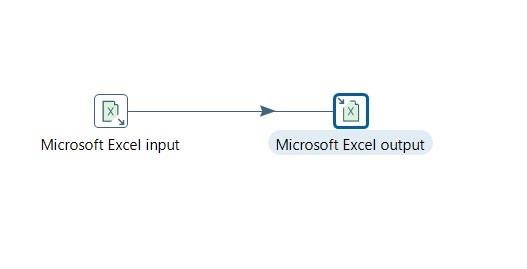
**Pentaho Assignment**

**● Assignment on input/output:**

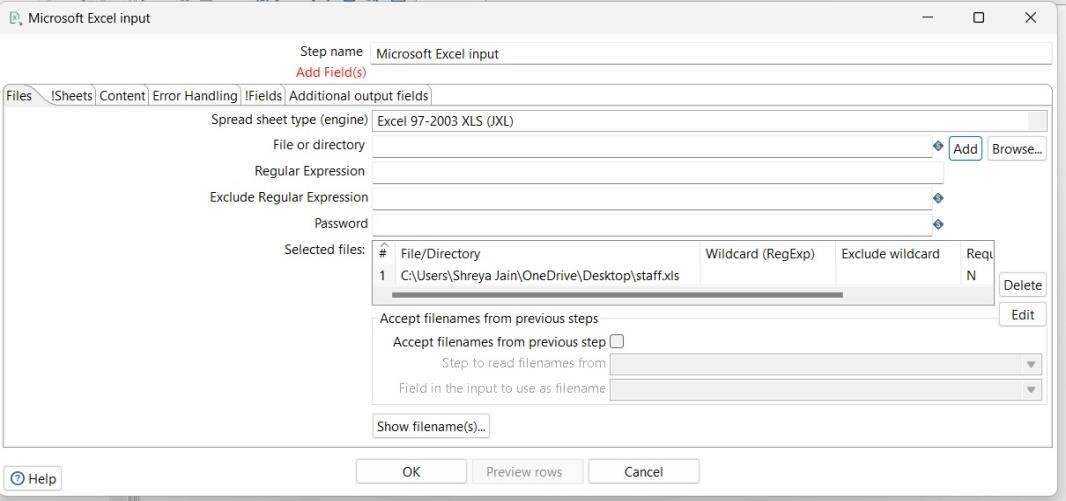
**1) Retrieving and load Data from a Flat File (Excel) to data warehouse**

**Steps:**

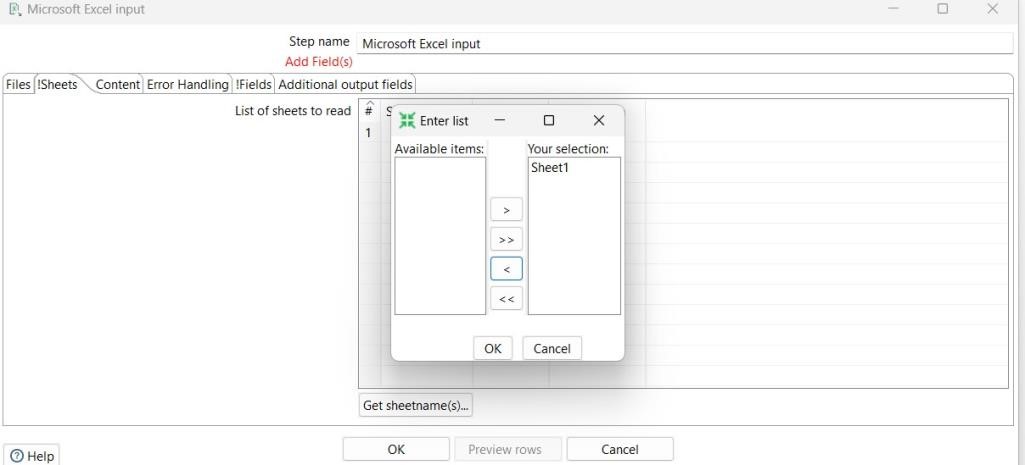
1. Open Data integration folder and launch the Pentaho Data Integration software by clicking on Window Batch File.
2. Click on File > New Transformations.
3. An Input folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel input file into the transformation area
4. Similarly an Output folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel output file into the transformation area.
5. Now create a new hop connection between the two steps by clicking on the output connector from the input excel file.



