**Data Warehousing & Business Intelligence**

**Week-3 | Assignment 3 Documentation**

**Vedant Desai (001830150)**

Creation Date: 01/31/2019 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 | 01/31/2019 | Vedant | Ayush, Professor Vincent |

1. **INTRODUCTIONS**
   1. **Assignment Overview**

This assignment requires us to create two tables in Toad modeller. 1) Male\_Emp and 2) Female\_Emp with common attributes such as Firstname, Lastname, Gender, loginid.

Secondly, we need to create a SSIS package for reading employee table and splitting into Male and Female workers based on Gender by using various transformations.

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

1. **ARCHITECTURE OVERVIEW**

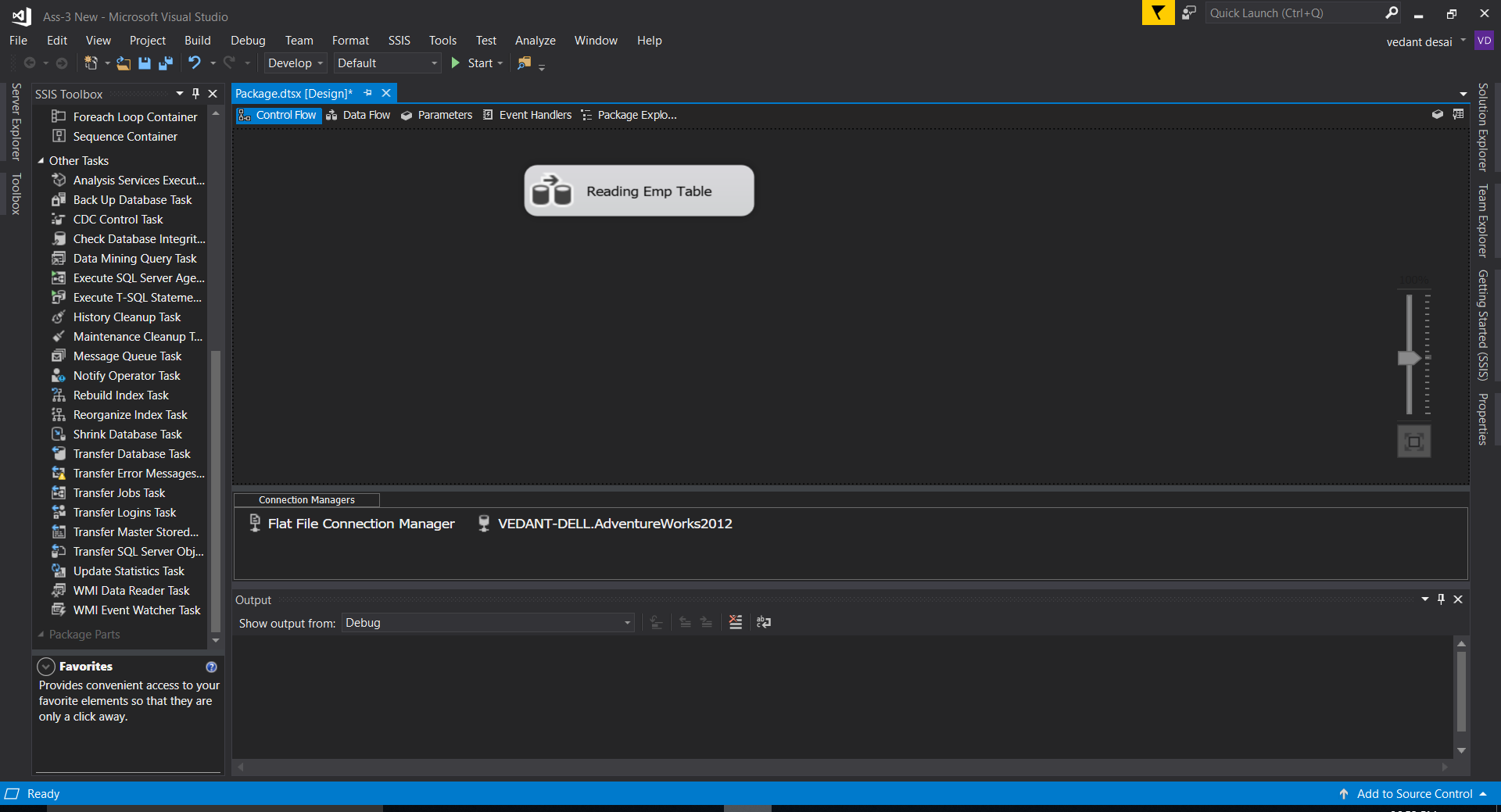
For initial read we need to use two tables 1) Employee 2) Person, as we need to join two tables. Secondly for lookup we will create a Gender\_lookup table for validation of gender.

**Databases**

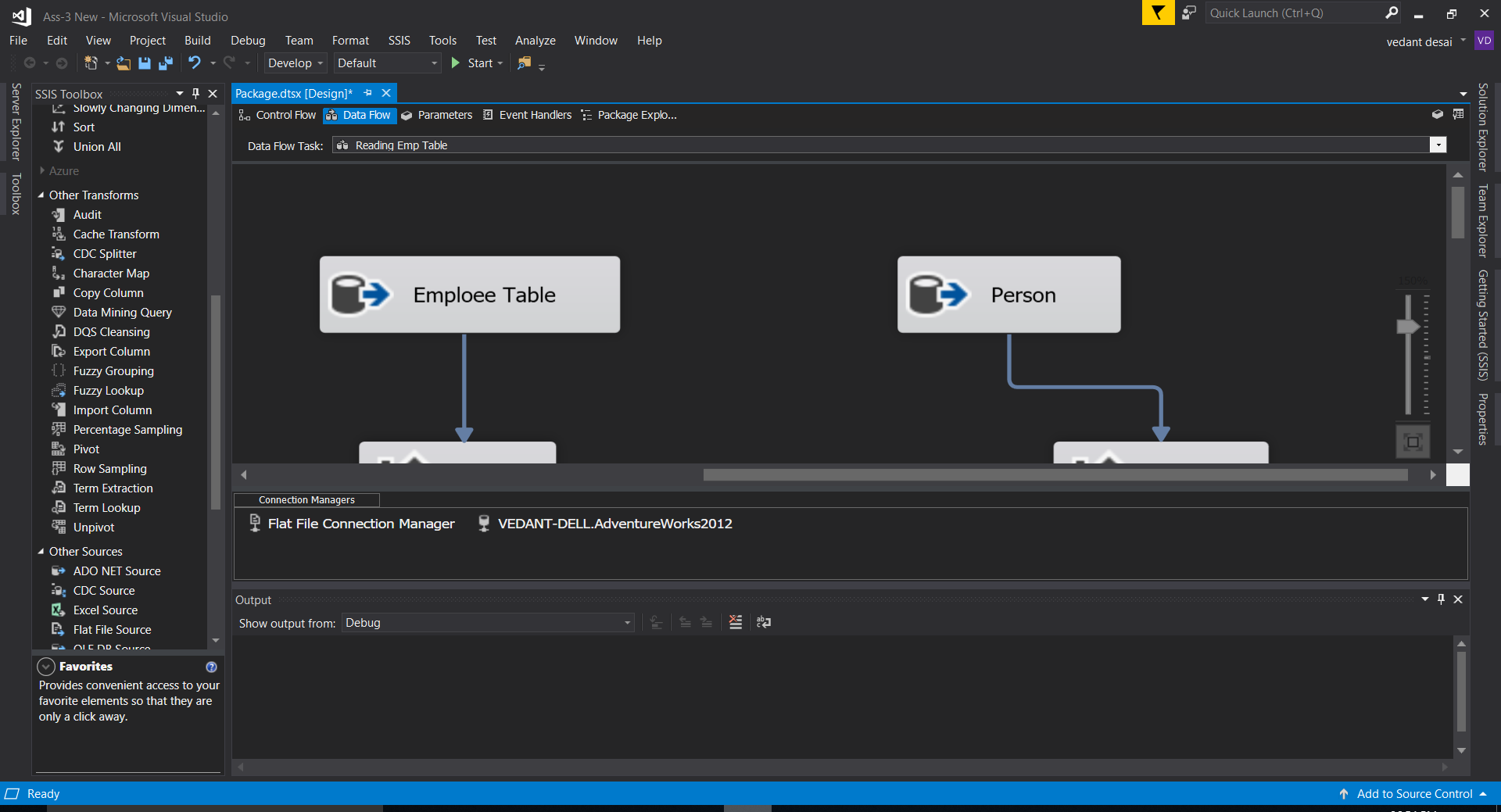
|  |  |  |
| --- | --- | --- |
| Database | Table | Description |
| AdventureWorks | HumanResource.Employee | Contains information of all employees and will be used for export to CSV |
| AdventureWorks | Person.Person | To read some columns |
| Pubs | Employee | For BCP in and out data |

