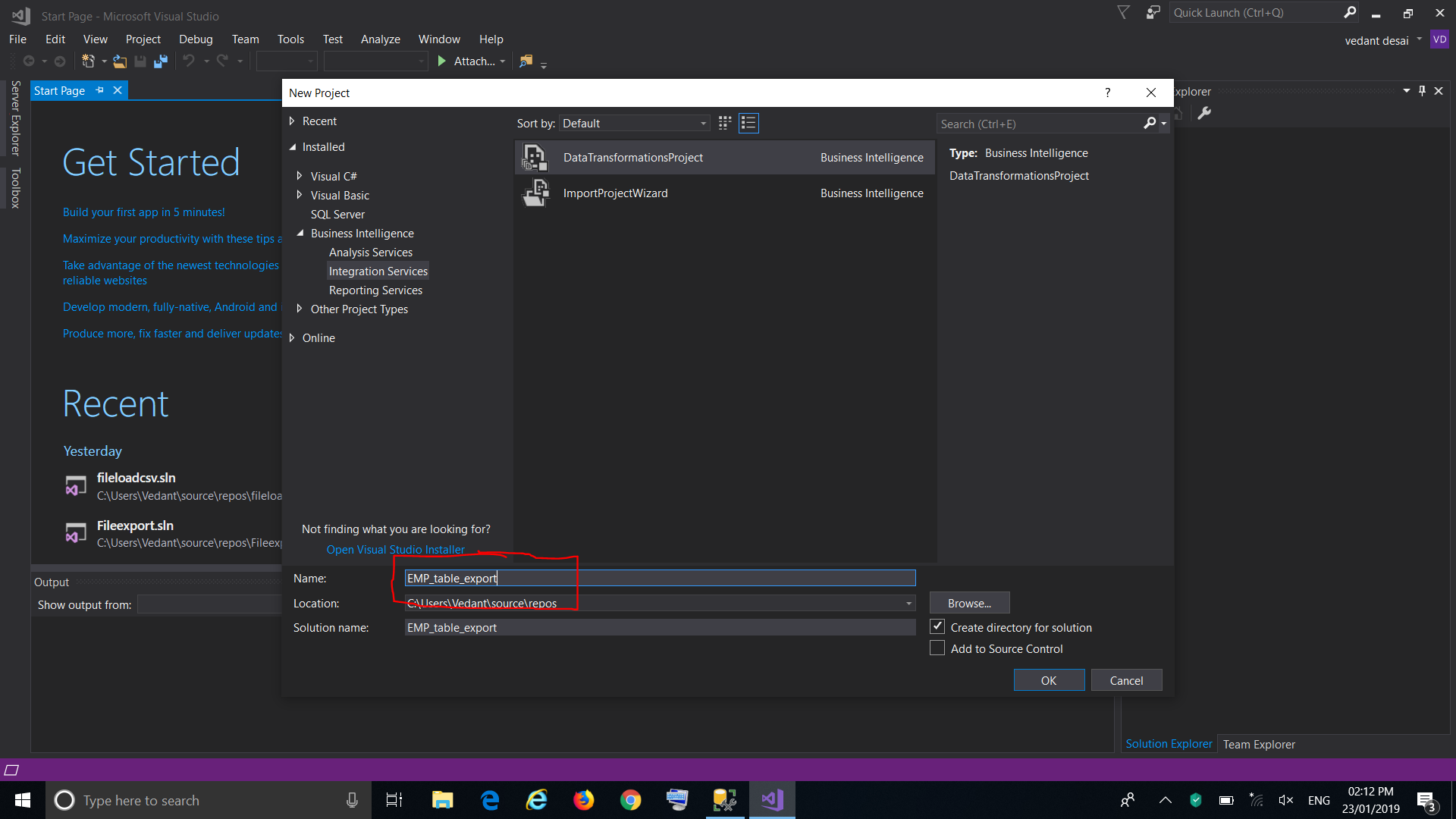
For Staging table, I created HumanResources.Emp\_staging and worked on it.