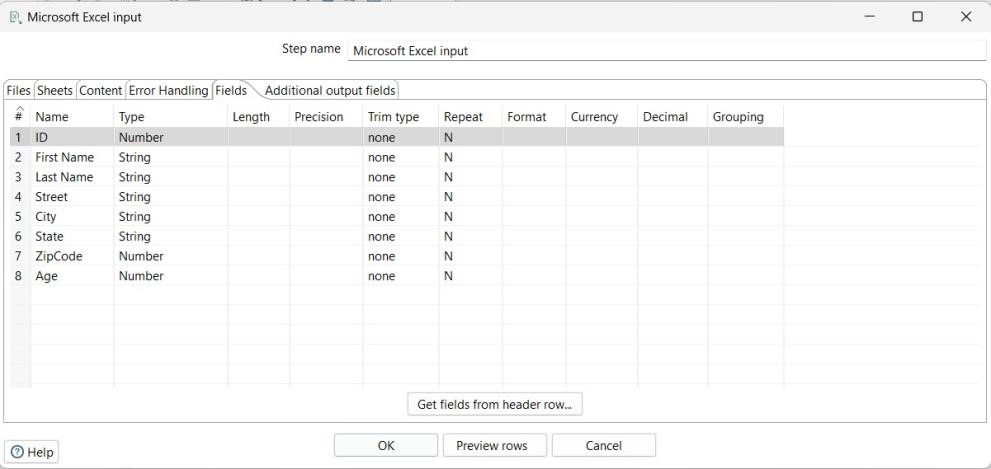
1. Double click on Microsoft excel input file and Browse the excel file from which you need to perform all operations.



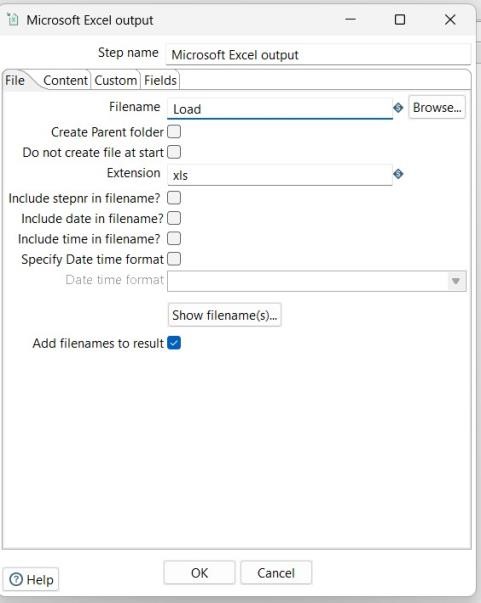
1. Select the sheet in which your data is stored.



1. Now get all the fields from that sheet by clicking on the Get fields from the header row.

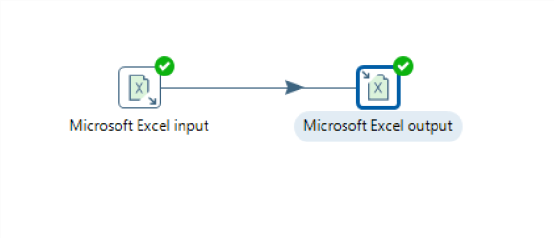


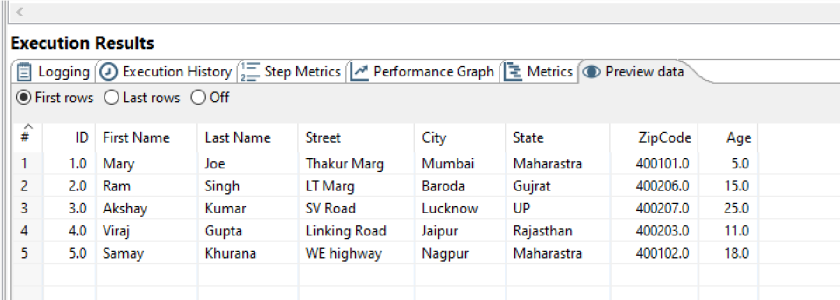
1. Select OK , Your data is now retrieved from the the excel file.
2. Now to load the data in another file, click on the Microsoft Excel Output file and save it by giving a suitable filename and click OK.



