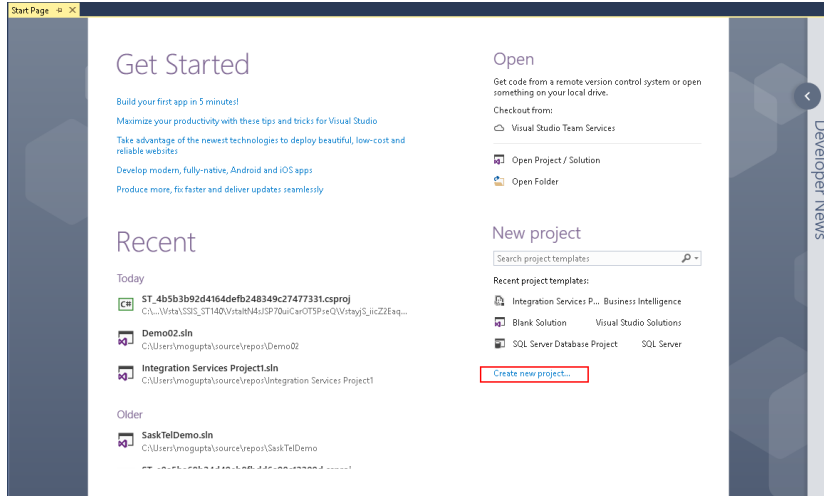


SQL SERVER INTEGRATION SERVICES

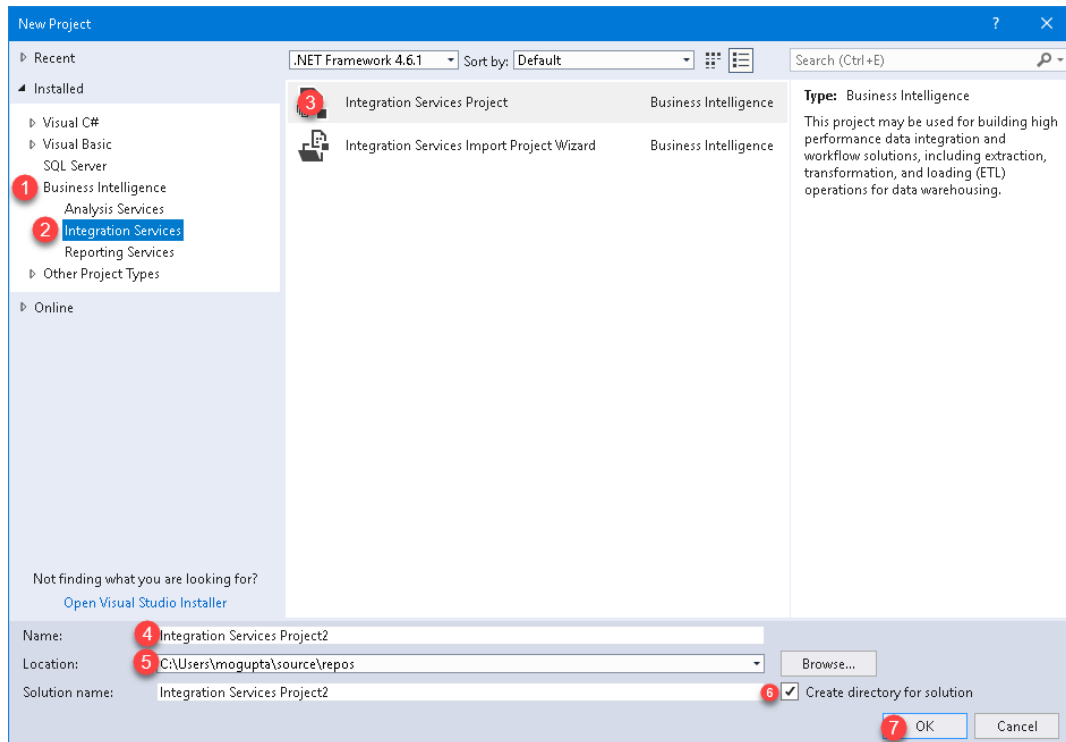
MODULE 03: DATA FLOW

Note: Save this lab once completed, we will need this lab at later time.

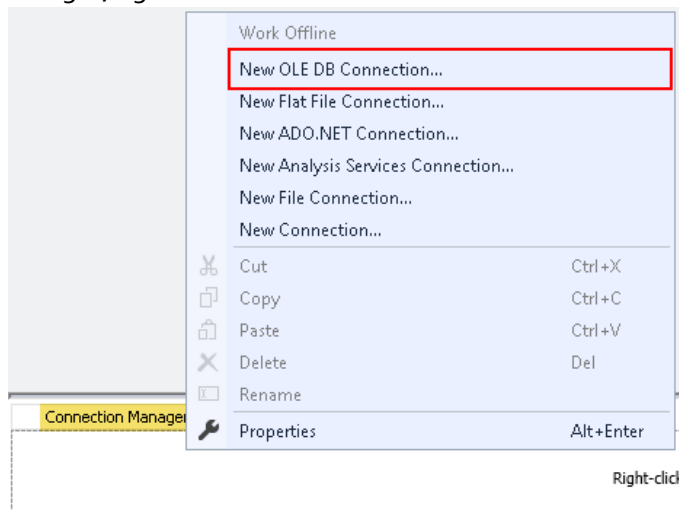
1. Launch SQL Server Data Tools (SSDT), under start menu look for Visual Studio 2017 (SSDT).
2. In the Start Page, click Create new project.



