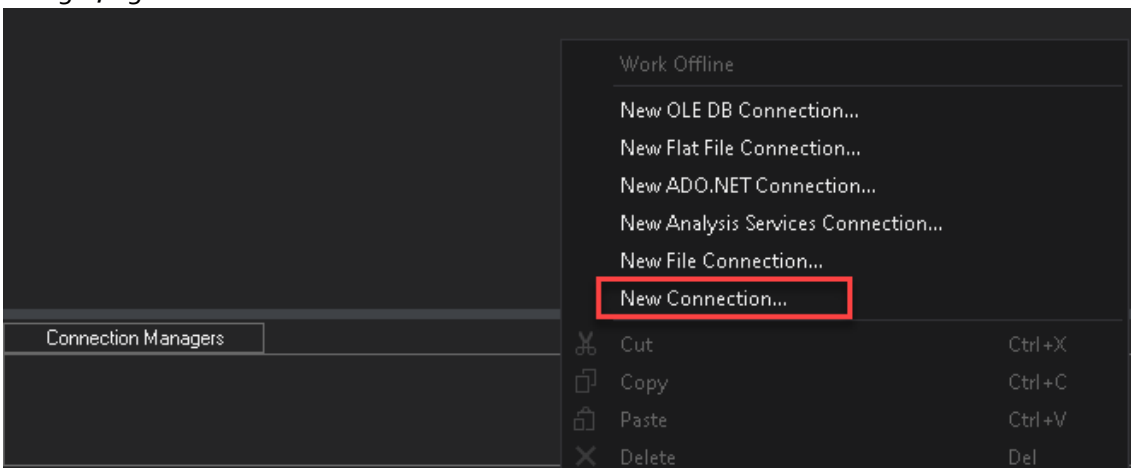
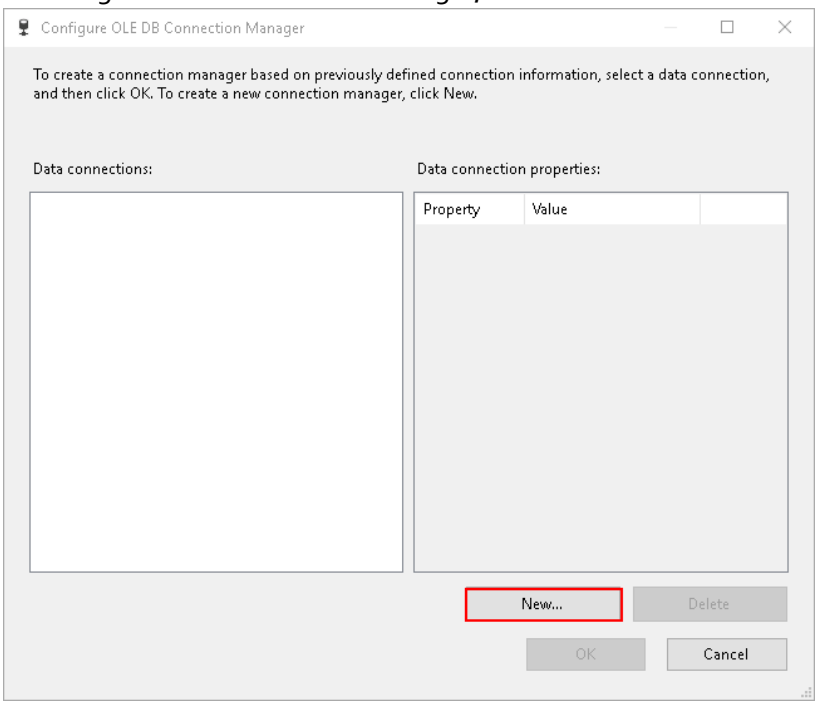
SQL SERVER INTEGRATION SERVICES