1. Now run your project and you will see the following output.

**Output:**



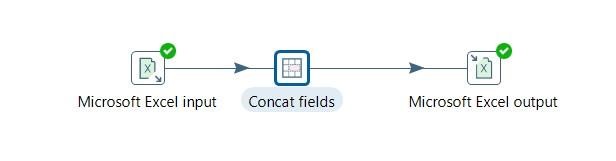


**● Assignment on transformation:**

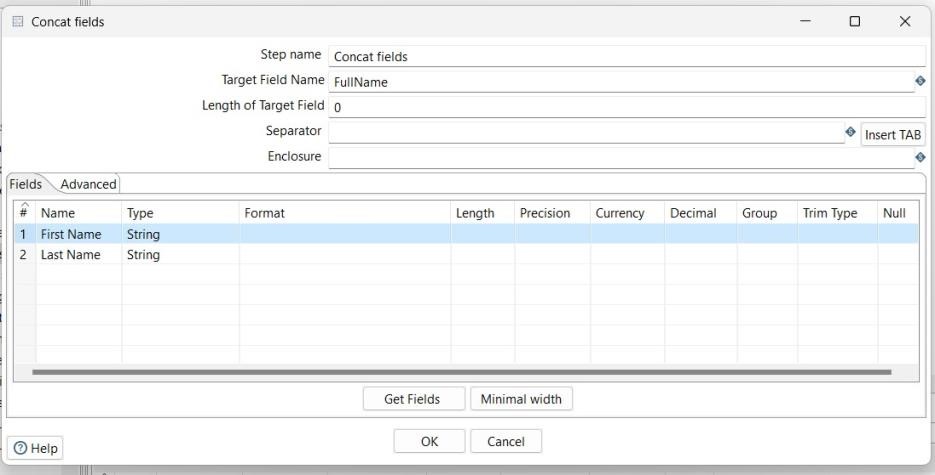
**1) Retrieve data from file staff.accdb and concat two columns firstname and lastname as fullname of teacher table and load data to data warehouse**.

**Steps:**

1. Open Data integration folder and launch the Pentaho Data Integration software by clicking on Window Batch File.
2. Click on File > New Transformations.
3. An Input folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel input file into the transformation area
4. Similarly an Output folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel output file into the transformation area.
5. From the Transform folder Drag and drop **Concat fields** into the transformation area.
6. Now create a new hop connection between all the three steps by clicking on the output connector from the input excel file.

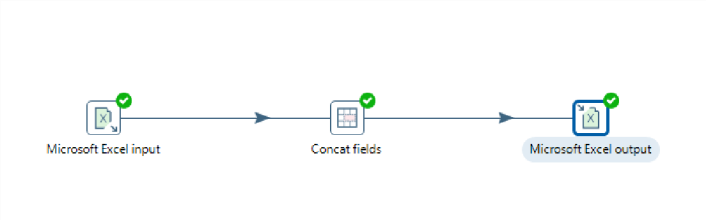


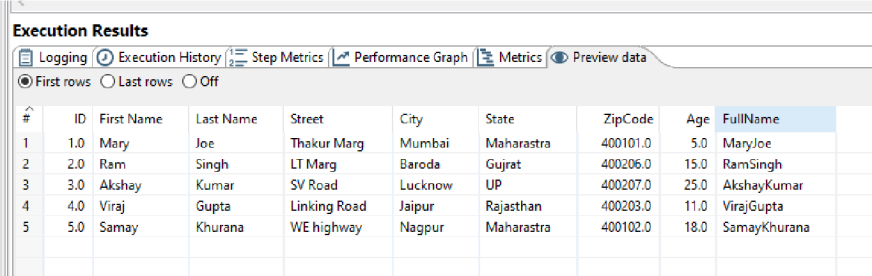
1. Double click on the concat field and click on get fields and delete from the list any fields you do want to contact, leaving only first name.



1. Click on the Microsoft Excel Output file and save it by giving a suitable filename and click OK.
2. Now run your project and you will see the following output.