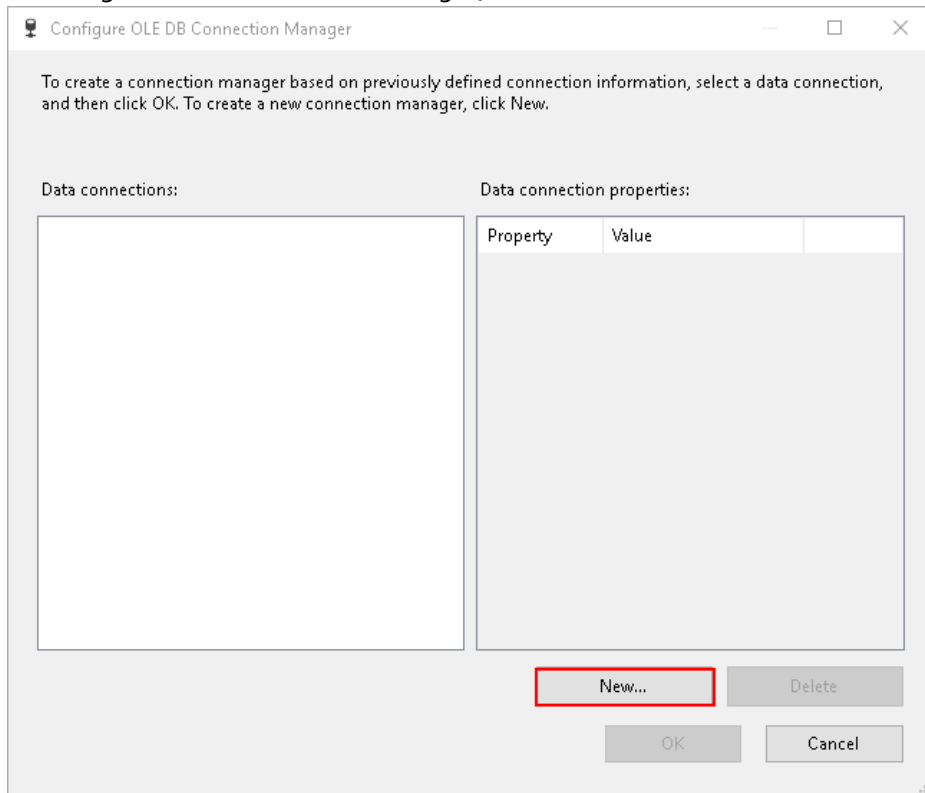
3. In New Project dialog box select Business Intelligence > Integration Services > Integration Services Project. On the bottom enter in project name and location you wish to save the project. Make sure "Create directory for solution" is selected and click OK.



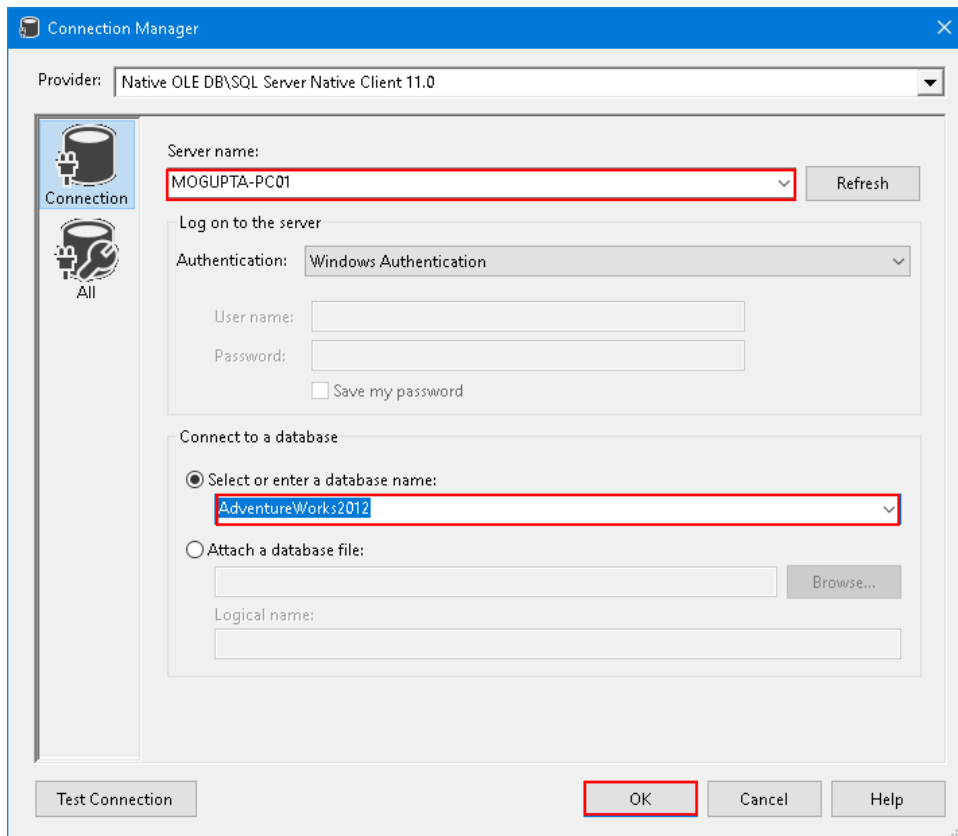
- Let's set up a connection manager to our database. In the bottom center pane under Connection Manager, right-click select New Ole-DB Connection.



5. In Configure OLE DB Connect Manager, click New.



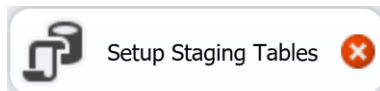
6. In connection manager, type the server name, select the database “AdventureWorks2012” and click OK.