* 1. **Extract Transform and Load (ETL)**

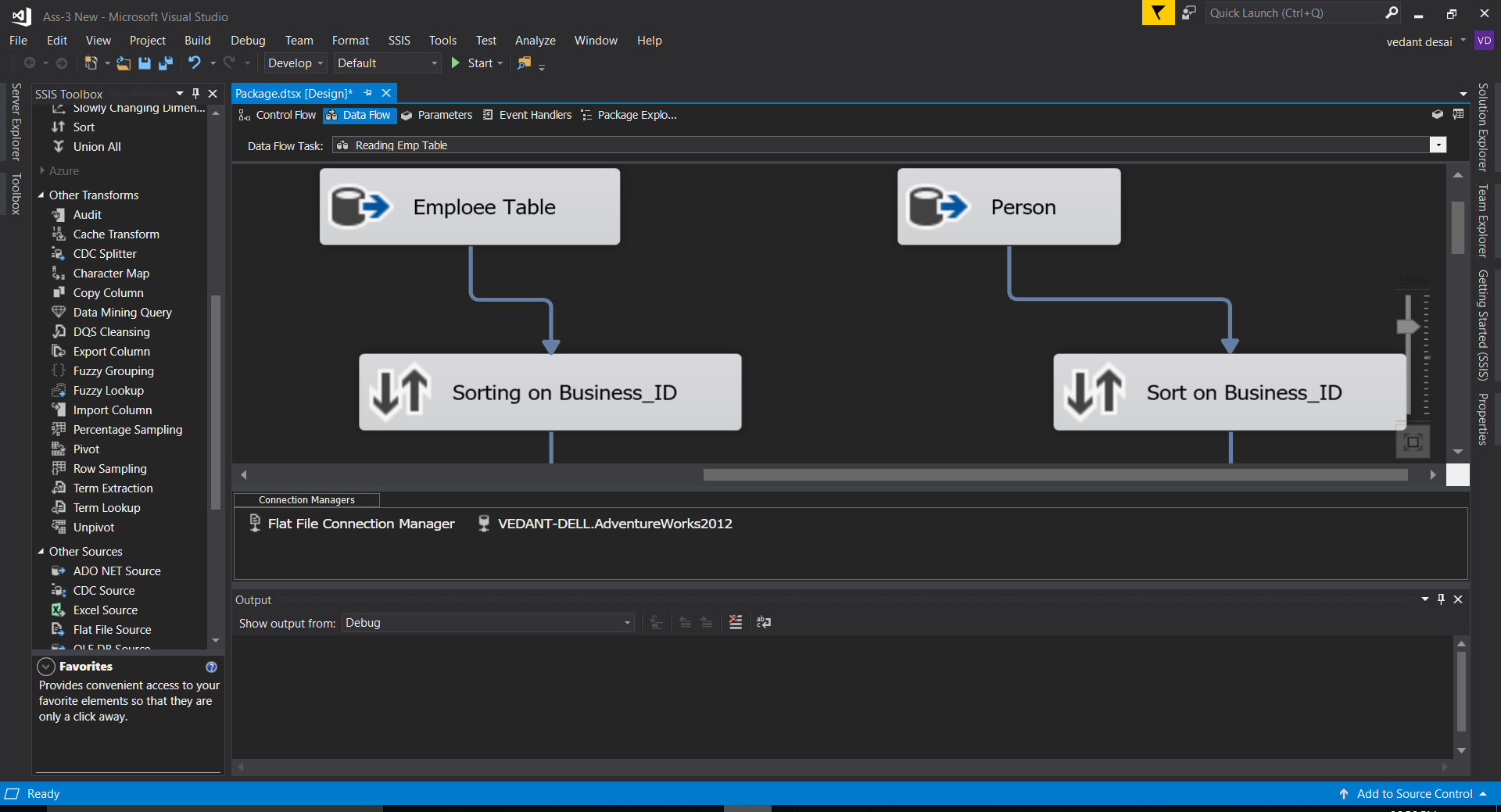
1. First, we will create a project for initializing a package.
2. For any export or import we can use wizard for ease, but here we will create a Data flow task from “control Flow” to allow data to flow from database to any source. Task is to read the Employee table and split into male and female employees.



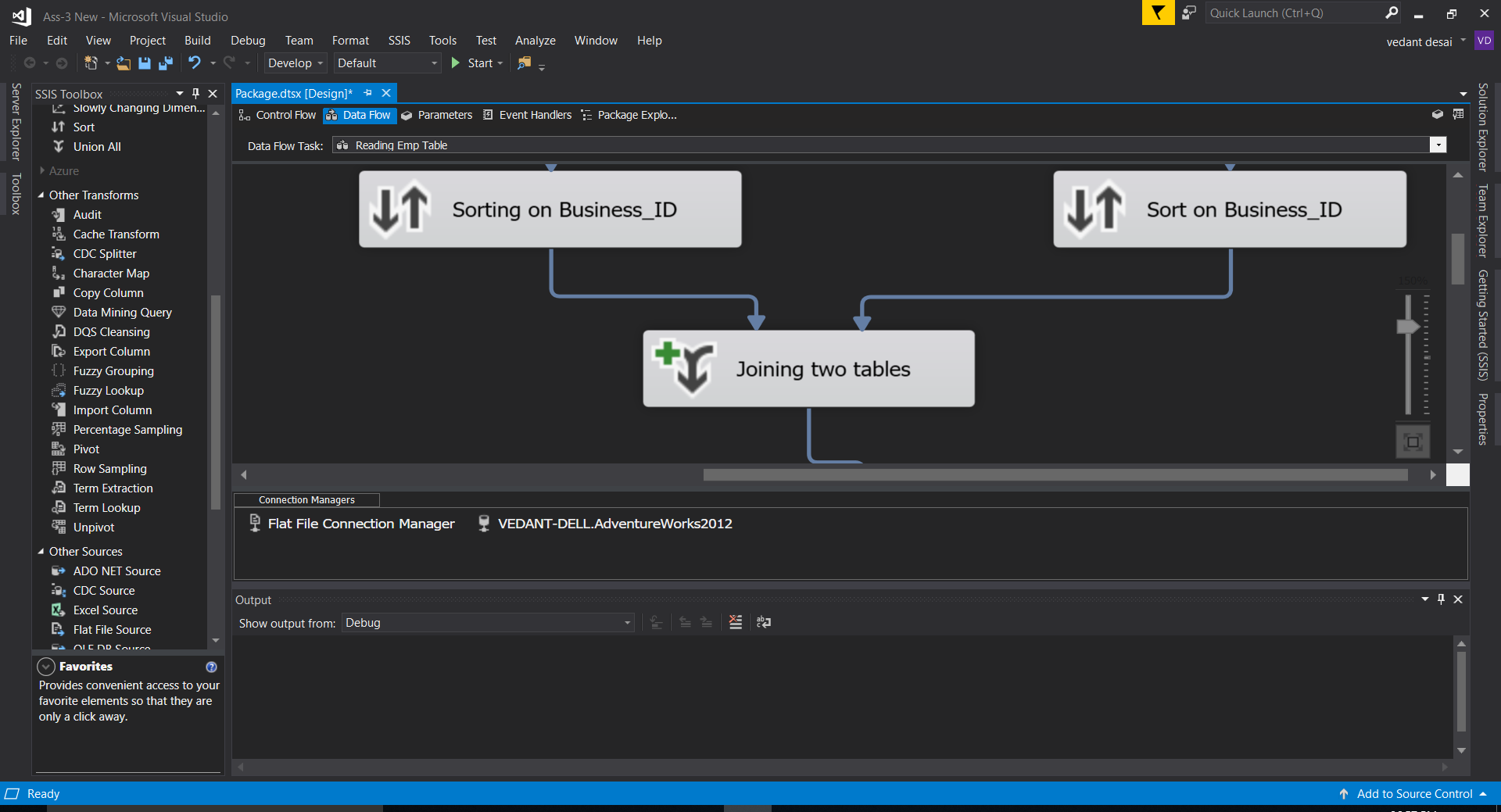
1. After Data flow task we will read data from employee table and person table as we need some columns that is not present in employee table.