MODULE 04 – LAB 01: ERROR HANDLING IN DATA FLOW

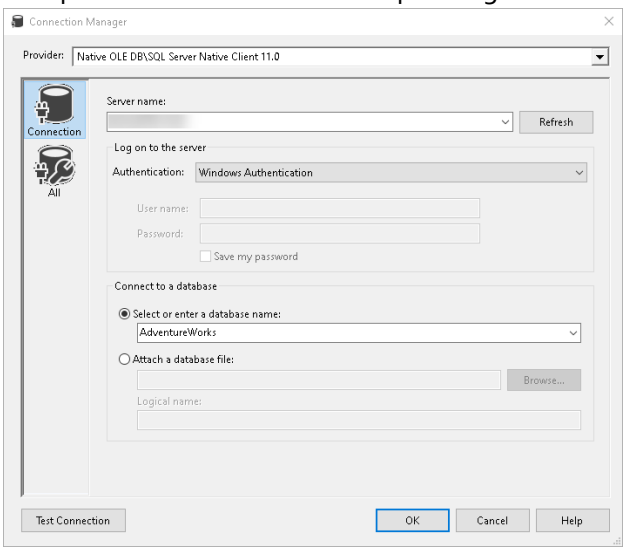
1. In the SQL Server Management Studio, connect to AdventureWorks database.
2. Run the SQL Server script to Create EmployeeData table (Create_EmployeeData.sql).
3. Launch Visual Studio 2019.
4. Create new integration services project.
5. Set up a connection manager for the database. In the bottom center pane under Connection Manger, right-click select New Ole-DB Connection.



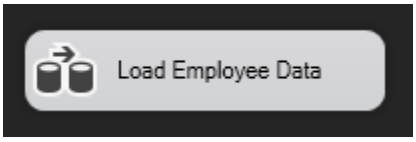
6. In Configure OLE DB Connect Manager, click New.



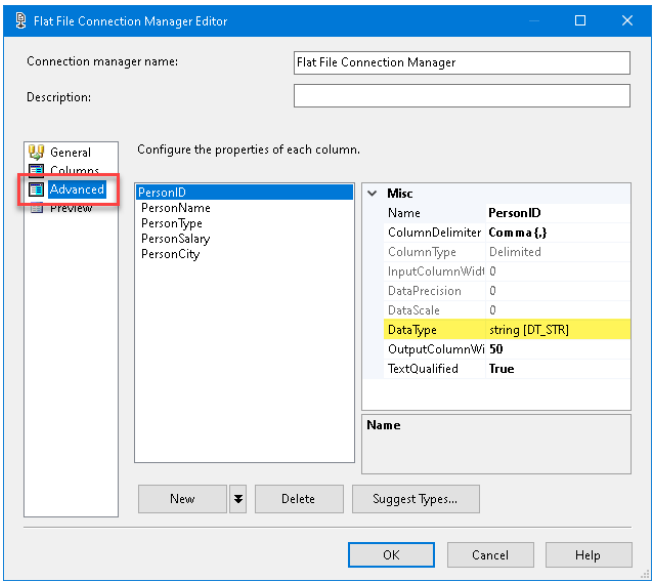
7. Setup connection to SQL Server pointing to the AdventureWorks database.



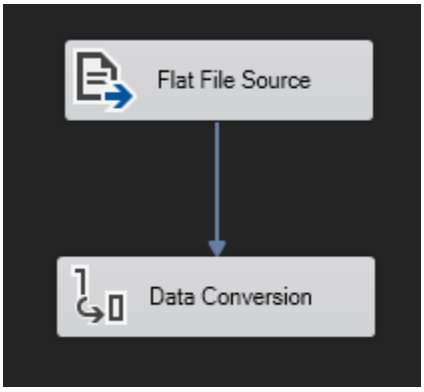
8. Next, create a new Data Flow Task, name it, Load Employee Data.