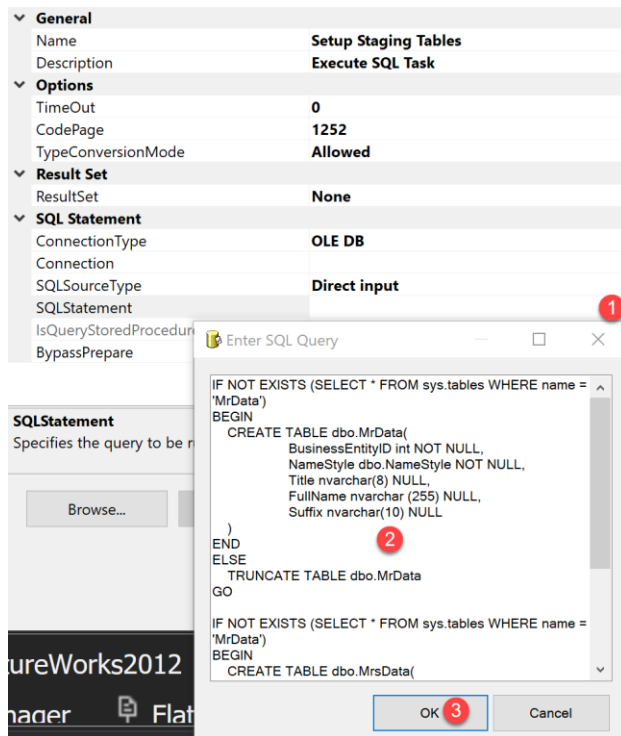
7. Click OK in Configure OLE DB Connection Manager. You should see a single connection under Connection Manager.

Hint: Rename the connection manager to OLEDB.ServerName.DatabaseName. This will make it easier to identify which driver is being used for the driver.

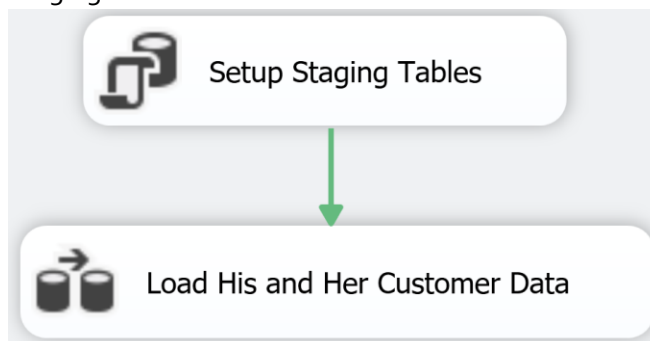
8. Create an Execute T-SQL Task. Rename it to “Setup Staging Tables”.



9. Copy the T-SQL Code in Labs\Lab_Files\Module03\Labo8\CreateTablesForLab.sql and paste it in the SQLStatement.

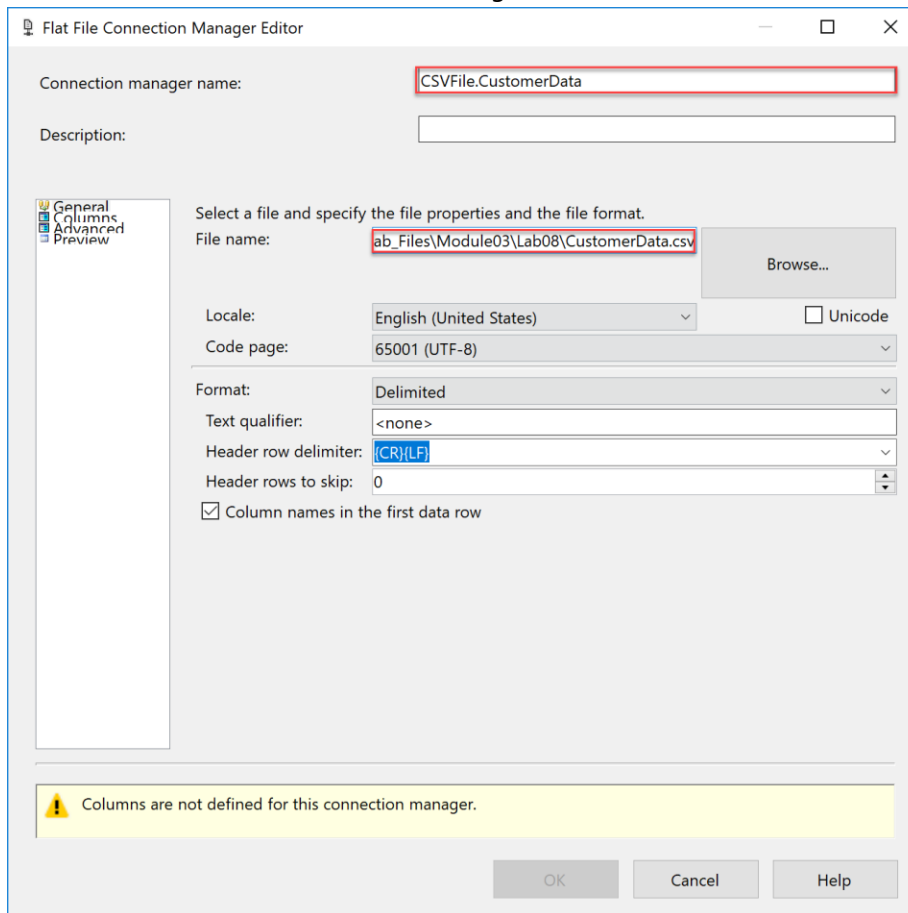


10. Update the Connection to point to connection created in step #7. Right click on the "Setup Staging Tables" Task and select execute, as these tables will be required at a later step.
11. Create a Data Flow task and rename it to "Load His and Her Customer Data". Link the "Setup Staging Tables" to data flow task.