**Output:**

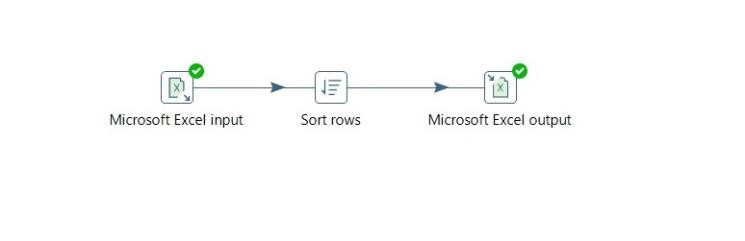




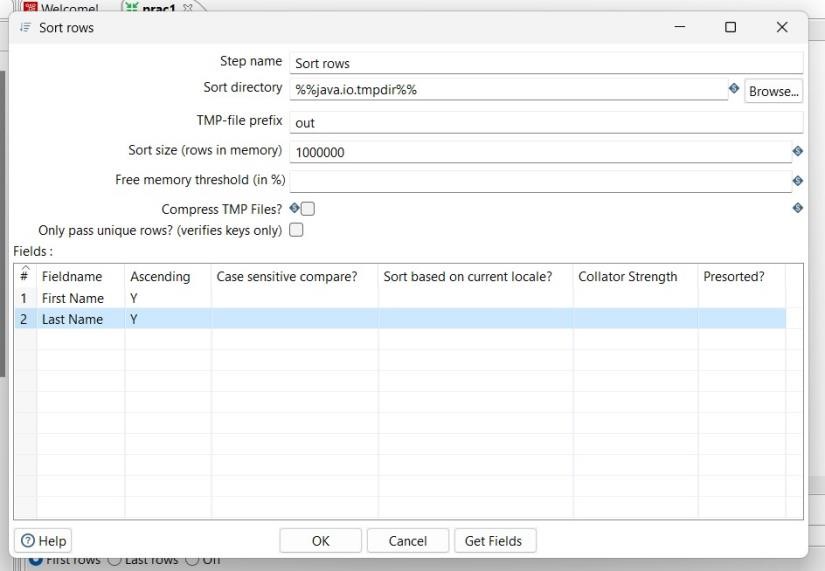
1. **Sort column fullname in descending order and load data to data warehouse.**

**Steps:**

* 1. Open Data integration folder and launch the Pentaho Data Integration software by clicking on Window Batch File.
  2. Click on File > New Transformations.
  3. An Input folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel input file into the transformation area
  4. Similarly an Output folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel output file into the transformation area.
  5. From the Transform folder Drag and drop **Sort rows** into the transformation area.
  6. Now create a new hop connection between all the three steps by clicking on the output connector from the input excel file.

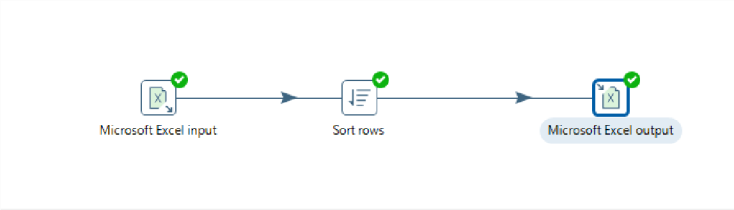


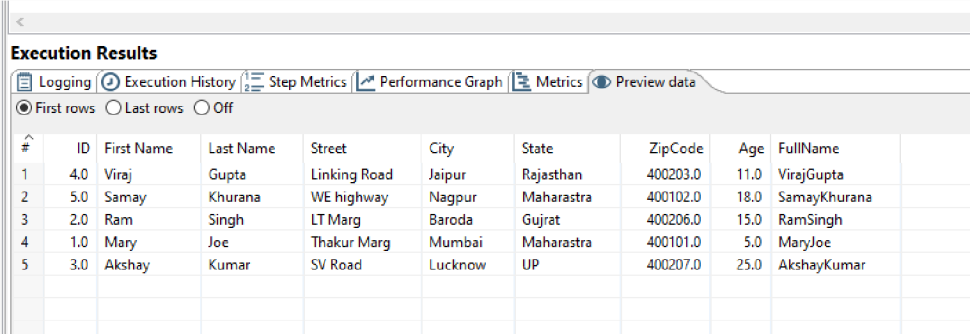
* 1. Double click on the sort row and click on get fields and delete from the list any fields you do want to contact, leaving only first name.



* 1. Click on the Microsoft Excel Output file and save it by giving a suitable filename and click OK.
  2. Now run your project and you will see the following output.