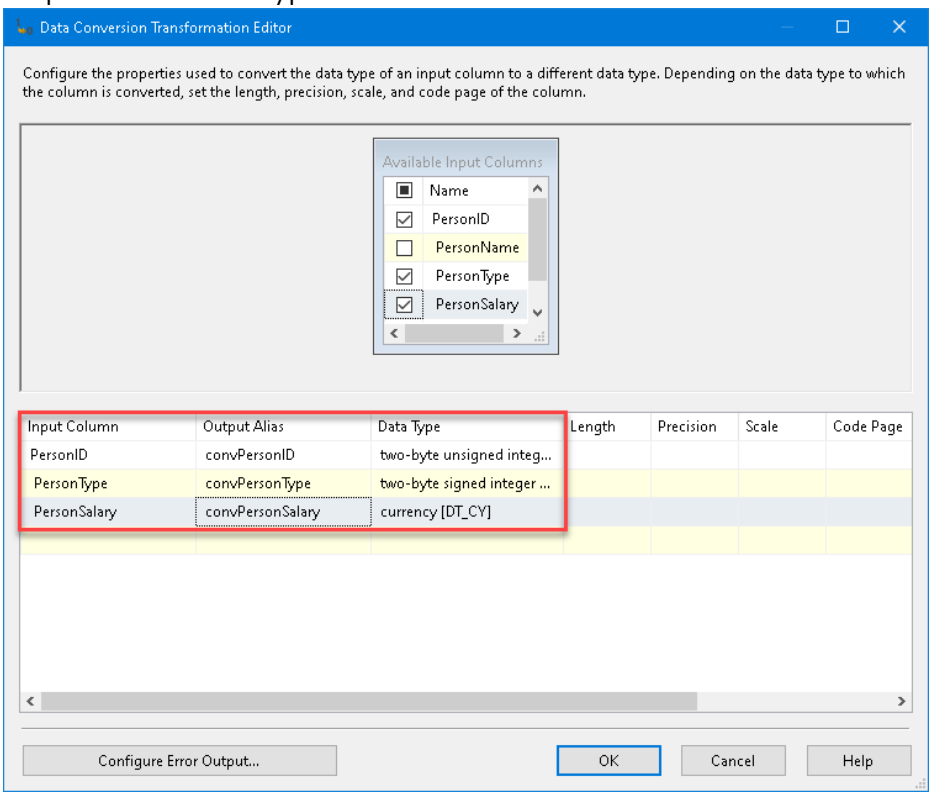
9. Double click on Data Flow Task, to switch to data flow designer.
10. Add a Flat-File Source, point it to the CSV file (PeopleData.csv). Review the default data type selected under Advanced, review DataType field. All fields' data type default to string [DT_STR].



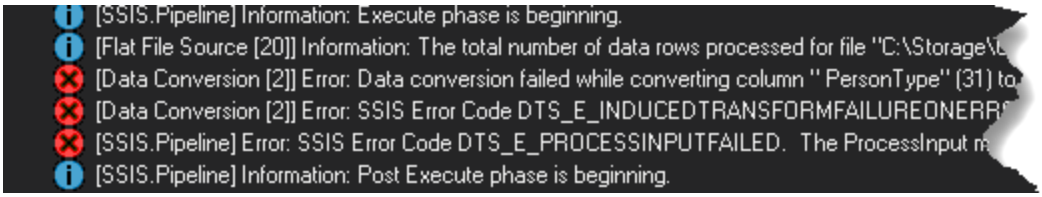
11. The default types are not correct. To correct this, add the Data Conversion step.



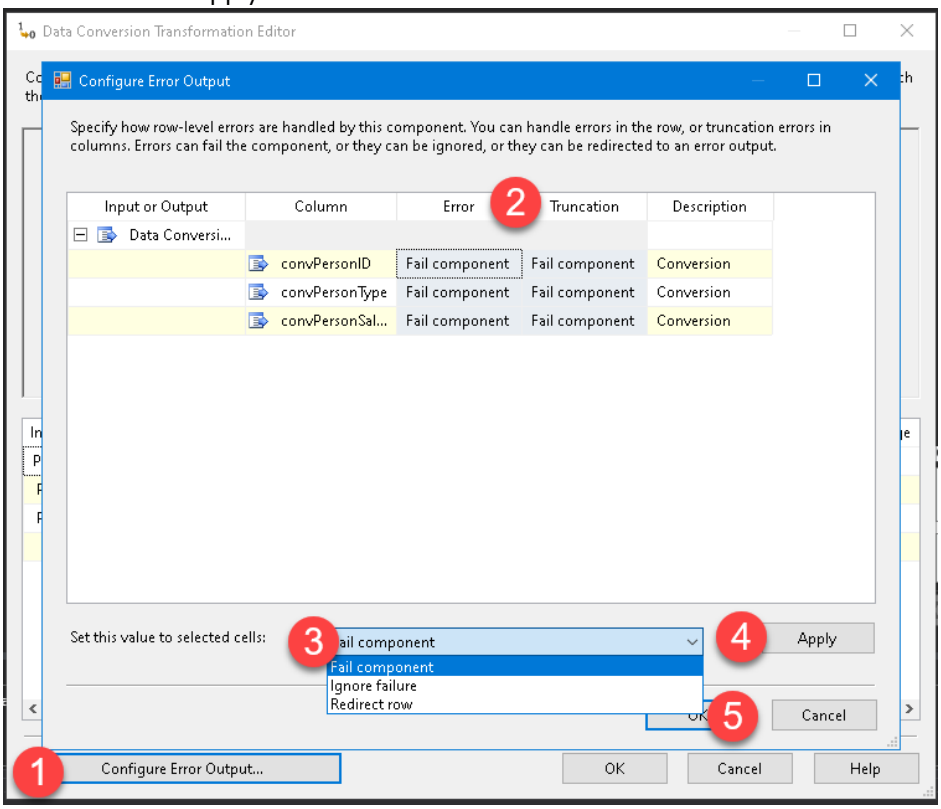
12. Create a mapping for fields that need the data type updated (i.e., fields that do not need DT_STR data type), double click on the data conversion task, and create mapping. Update the output alias and data type.