* 1. Databases

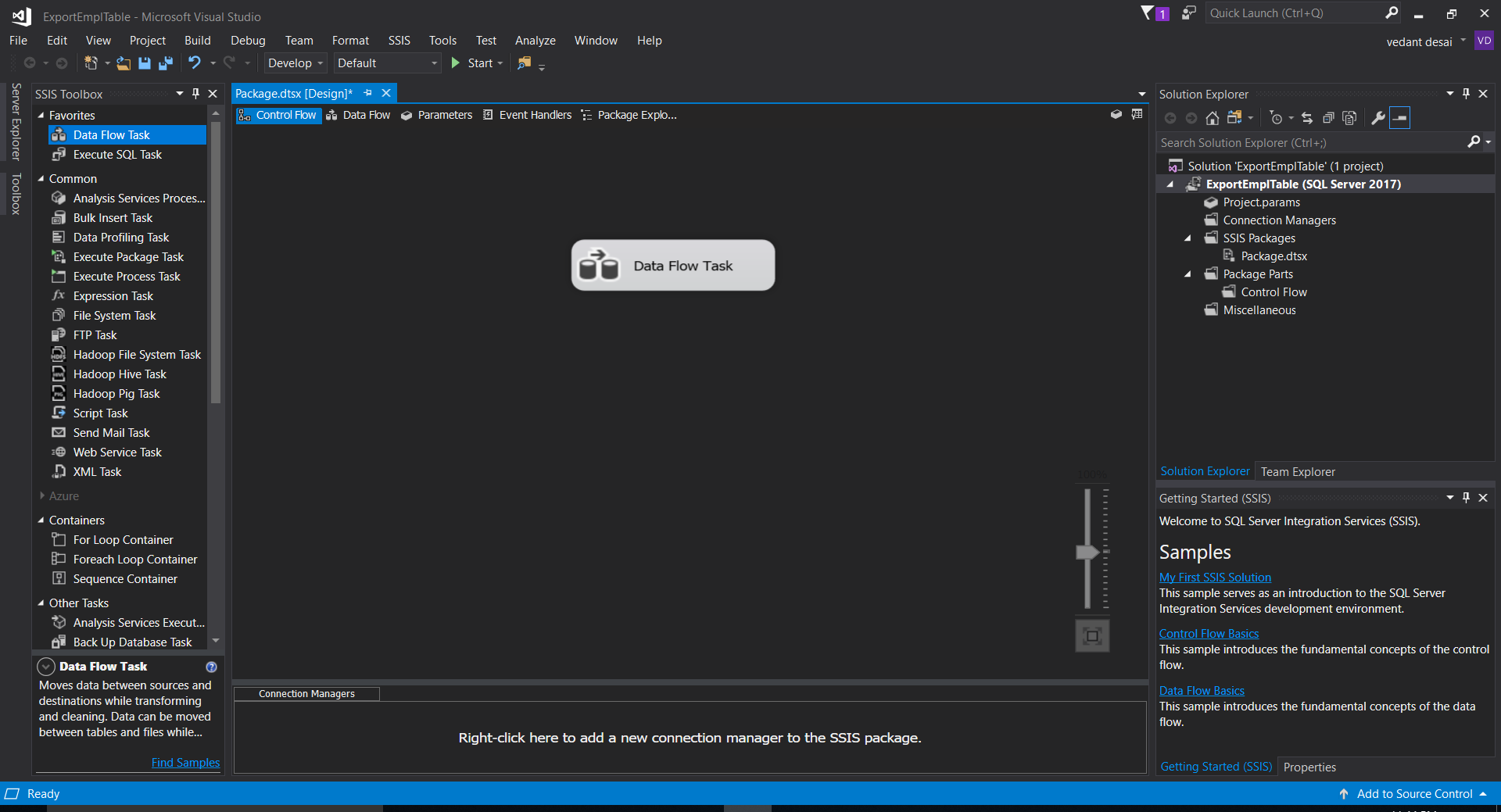
|  |  |  |
| --- | --- | --- |
| Database | Table | Description |
| AdventureWorks | HumanResource.Employee | Contains information of all employees and will be used for export to CSV |
| AdventureWorks | HumanResources.Emp\_staging | Created staging table for Employees |

