Objective: We have a csv file containing following columns

EmpId, Name, HireDate, Phone No and Salary

We need to perform the following validations on the data

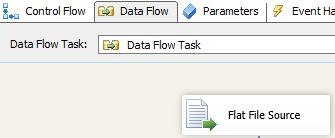
1. Filter the data if
   1. EmpId, Name, HireDate is null
   2. Salary is nonnumeric
   3. Phone No. length is not equal to 10

File:

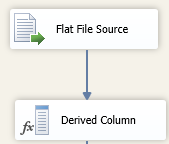


Steps:

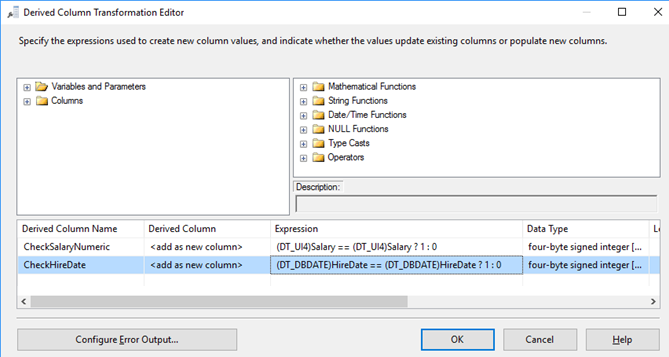
1. Create a new package
2. Add a Data Flow Task
3. In Data Flow Task, Add a Flat File Source



1. Add a derived column, where-in we will check for datatypes

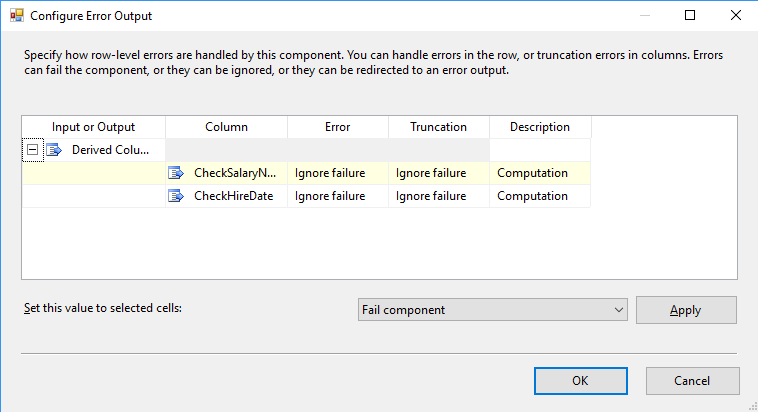


1. Right click on Derived Column, and add the following columns
   1. CheckSalaryNumeric, with expression: (DT\_UI4)Salary == (DT\_UI4)Salary ? 1 : 0
   2. CheckHireDate, with expression: (DT\_DBDATE)HireDate == (DT\_DBDATE)HireDate ? 1 : 0



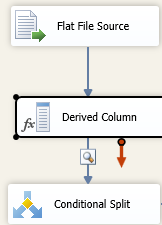
Now, incase it contains any invalid data, then it will throw the error, so we need to configure the output

Click on Configure Error Output,



and change the behavior on Error or Truncation to ignore failure and click on OK button

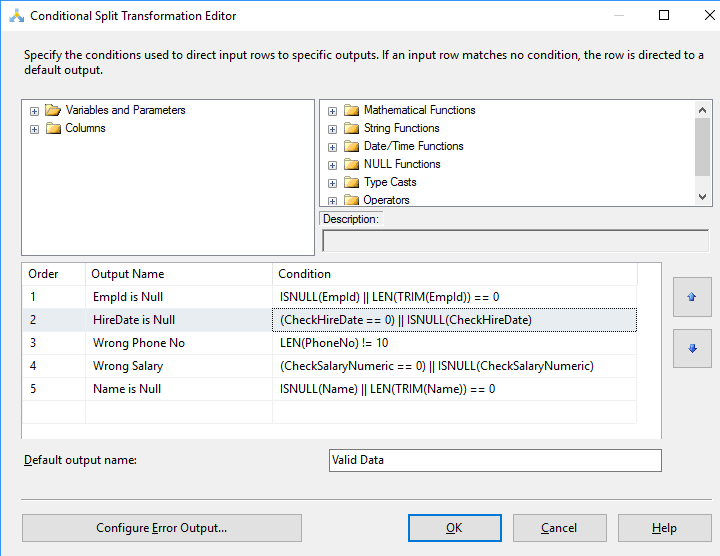
1. Now, add the transformation Conditional Split



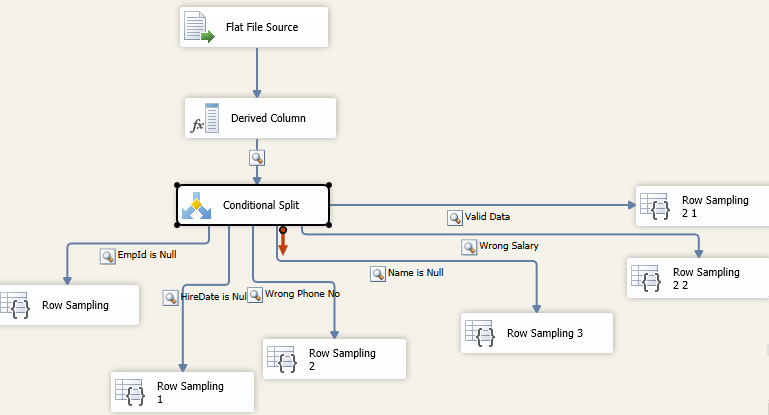
And add the following outputs

|  |  |  |
| --- | --- | --- |
| S.No | Output Name | Condition |
| 1 | EmpId is Null | ISNULL(EmpId) || LEN(TRIM(EmpId)) == 0 |
| 2 | HireDate is Null | (CheckHireDate == 0) || ISNULL(CheckHireDate) |
| 3 | Wrong Phone No | LEN(PhoneNo) != 10 |
| 4 | Wrong Salary | (CheckSalaryNumeric == 0) || ISNULL(CheckSalaryNumeric) |
| 5 | Name is Null | ISNULL(Name) || LEN(TRIM(Name)) == 0 |

And give the Default Output Name: Valid Data



Now join them with respective Row Sampling Components, and add dataviewer to precedence of each each row sampling component



Now if we run this, we will get the respective data directed to respective Row Sampling components