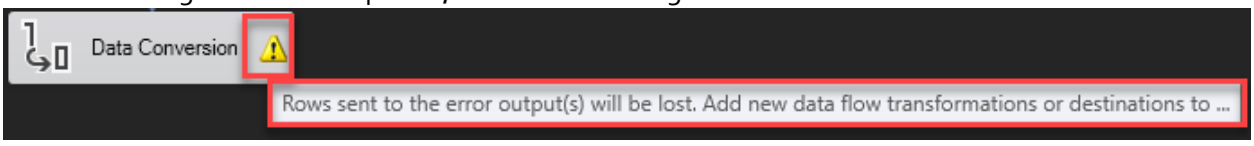
13. Try running the package now.
14. The data conversion fails, why? Click on the Progress tab and review the output. Look for errors like below, review and understand the errors.



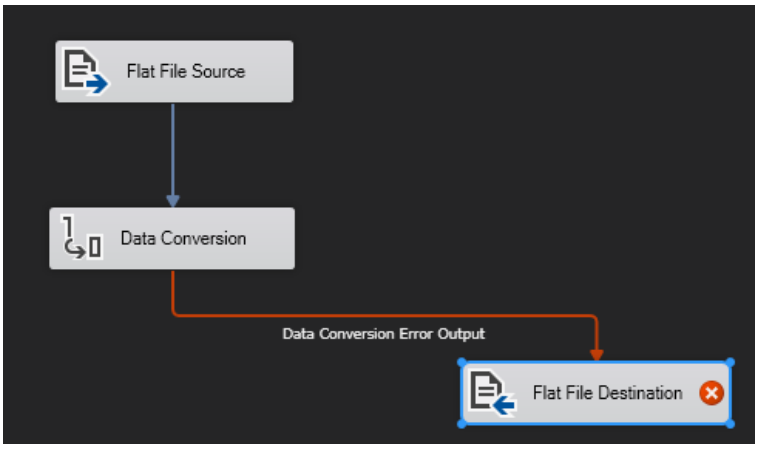
15. The type conversion failed. This is ***expected*** because the data is not clean. The error does not clearly identify the errored rows. Let's build an understanding of which rows failed. To do this, next capture the rows which have failed.
16. Go back to data conversion task in the data flow and click on Configure Error Output. Take note that default behavior is to "Fail component". This makes it difficult to troubleshoot. Change the default behavior for both Error and Truncation (select all the cells), select "Redirect rows" and click Apply.