* 1. Extract Transform and Load (ETL)

1. First, we will create a project for initializing a package.

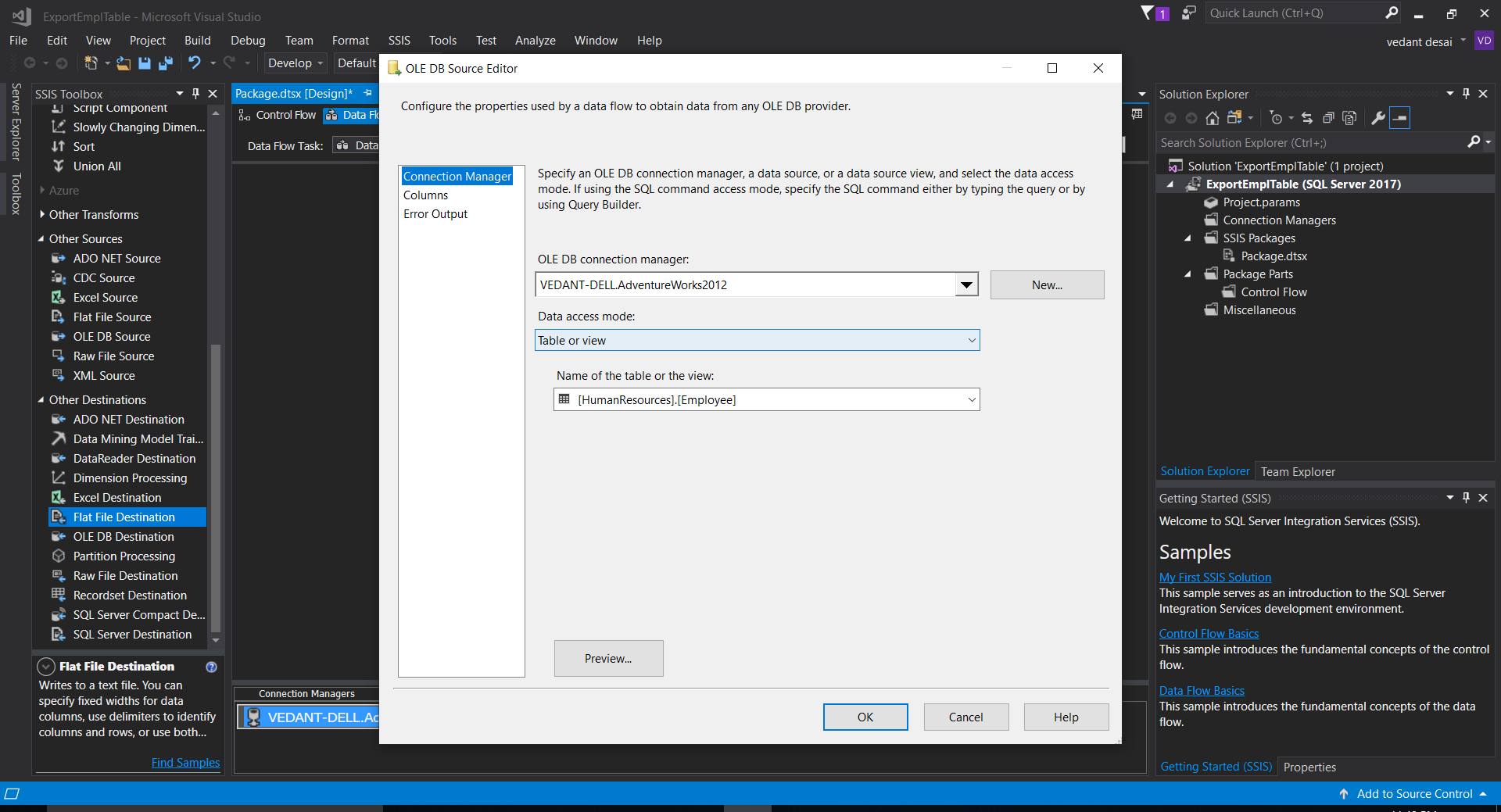


1. For any export or import we can use wizard for ease, but here we will create a Data flow task from “control Flow” to allow data to flow from database to any source.