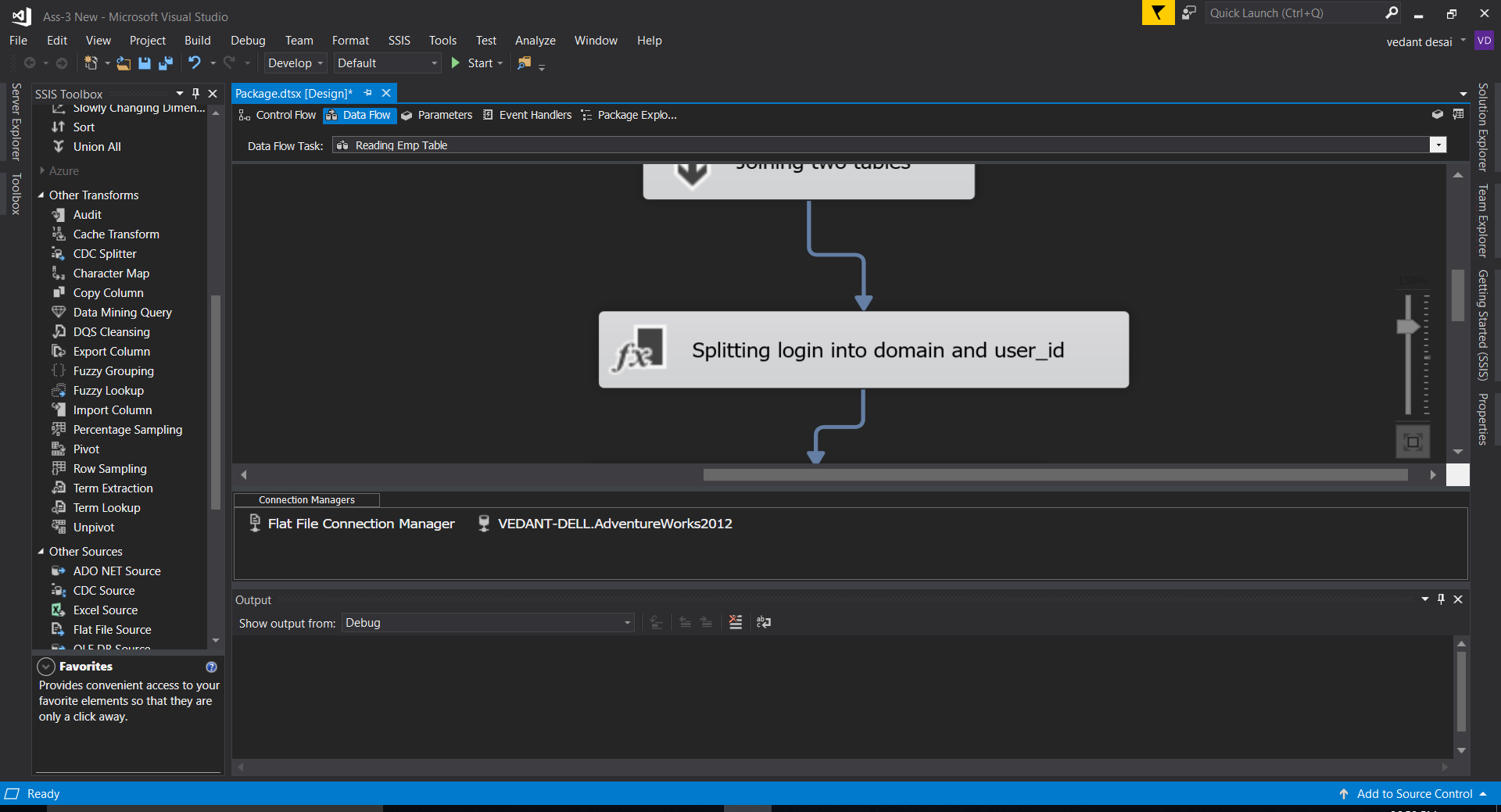
1. After Initialization of source we will sort the primary key of both the tables in ascending order as we will be merging in next step