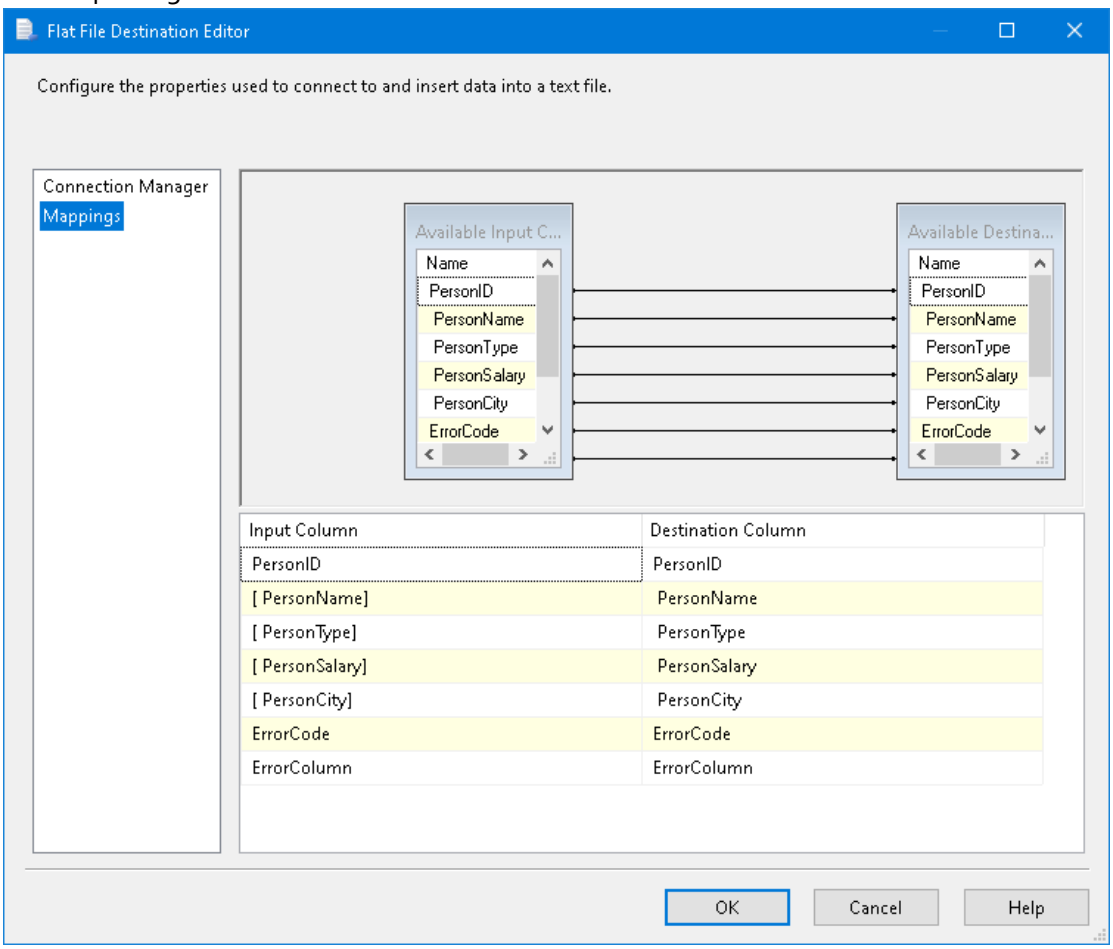
17. After the configuration is completed, notice the warning on the Data Conversion.



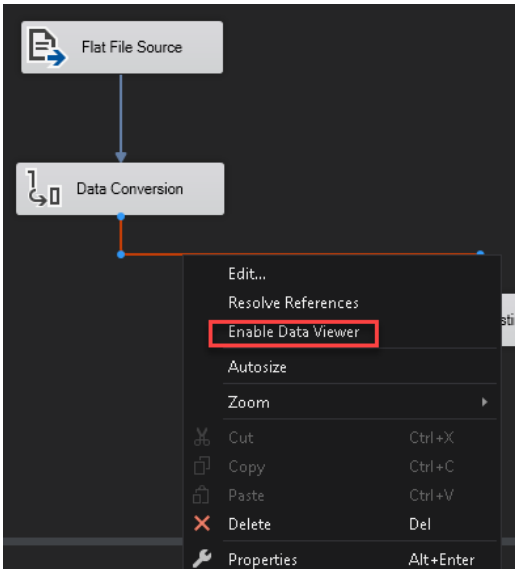
18. Add a flat file destination to redirect error rows and connect it to Data Conversion error output (red arrow).