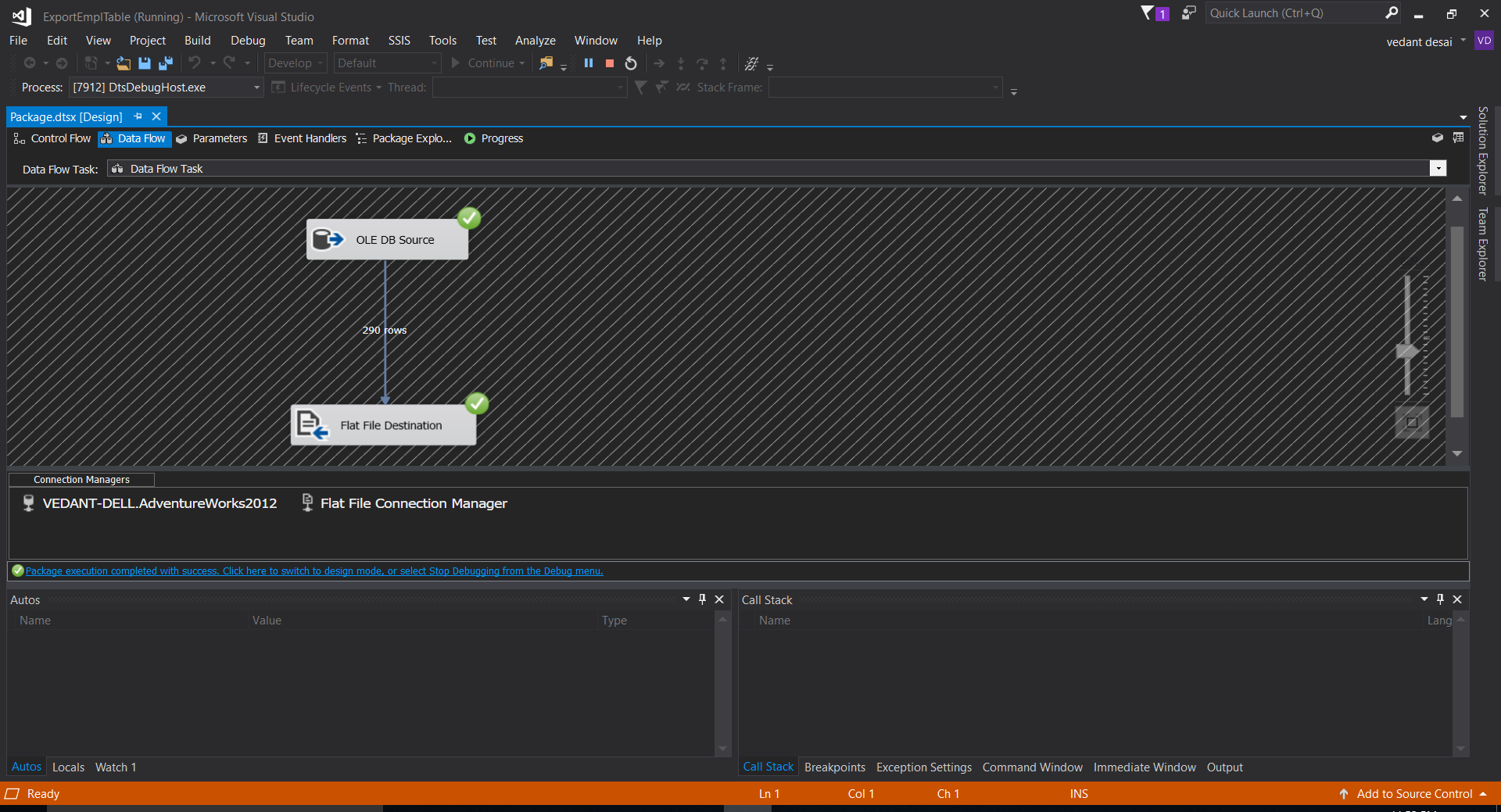


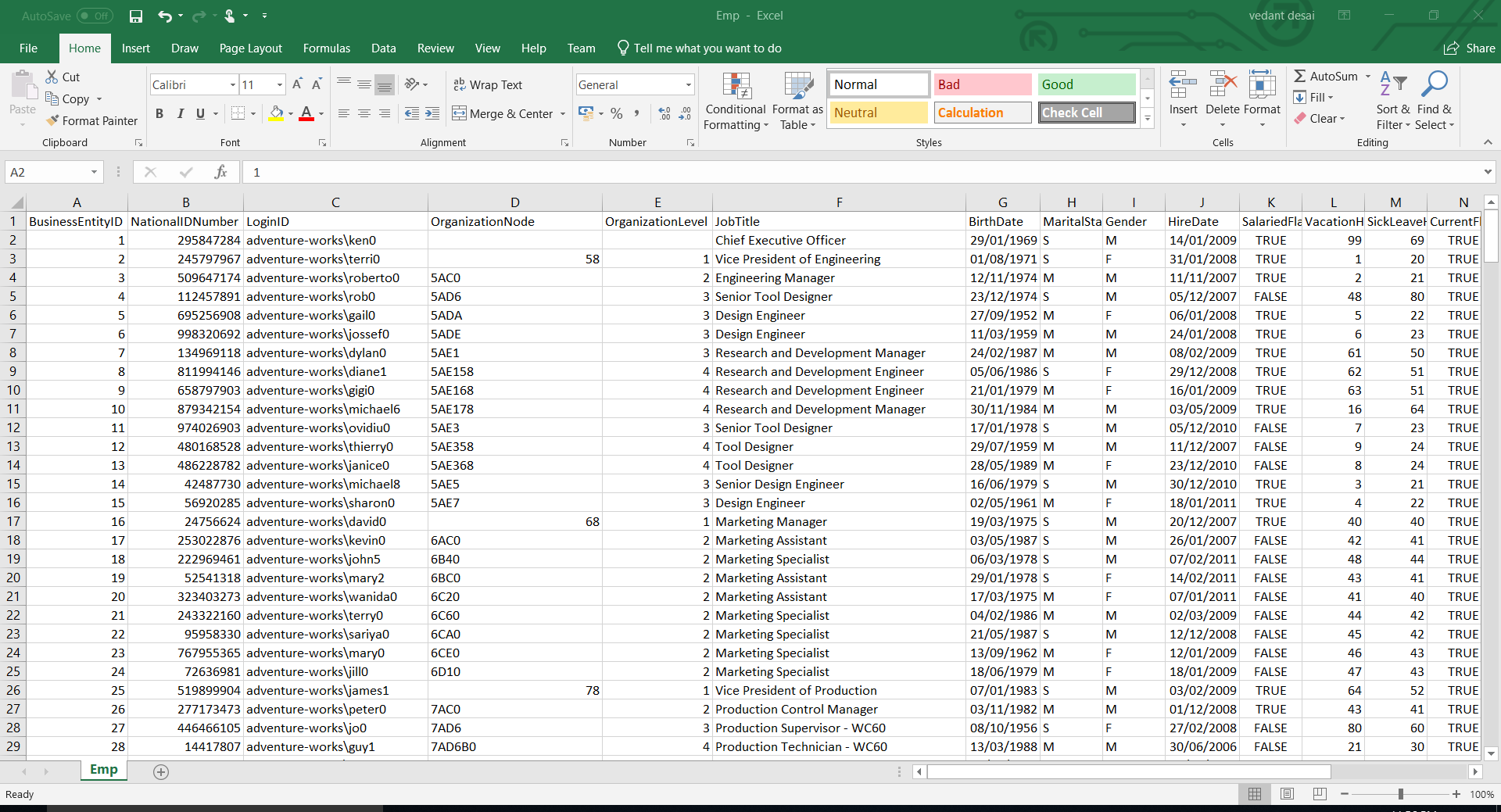
1. After Data flow we need to decide what’s the source data and target destination.

Here I used OLE DB as a Source and CSV file a destination.