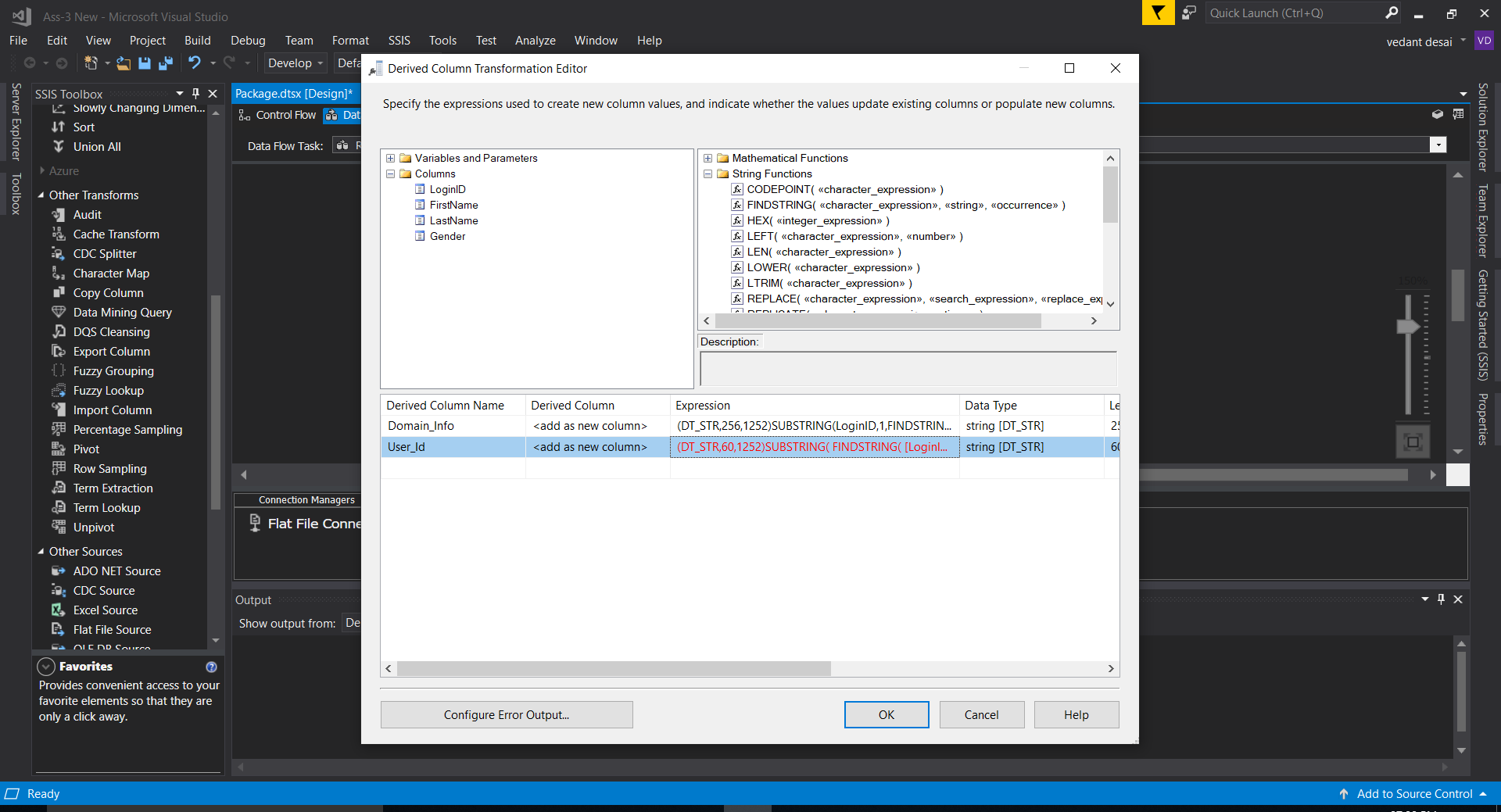
1. Now, we need to inner join both the tables and get firstname and lastname from person table and join with gender of employee table.



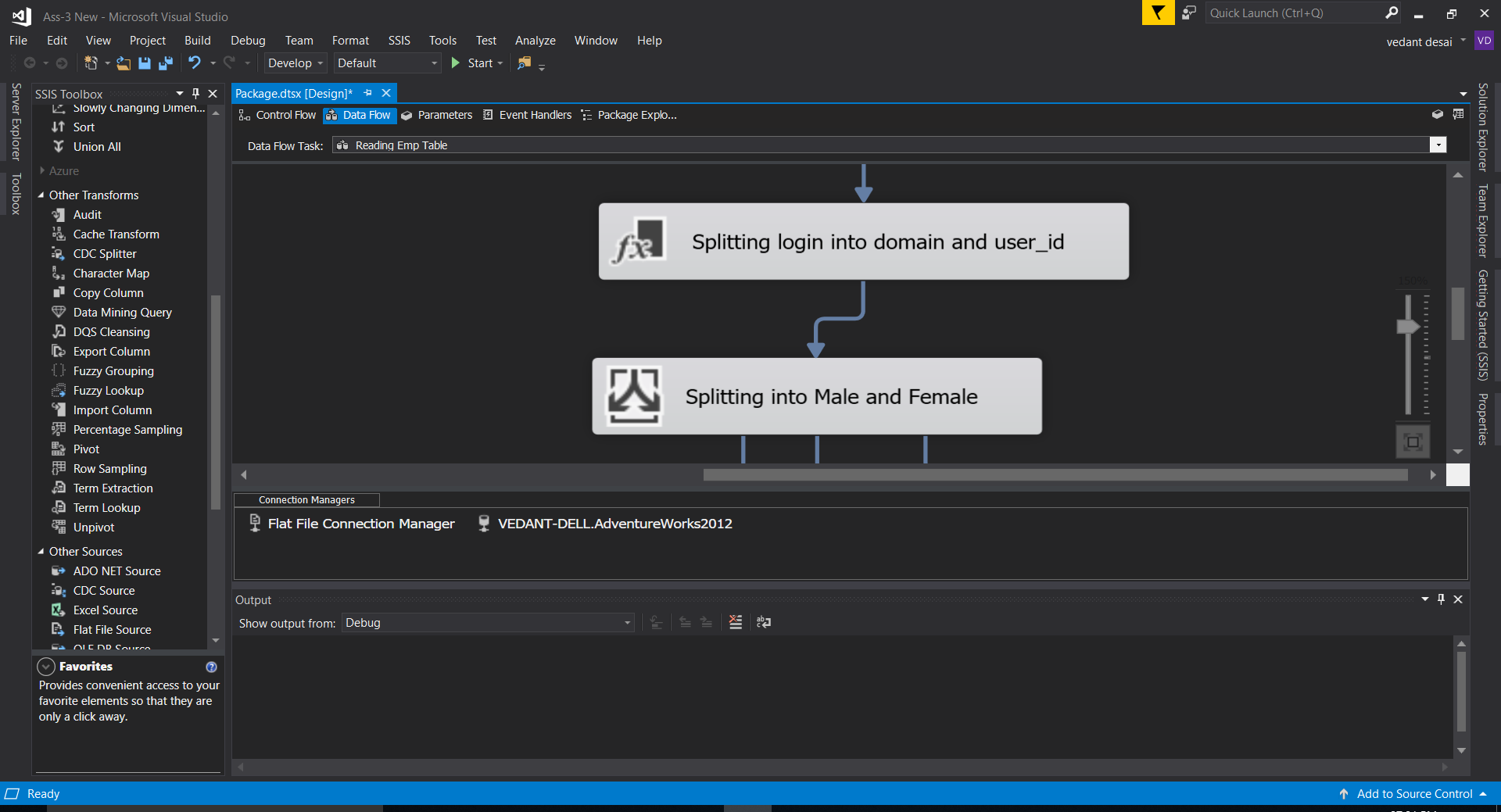
1. After joining the tables our task is to split the table column “login\_id” into domain and user\_id. For that we will use derived column transformation.