12. Double-click on the data flow task, to start building the data flow for loading CSV data, removing the duplicates, distributing the data into multiple streams, splitting the data by His and Her dataset, combining the respective data from His and Her streams, build a new derived column for each set, sort each set, and save it in their respective tables.
13. So, to start, please create a Flat File Source task. Rename it to "Load Customer CSV Data". Double-click it to configure the task and click "New" to build a new connection manager object.

14. Rename the object to CSVFile.CustomerData. Point the file to C:\SSIS\<username>\Lab_Files\Module03\Lab08\CustomerData.csv.



The image shows the 'Flat File Connection Manager Editor' dialog box. The 'Connection manager name' is 'CSVFile.CustomerData'. The 'Description' field is empty. On the left, the 'General' tab is selected. The 'File name' is 'ab_Files\Module03\Lab08\CustomerData.csv'. The 'Locale' is 'English (United States)' and 'Unicode' is unchecked. The 'Code page' is '65001 (UTF-8)'. The 'Format' is 'Delimited'. The 'Text qualifier' is '<none>'. The 'Header row delimiter' is '{CR}{LF}'. The 'Header rows to skip' is '0'. The checkbox 'Column names in the first data row' is checked. A yellow warning bar at the bottom states 'Columns are not defined for this connection manager.' The 'OK', 'Cancel', and 'Help' buttons are at the bottom right.

Flat File Connection Manager Editor

Connection manager name: CSVFile.CustomerData

Description:

General
Columns
Advanced
Preview

Select a file and specify the file properties and the file format.

File name: ab_Files\Module03\Lab08\CustomerData.csv Browse...

Locale: English (United States) ☐ Unicode

Code page: 65001 (UTF-8)

Format: Delimited

Text qualifier: <none>

Header row delimiter: {CR}{LF}

Header rows to skip: 0

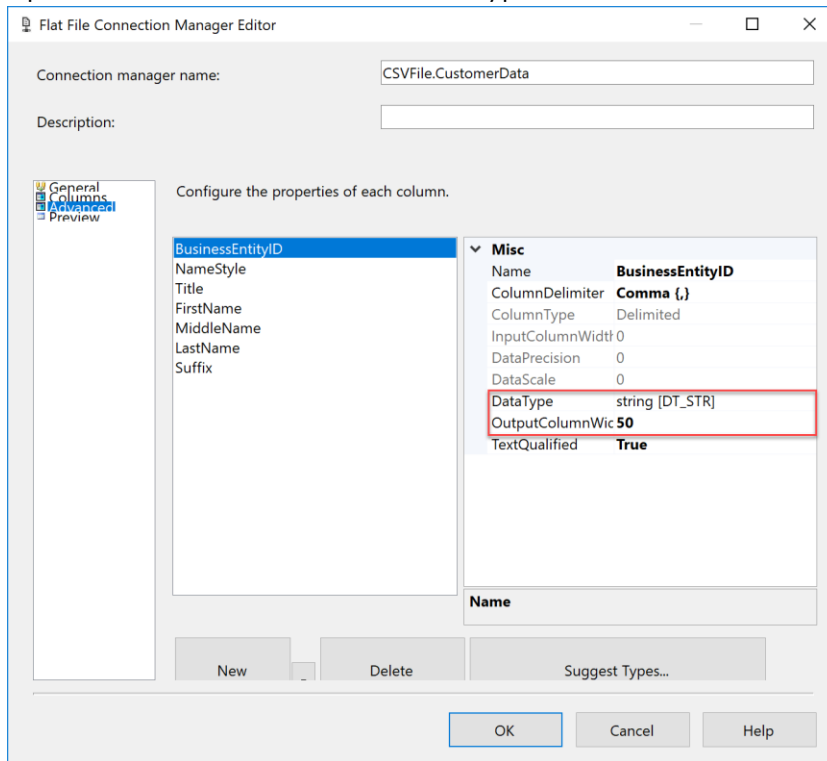
☒ Column names in the first data row

Columns are not defined for this connection manager.

OK Cancel Help

15. Next click on Advanced on the left side, to configure the columns.

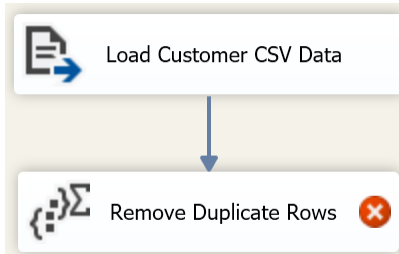
16. Update each column to make sure the Type and Width information matches the table below.



Column Name	Data Type	Output Column Width
BusinessEntityID	four-byte signed integer [DT_I4]	0
NameStyle	single-byte signed integer [DT_I1]	0
Title	Unicode string [DT_WSTR]	8
FirstName	Unicode string [DT_WSTR]	50
MiddleName	Unicode string [DT_WSTR]	50
LastName	Unicode string [DT_WSTR]	50
Suffix	Unicode string [DT_WSTR]	10

17. After updating click OK to save the connection manager. In the Flat File Source Editor, click on the Columns to update the column mapping and click OK.