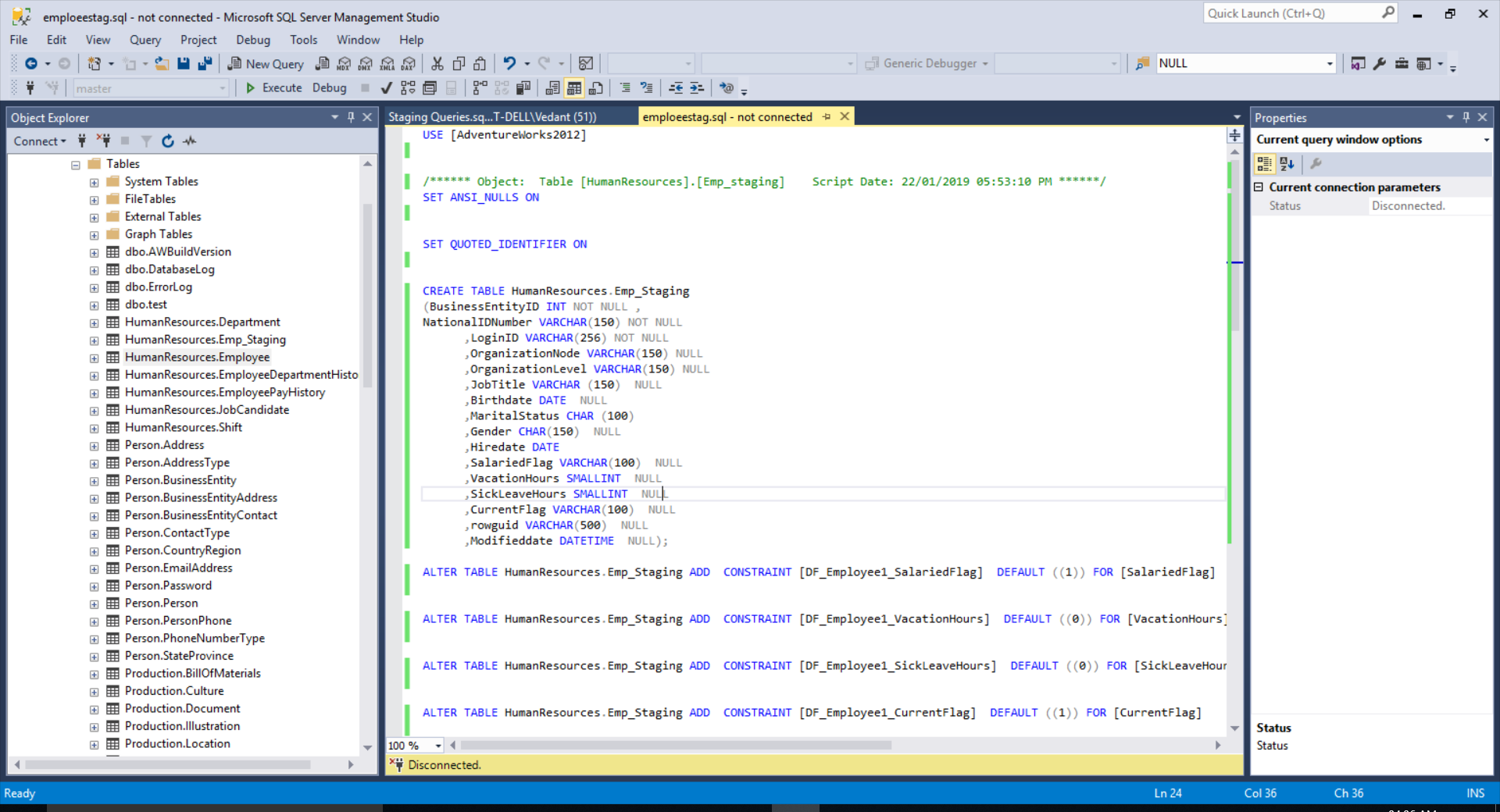
1. After Initialization of source and target, data will flow from database to CSV and when I’d debugged 290 rows from source to target are already populated in CSV.