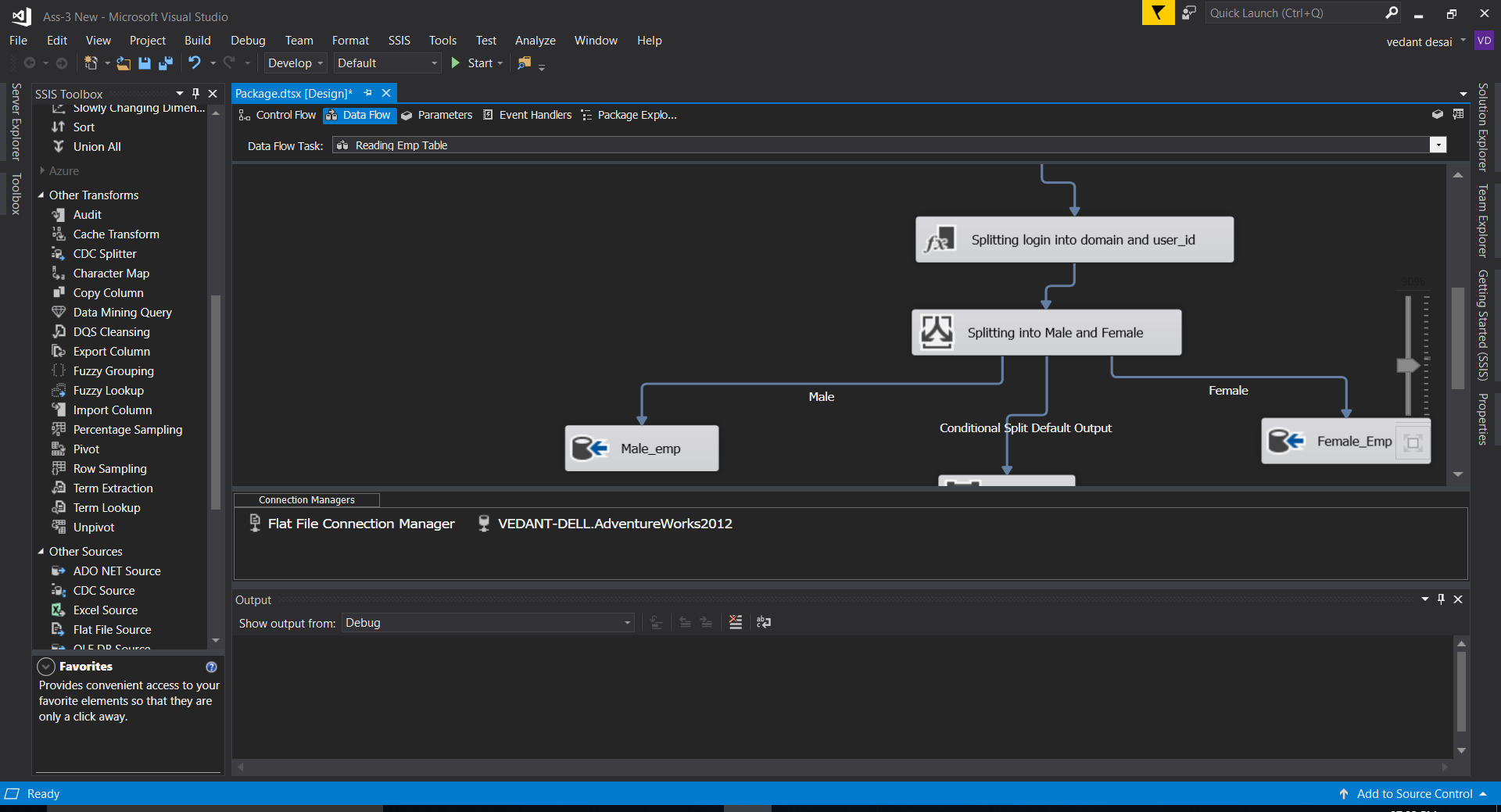


1. We will use the substring function for derivation and converting into string datatype using type\_caste()

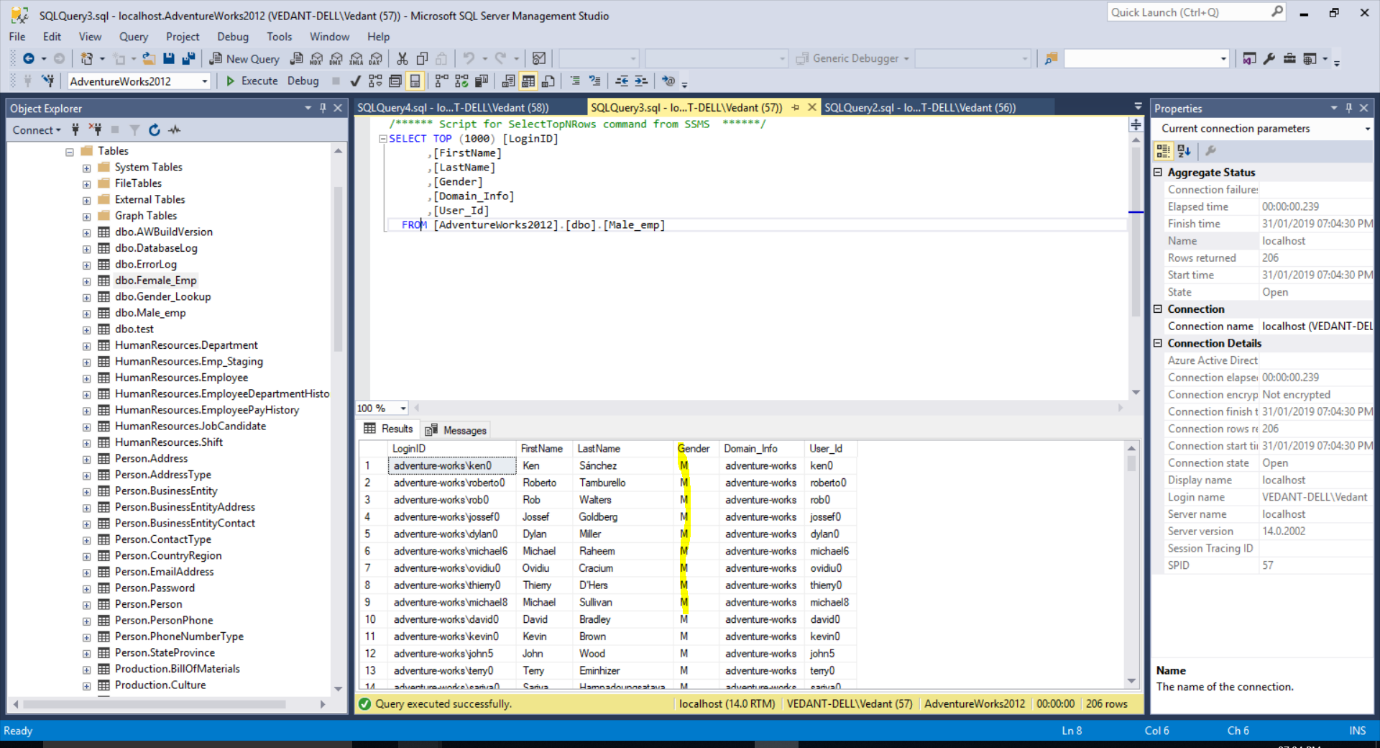


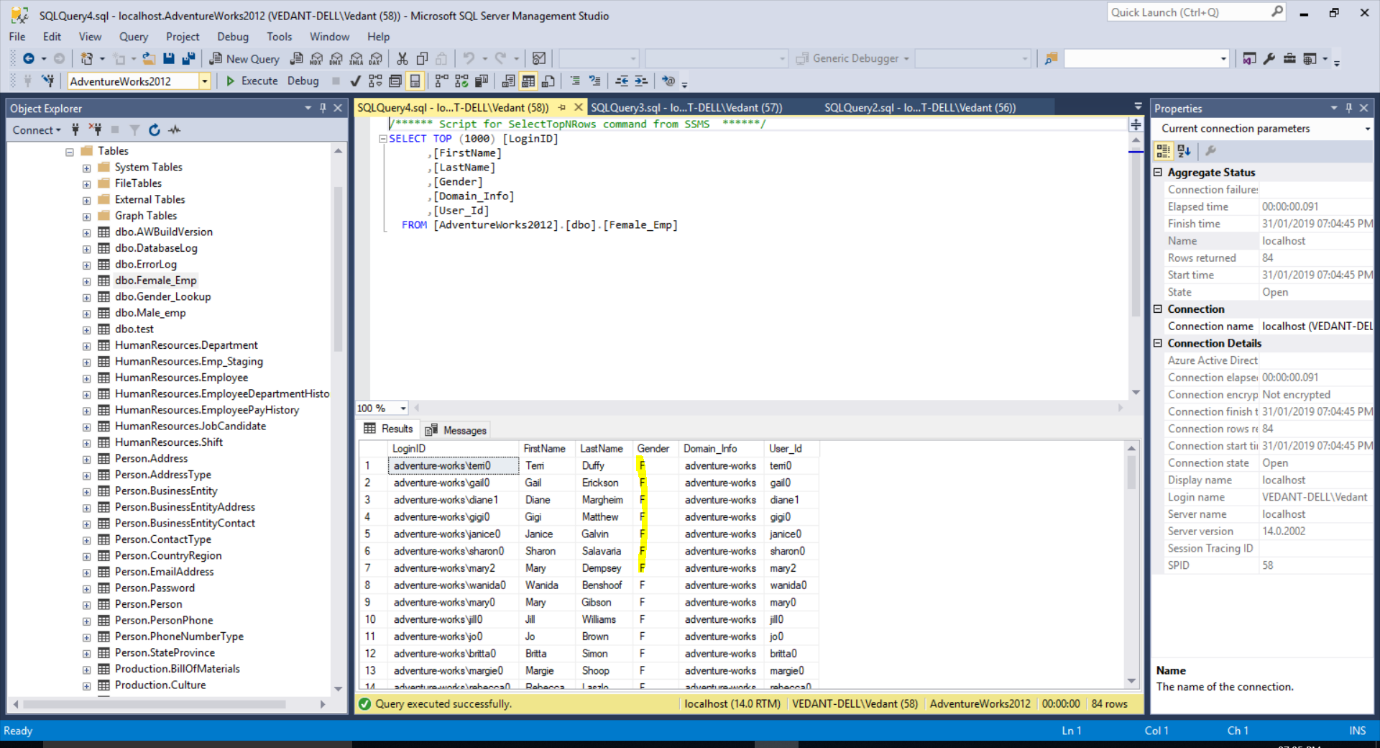
1. Now we will use Conditional split transformation to split the merged table to Male\_Emp and Female\_Emp