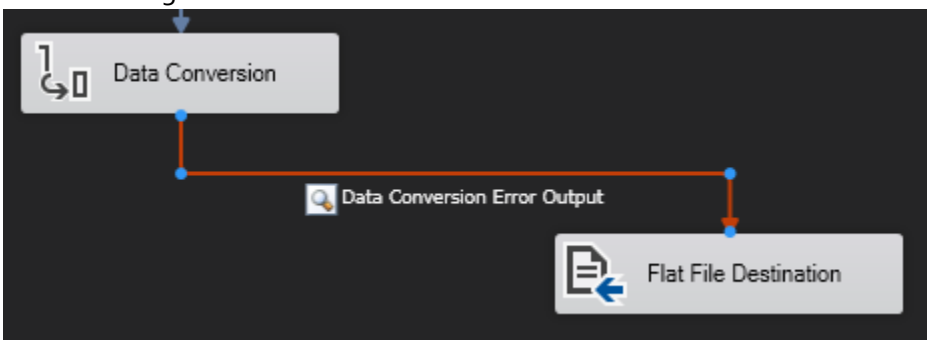
19. Configure the flat file destination to new file, ErrorData.csv. After creating new data source click on Mapping in "Flat File Destination Editor". New fields are introduced in the output for error reporting.



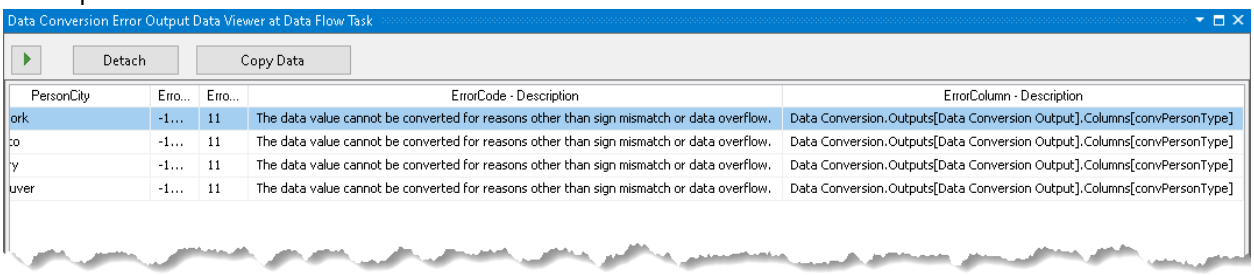
20. Try out the package and review the ErrorData.csv file to failed records. The csv gives the error code and column but doesn't provide the description. The column number is the internal ID value assigned to the column and not the actual position in data. The CSV data let us know which rows failed. To understand the error and columns go back to Data Flow designer. Right click on the Error Output, Enable Data Viewer, and run the package again.



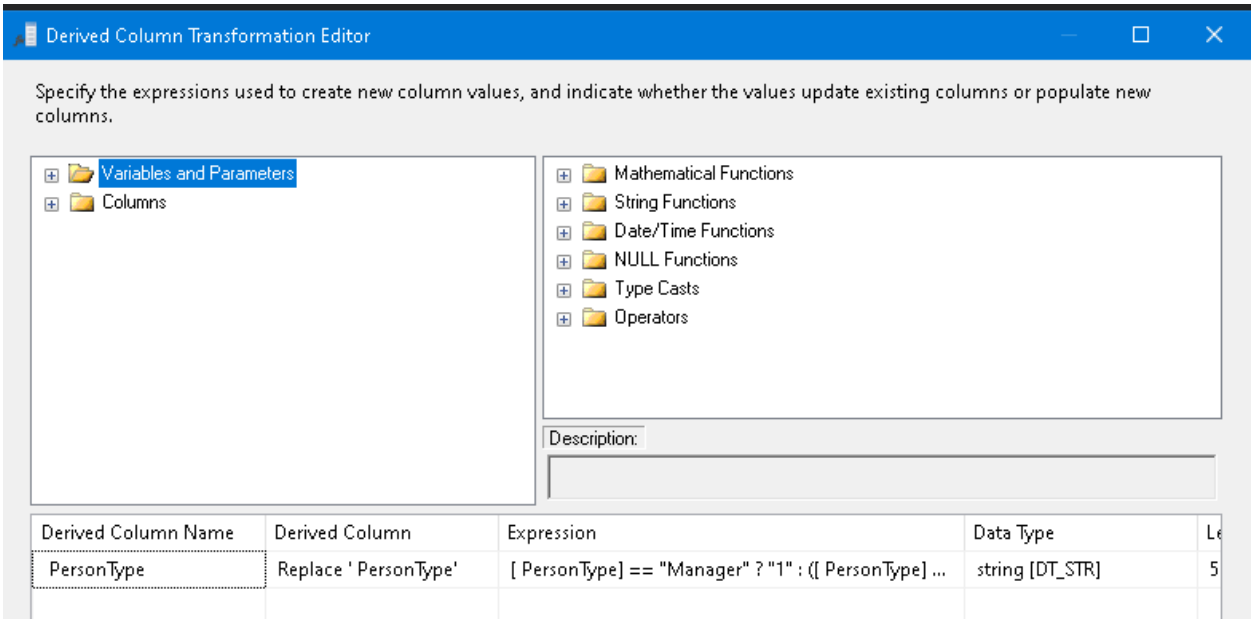
After enabling:



21. After running a new output window will popup, scroll to right to review the "ErrorColumn – Description" column.

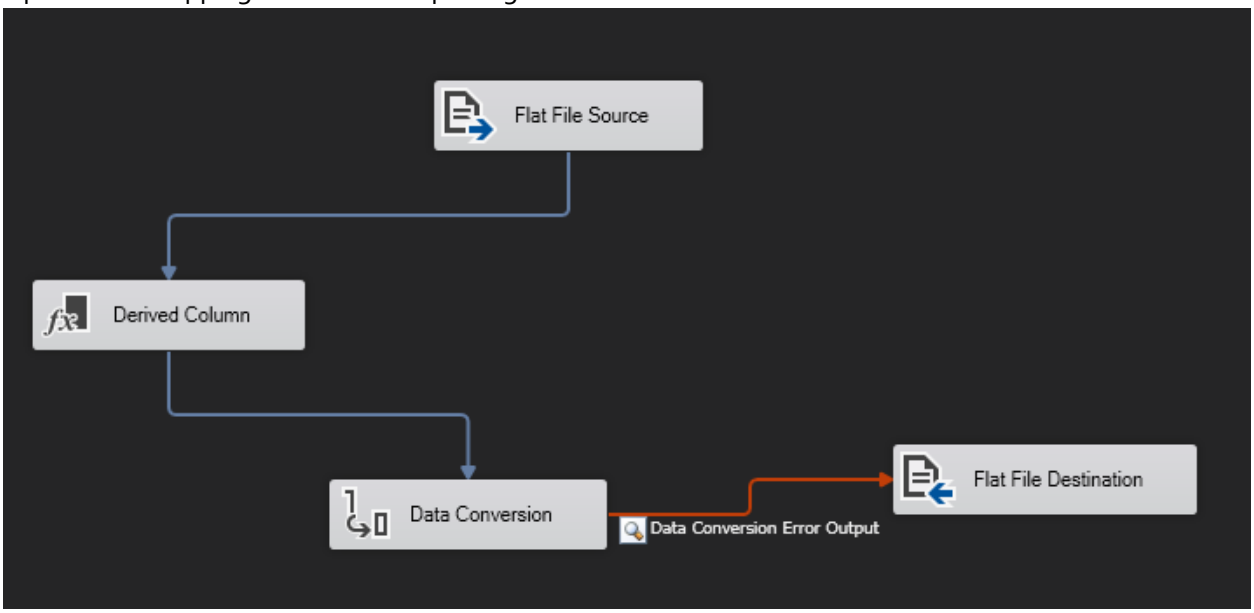


22. The conversion issue is with PersonType column. Looking at the SQL table created in step #1, it is expecting an integer. Therefore, some pre-clean up steps need to be applied before the data type conversion.
23. Add a new Drive Column task for cleaning up the PersonType column. Add the following formula to correct the field from description values to id values.



Expression: [PersonType] == "Manager" ? "1" : ([PersonType] == "Employee" ? "2" : [PersonType])

24. Update the mapping and run the export again.



25. The Data Explorer dialog will show there are still issues with our data. This time it is with Person Salary column. Review the data and see how you might consider fixing it?
26. After reviewing the data notice the salary value is in English. There is no way to protect or pre-process the data for this scenario. So, either correct the data in CSV or manage these rows manually.