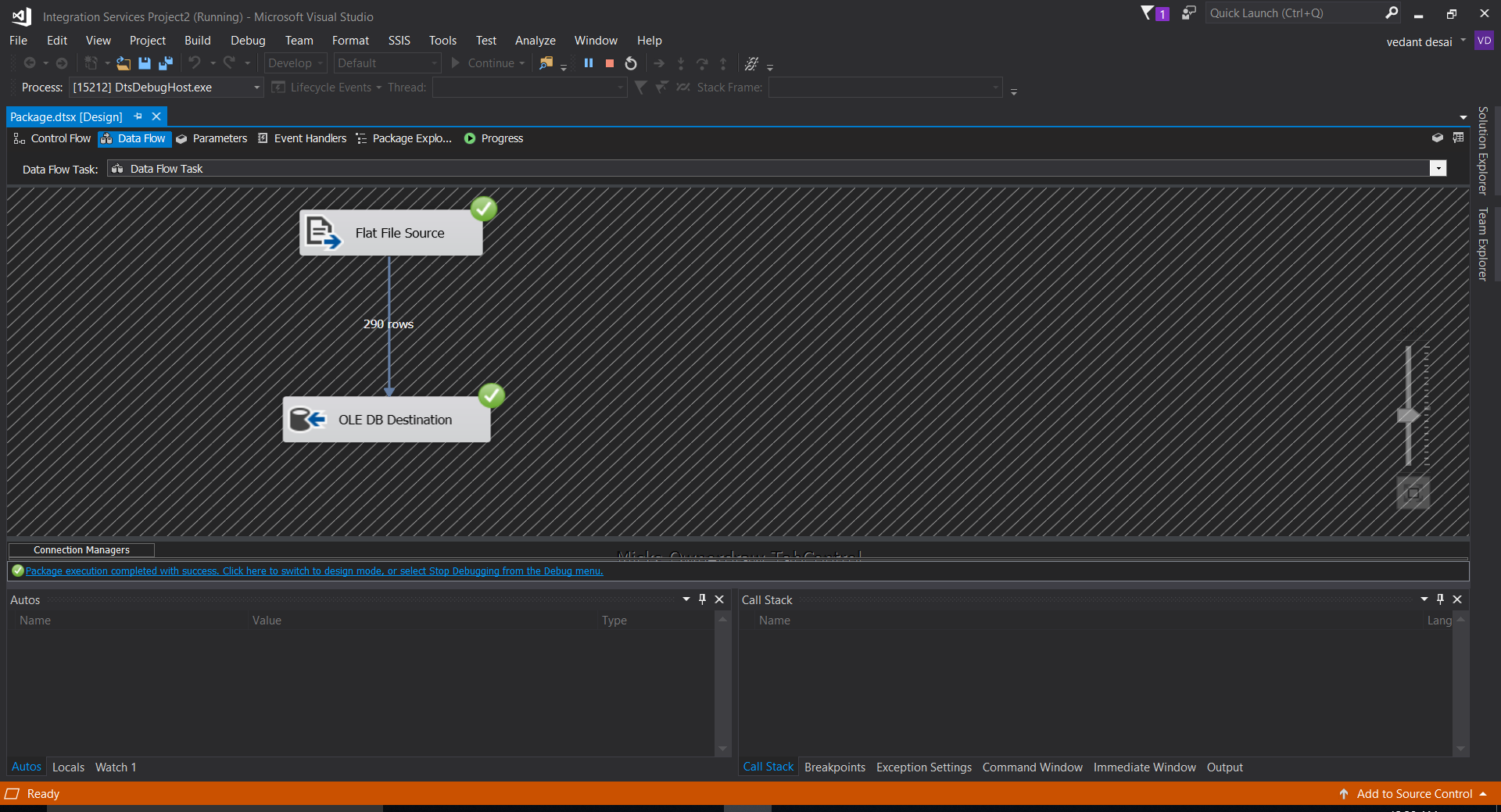
Preview of CSV file.

1. Now, we need to create a staging table with proper attribute names and their data types. Also, there are certain constraints which we must follow.