18. Add a new task Aggregate and rename it to “Remove Duplicate Rows” and link it to “Load Customer CSV Data”.

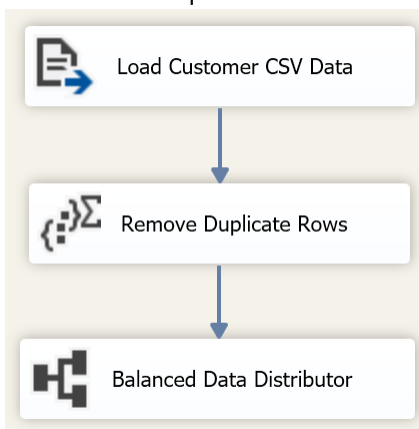


19. Double-click on Remove Duplicate Rows to configure. Select all the rows make sure option is set to group-by.

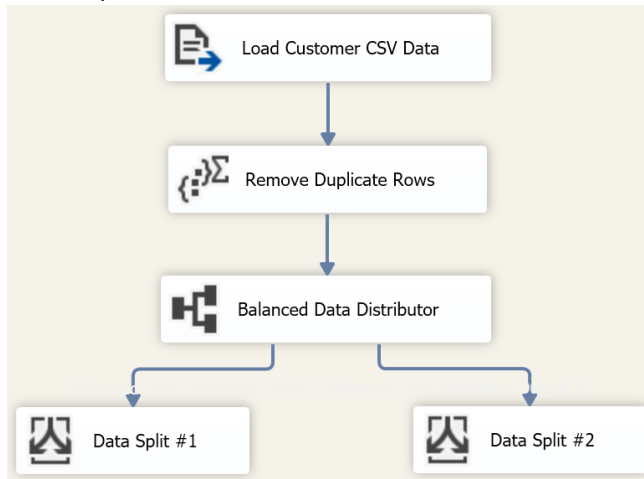
The screenshot shows the configuration window for the 'Remove Duplicate Rows' task. At the top, there is a list of 'Available Input Columns' with checkboxes next to each: Name, BusinessEntityID, NameStyle, Title, FirstName, MiddleName, LastName, and Suffix. Below this is a table with three columns: 'Input Column', 'Output Alias', and 'Operation'. The table lists the input columns and their corresponding output aliases, with the 'Operation' column set to 'Group by' for each.

Input Column	Output Alias	Operation
BusinessEntityID	BusinessEntityID	Group by
NameStyle	NameStyle	Group by
Title	Title	Group by
FirstName	FirstName	Group by
MiddleName	MiddleName	Group by
LastName	LastName	Group by
Suffix	Suffix	Group by

20. Next add Balance Data Distributor. Link it to “Remove Duplicate Rows” We are doing to help us divide and conquer the data.



21. Next add two Conditional Split tasks, rename first one to “Data Split #1” and second one to “Data Split #2”. For each task link it back to Balanced Data Distributor.



22. We need to configure EACH Data Split task, with following expressions and rules. Must be completed for BOTH Data Split #1 and Data Split #2.

Conditional Split Transformation Editor

Specify the conditions used to direct input rows to specific outputs. If an input row matches no condition, the row is directed to a default output.

Variables and Parameters

Columns

- BusinessEntityID
- NameStyle
- Title
- FirstName
- MiddleName
- LastName
- Suffix

Mathematical Functions

- String Functions
- Date/Time Functions
- NULL Functions
- Type Casts
- Operators

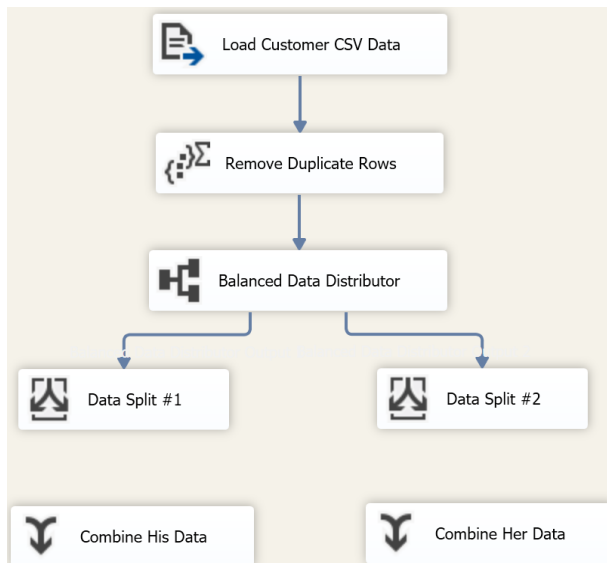
Description:

Order	Output Name	Condition
1	Mr	[Title] == "Mr"
2	Mrs	[Title] == "Mrs"