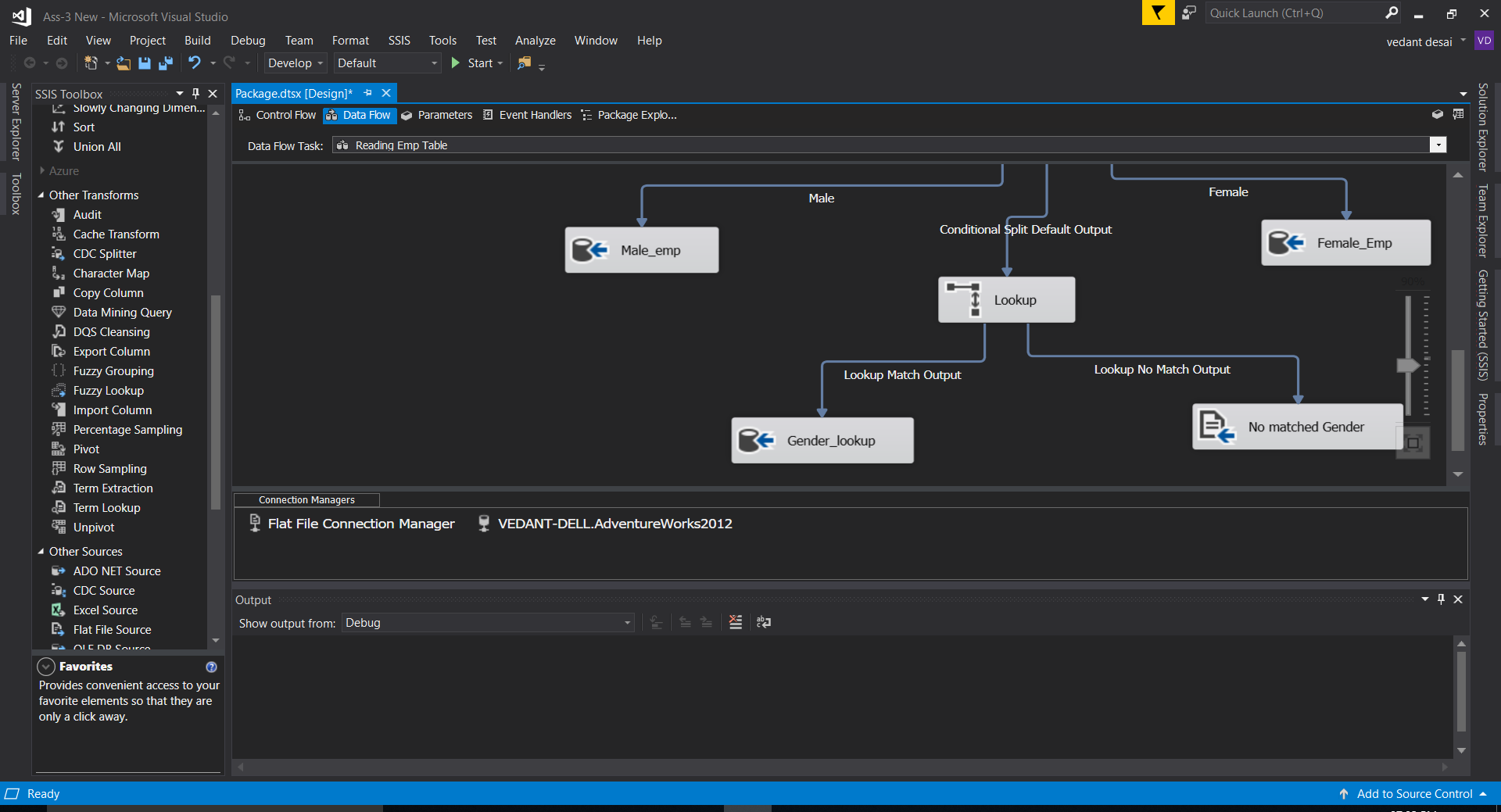


Below attached are the tables that is derived from above transformation.