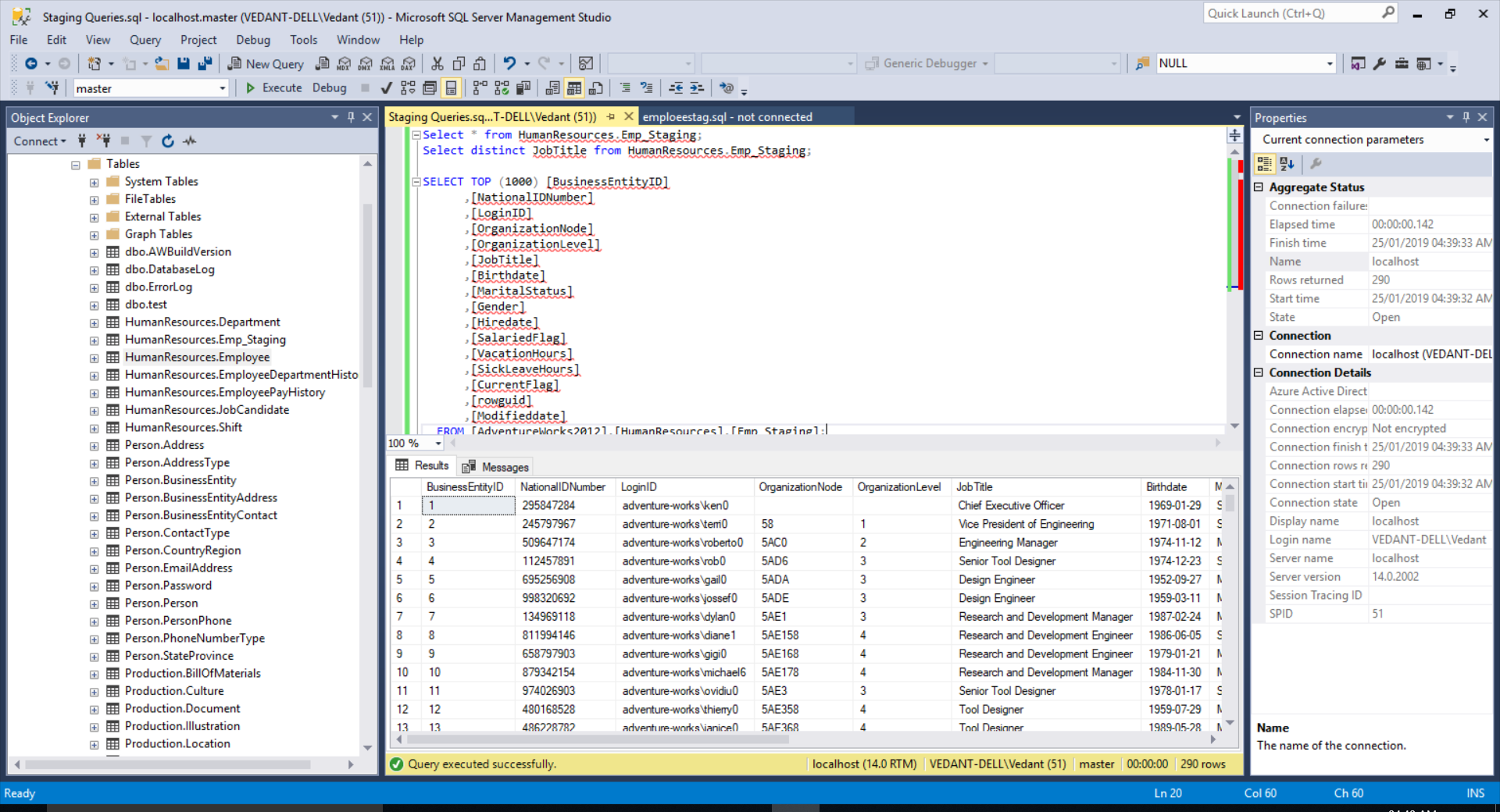


1. After creating staging table, we will import table from CSV file to the staging table.

For that we will create a package that will show the dataflow from source to destination.



1. We will check in SQL Server if staging table is populated with data.



1. MONITORING AND ALERTING

A repository is created on GITHUB so that one can have access to it.

Link: [https://github.com/Vedant1065/Data-Warehousing](https://github.com/Vedant1065/Data-Warehousing%20)