**Output:**

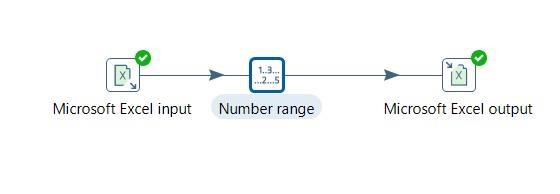




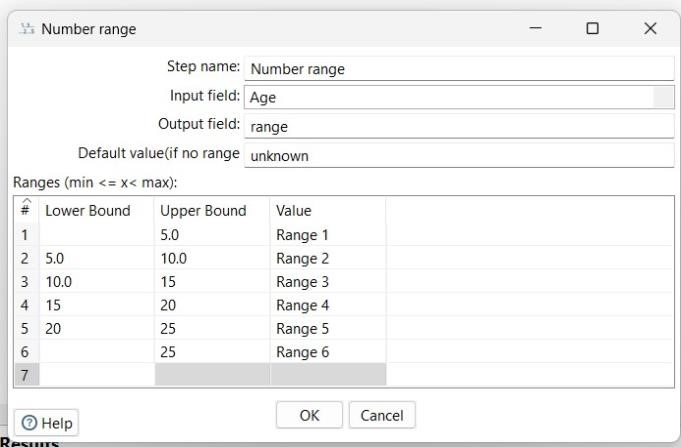
1. **Retrieve data from file and descritize age column of class table with <5,** **5-10,10-15,15-20, 20-25, > 25 and load data to data warehouse**

**Steps:**

1. Open Data integration folder and launch the Pentaho Data Integration software by clicking on Window Batch File.
2. Click on File > New Transformations.
3. An Input folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel input file into the transformation area
4. Similarly an Output folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel output file into the transformation area.
5. From the Transform folder Drag and drop **Number range** into the transformation area.
6. Now create a new hop connection between all the three steps by clicking on the output connector from the input excel file.

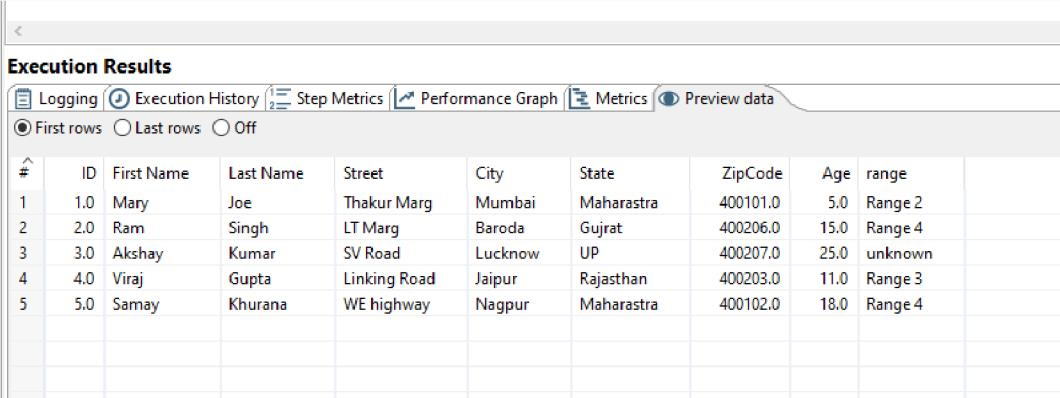
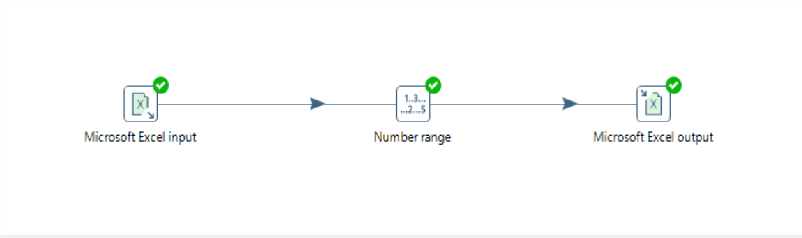


1. Double click on Number range and choose the input field as age.Either use default ranges or modify as per requirement.



1. Click on the Microsoft Excel Output file and save it by giving a suitable filename and click OK.
2. Now run your project and you will see the following output.