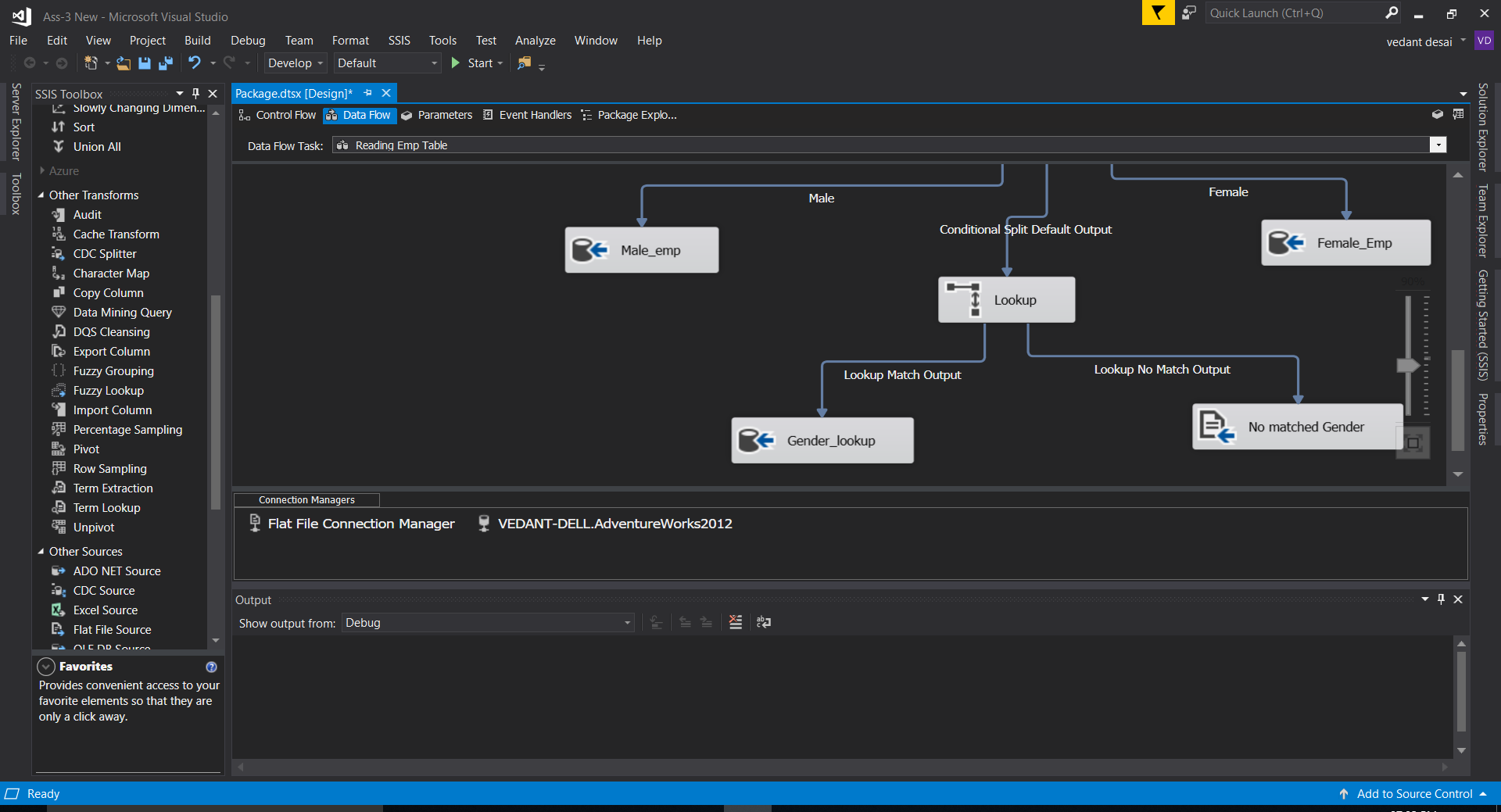




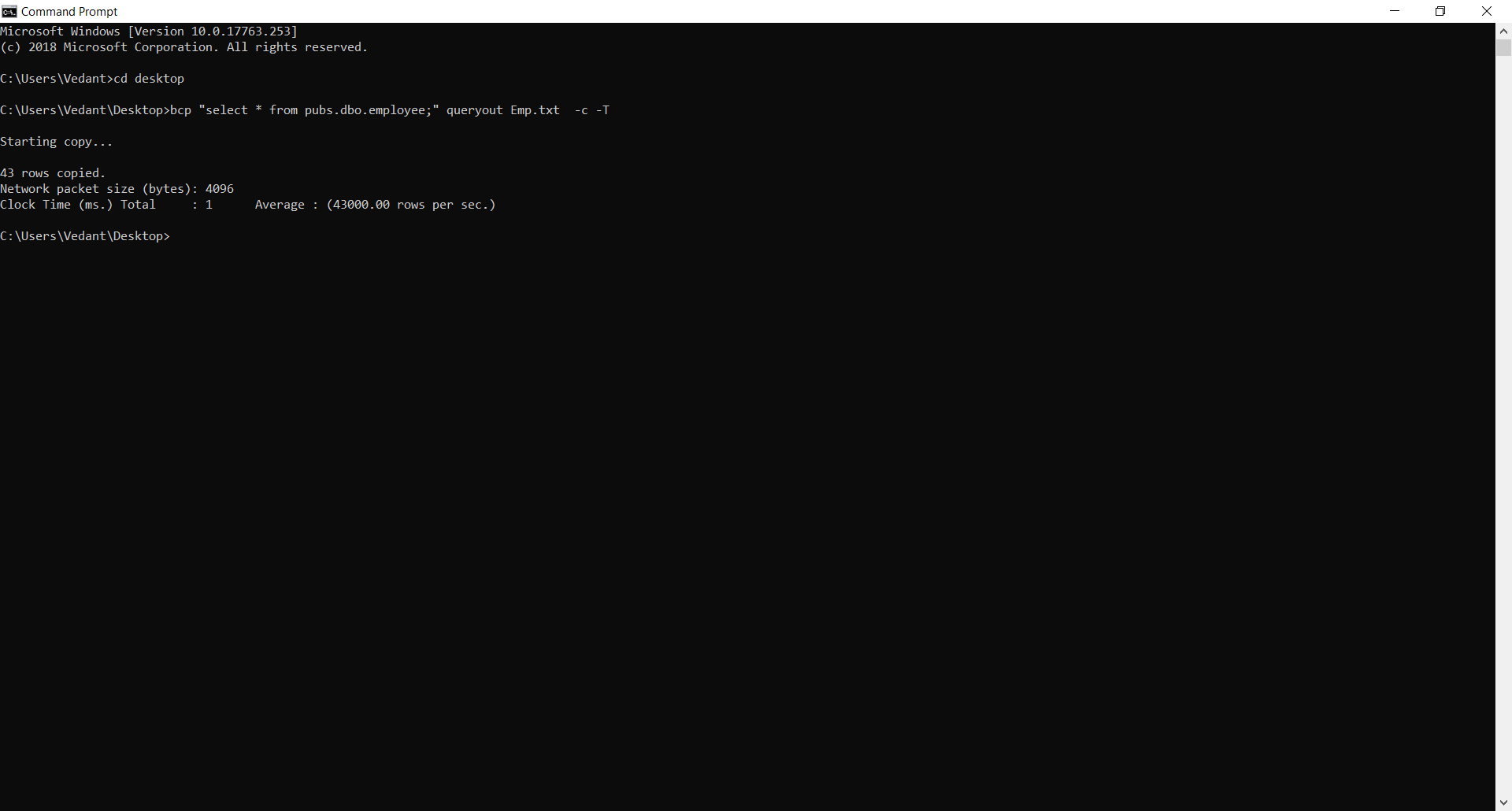
1. Now we will use the look\_up transformation to validate the Gender and creating a gender\_lookup table for validation purpose.



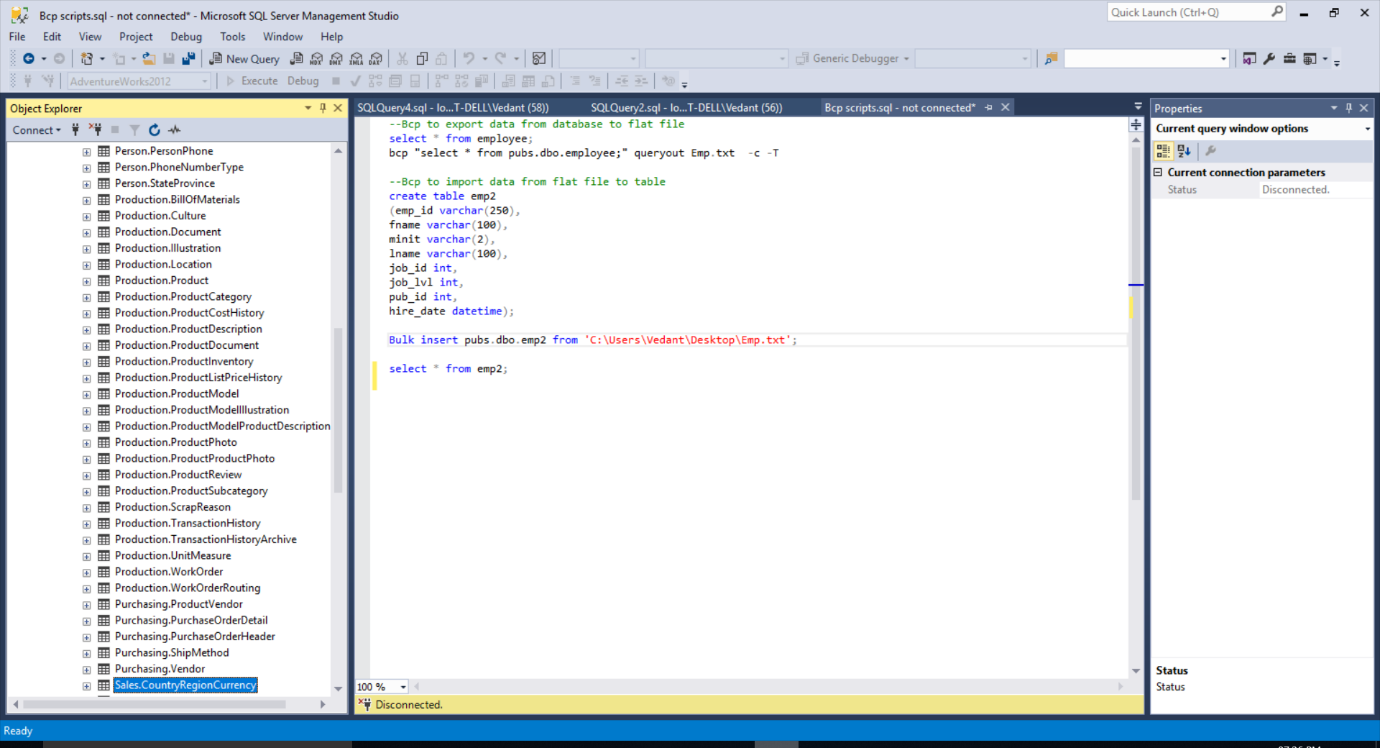
1. Now we must handle the error if there will be an unknown Gender, so we have created a flat file to handle unknown gender in table.