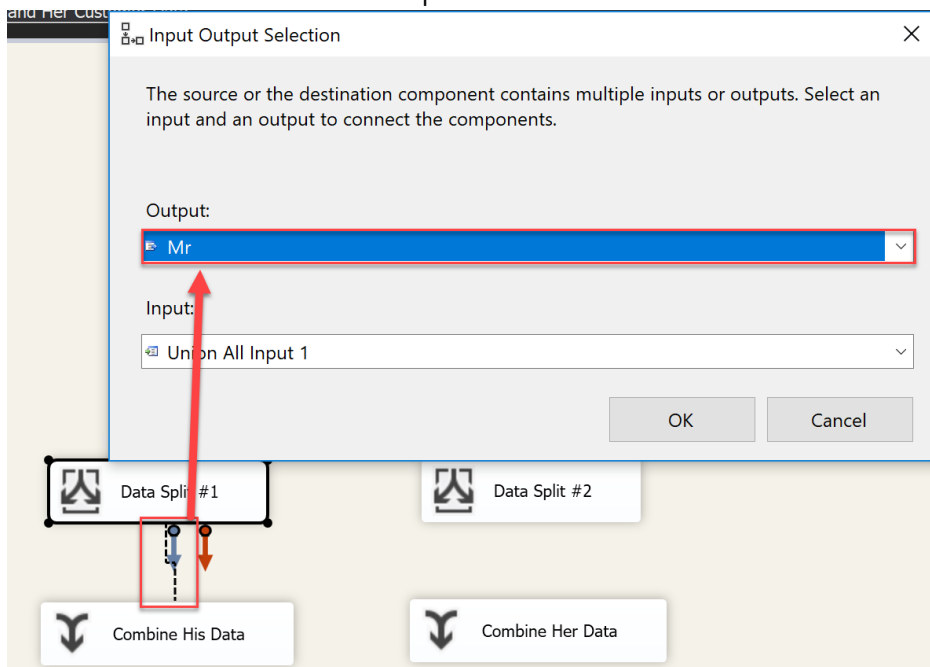
Default output name: Conditional Split Default Output

Configure Error Output... OK Cancel Help

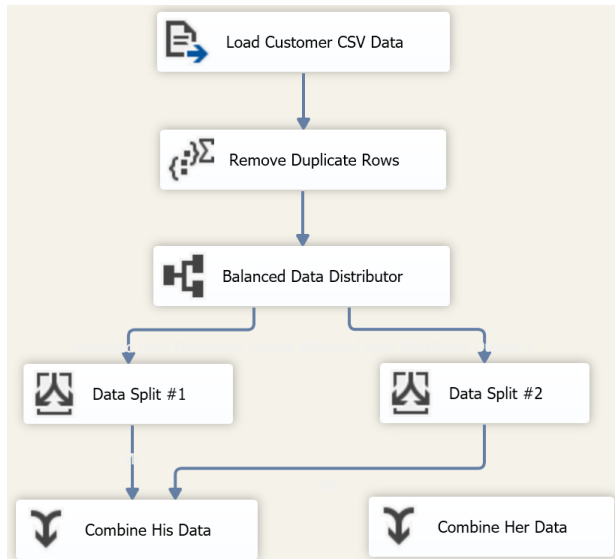
23. Create two new tasks for "Union All". Rename one to "Combine His Data" and other to "Combine Her Data".



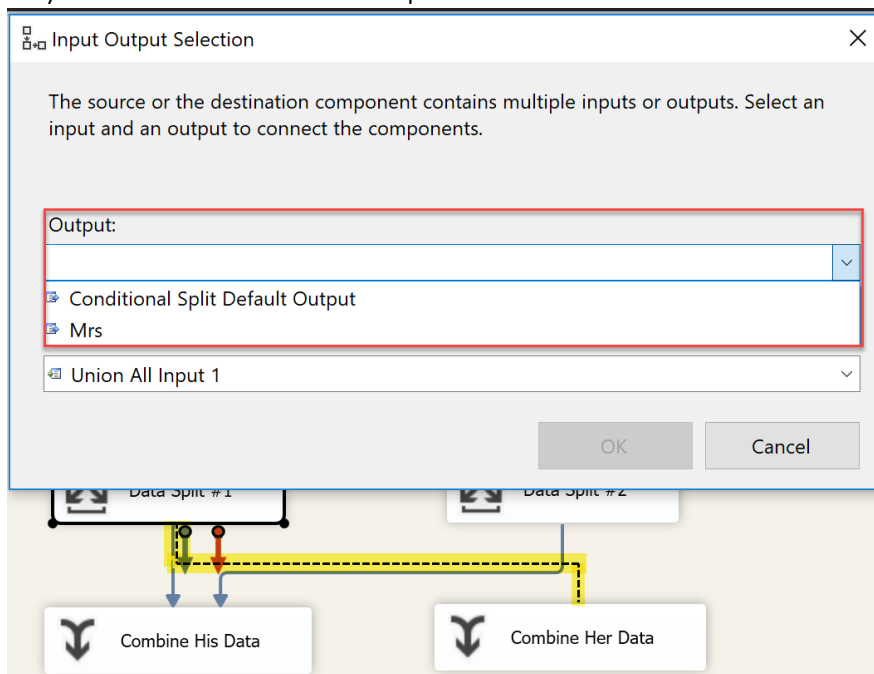
24. Drag an arrow from Data Split #1 to Combine His Data, when you do you'll get a below dialog box. Select "Mr" stream in the Output and click OK.



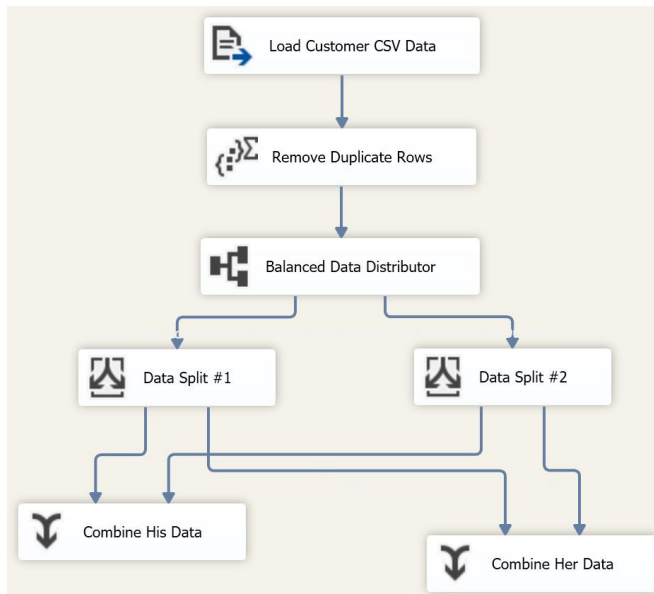
25. Drag an arrow from Data Split #2 and connecting to Combine Hist Data, again making sure to select "Mr" stream in output.



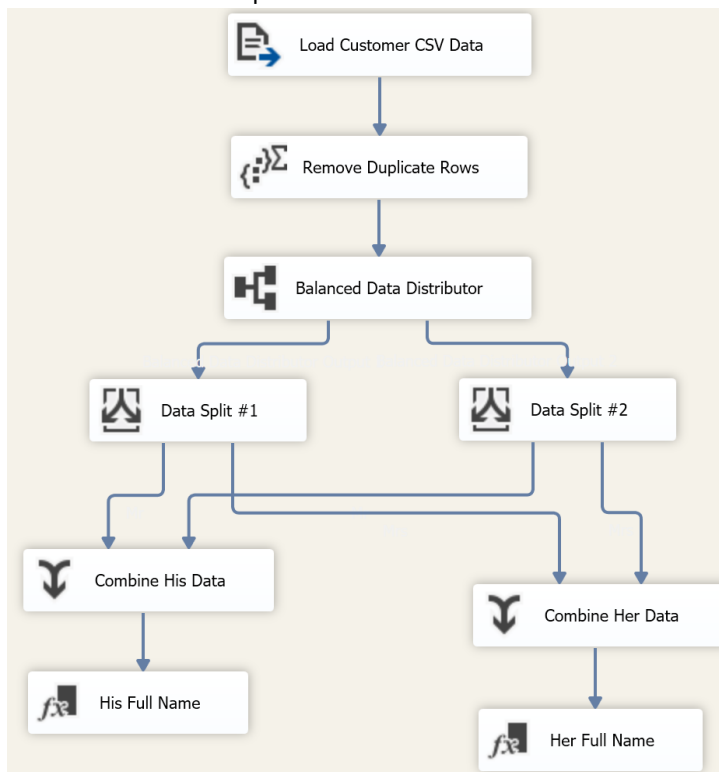
26. Do the same connection from Data Split #1 to Combine Her Data. This time notice dialog box only has "Her" stream left as an option.



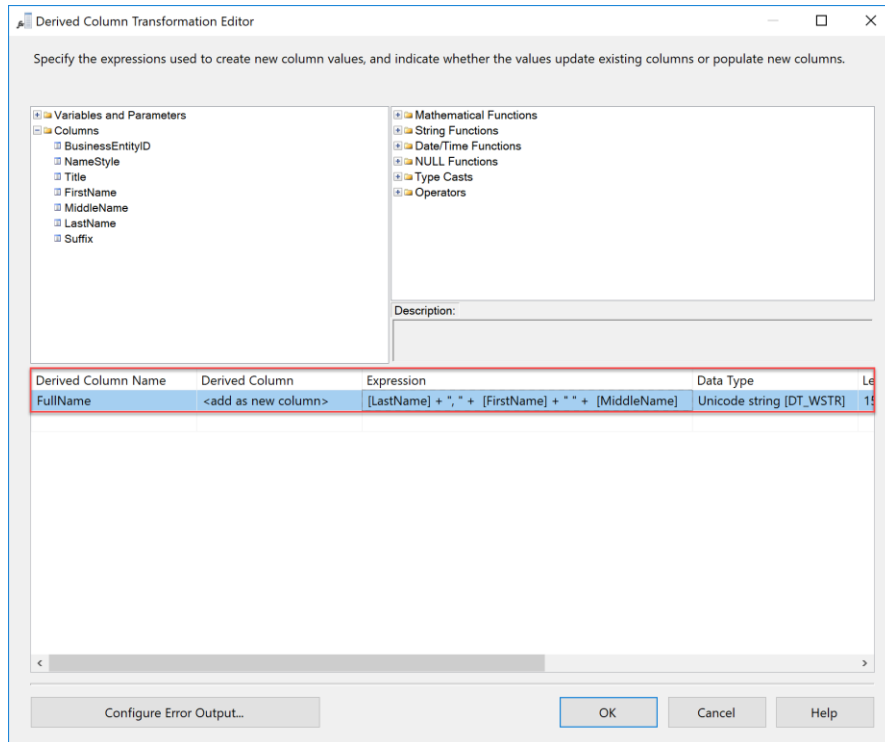
27. Next link the "Data Split #2" to "Combine Her Data", making sure to like "Mrs" stream in output.



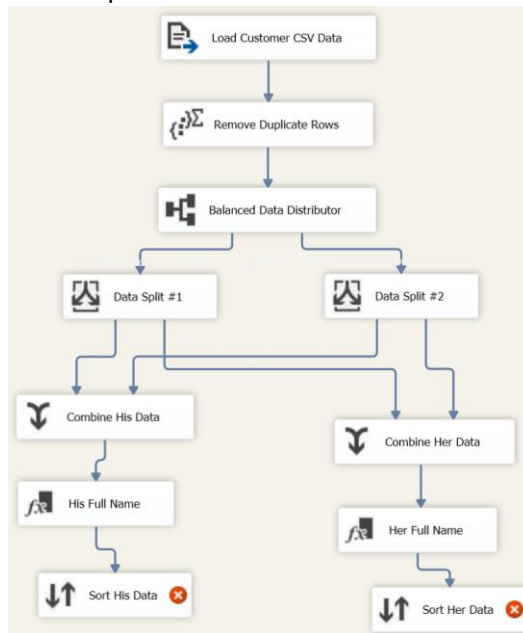
28. Add two “Derived Column” tasks. Rename one task to “His Full Name” another to “Her Full Name”. Link the respective tasks to their “Combine His Data” or “Combine Her Data”.