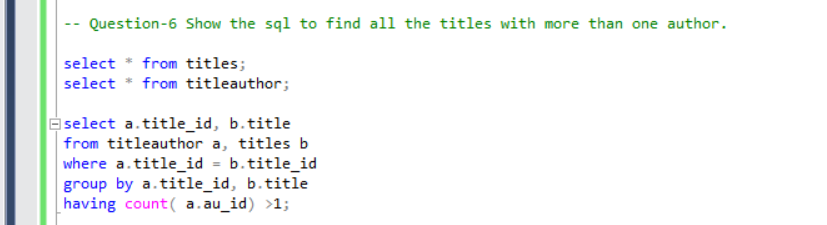
1. Now we have to use Pubs database and read the emp table and do BCP in and out of SQL. For that we will use Command prompt.



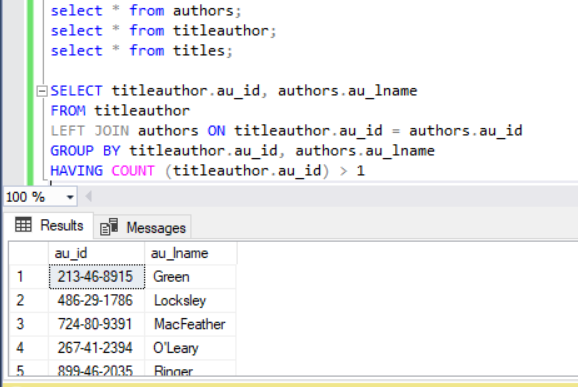
1. For Bulk insert we will use SQL command.



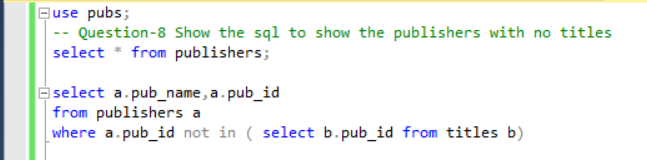
1. Show the sql to find all the titles with more than one author.



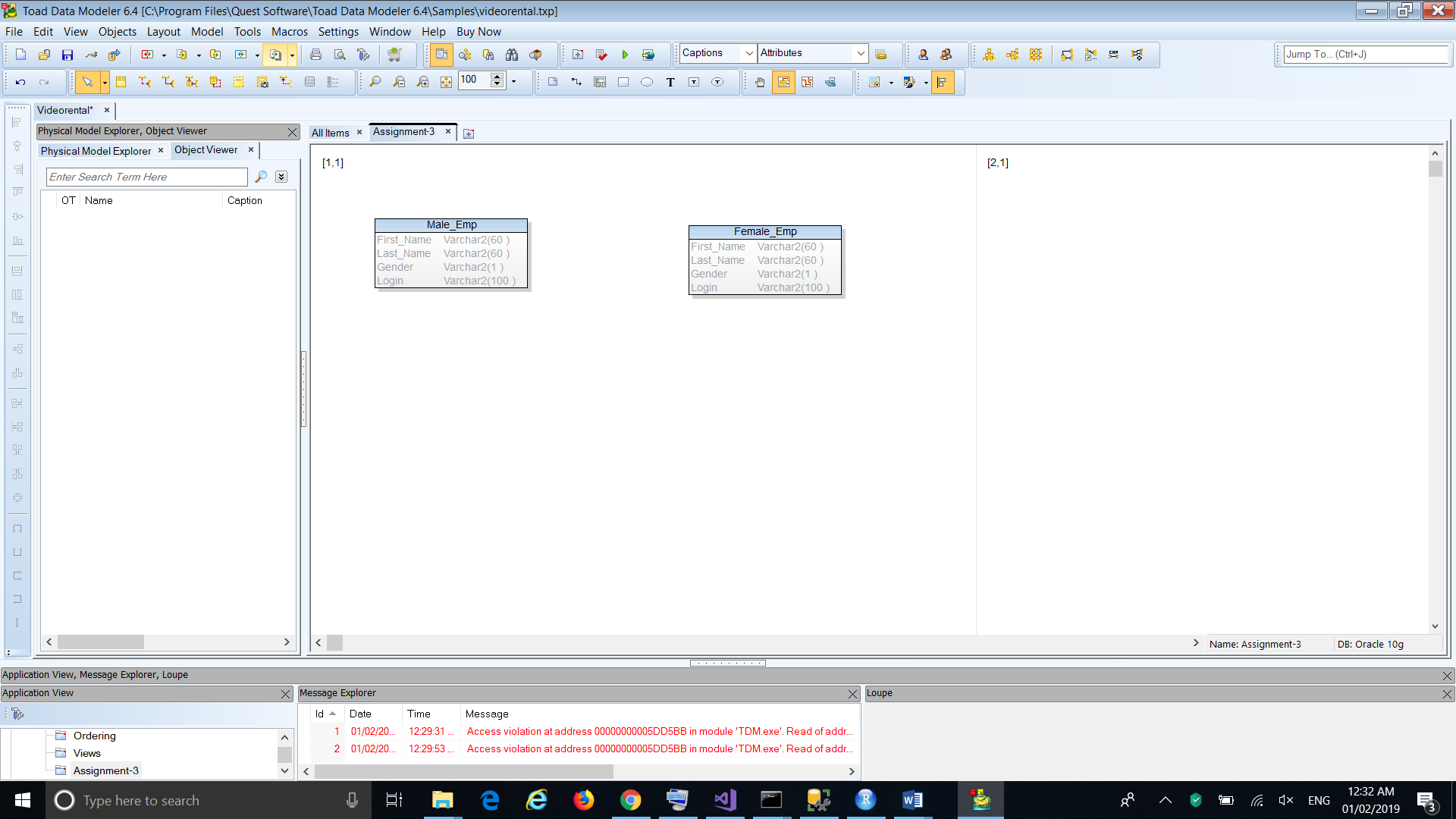
1. Show the sql to show all the authors with more than one book



1. Show the sql to show the publishers with no titles



1. Creating tables in Toad



1. **MONITORING AND ALERTING**

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

Link: <https://github.com/Vedant1065/Assignment-3.git>