29. Double-click on “His Full Name”, build a new derived column with expression in screenshot below.



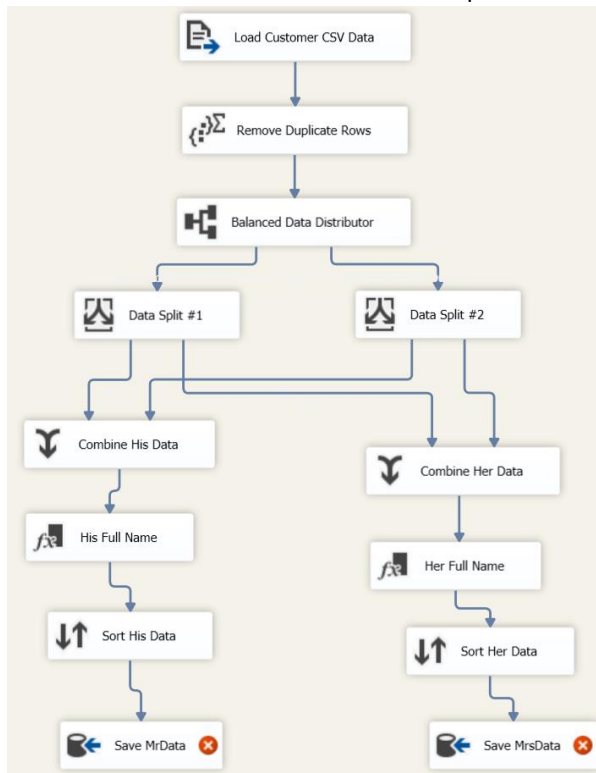
30. Create the same expression in “Her Full Name”.
31. Create two Sort Tasks, rename one to “Sort His Data” another to “Sort Her Data”. Link them to their respective derived column task.



32. Configure the sort and define the sort key as below for each task.

Input Column	Output Alias	Sort Type	Sort Order	Comparison
LastName	LastName	ascending	1	
FirstName	FirstName	ascending	2	

33. Create two new tasks for OLE DB Destination, rename one to “Save MrData” and other to “Save MrData”. Link them to their respective Sort tasks.



34. Double-click on "Save MrData". Connect it to connection string build in step #7 and change the table to "dbo.MrData". After connecting click on Column mapping to finalize mapping and click OK.

Connection Manager

Mappings

Error Output

3

Specify an OLE DB connection manager, a data source, or a data source view, and select the access mode. To use the OLE DB provider for Microsoft SQL Server, specify the data source as the name of the server. To use the OLE DB provider for Oracle, specify the data source as the name of the database. To use the OLE DB provider for ODBC, specify the data source as the name of the DSN. To use the OLE DB provider for Microsoft Access, specify the data source as the path to the database. To use the OLE DB provider for Microsoft Jet, specify the data source as the path to the database. To use the OLE DB provider for Microsoft Excel, specify the data source as the path to the spreadsheet. To use the OLE DB provider for Microsoft Visual Basic for Applications, specify the data source as the path to the spreadsheet. To use the OLE DB provider for Microsoft Access, specify the data source as the path to the database. To use the OLE DB provider for Microsoft Jet, specify the data source as the path to the database. To use the OLE DB provider for Microsoft Excel, specify the data source as the path to the spreadsheet. To use the OLE DB provider for Microsoft Visual Basic for Applications, specify the data source as the path to the spreadsheet.

OLE DB connection manager:

MOGUPTA-PC02.AdventureWorks2012 1

Data access mode:

Table or view - fast load

Name of the table or the view:

[dbo].[MrData] 2

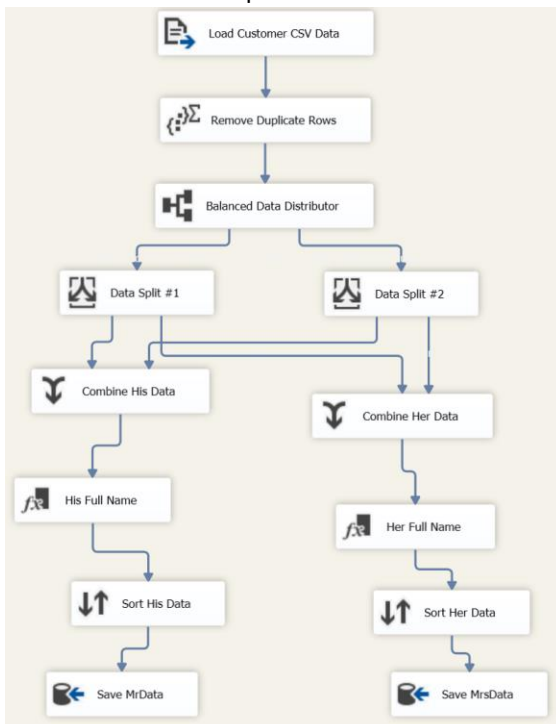
☐ Keep identity ☒ Table lock

☐ Keep nulls ☒ Check constraints

Rows per batch:

Maximum insert commit size:

35. Follow the same steps for "Save MrsData". Connecting to "dbo.MrsData" table.



36. Execute task. Go to SQL Server and verify data in each table by executing "SELECT * FROM dbo.MrData" and "SELECT * FROM dbo.MrsData".