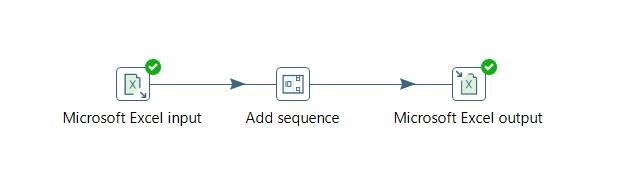
**Output:**



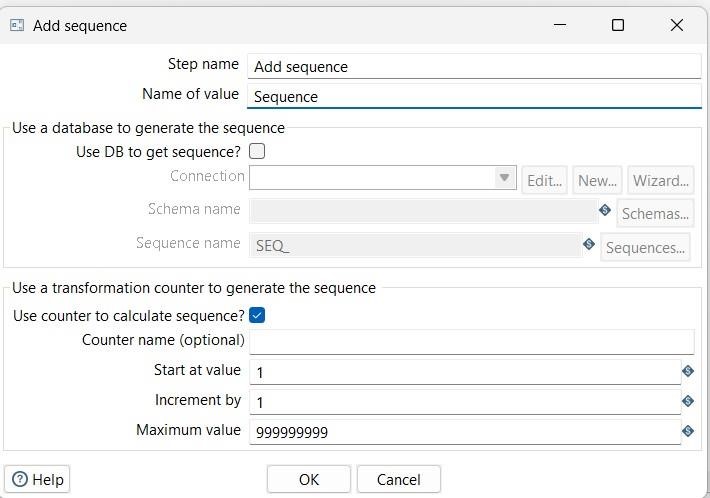
**4)Retrieve data from file and add sequence in the staff table.**

**Steps:**

* 1. Open Data integration folder and launch the Pentaho Data Integration software by clicking on Window Batch File.
  2. Click on File > New Transformations.
  3. An Input folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel input file into the transformation area
  4. Similarly an Output folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel output file into the transformation area.
  5. From the Transform folder Drag and drop **Add Sequence** into the transformation area.
  6. Now create a new hop connection between all the three steps by clicking on the output connector from the input excel file.

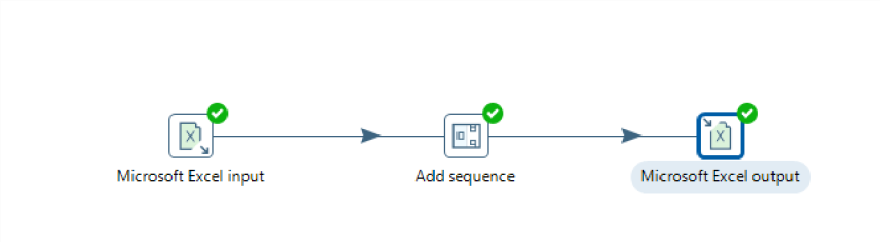


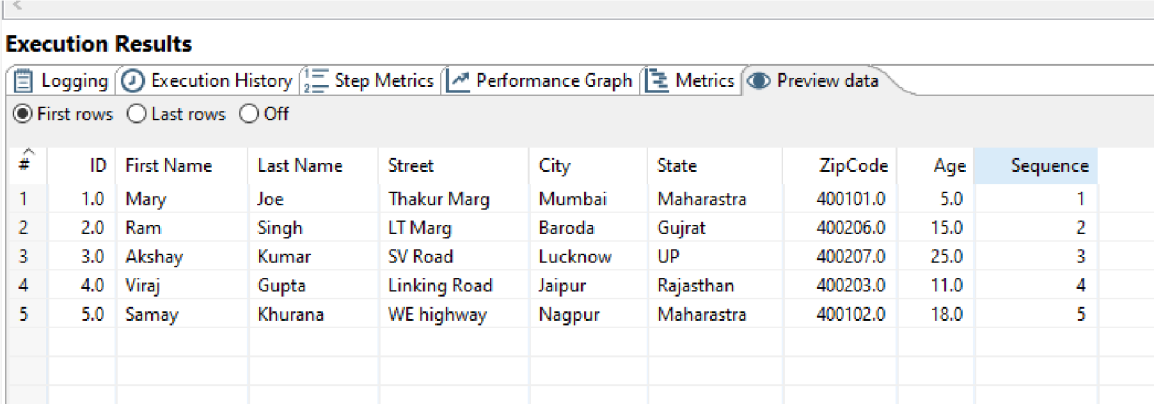
* 1. Double click on Add Sequence and specify “Name of value” for the field name you would want to give to the sequence.



* 1. Click on the Microsoft Excel Output file and save it by giving a suitable filename and click OK.
  2. Now run your project and you will see the following output.

**Output:**



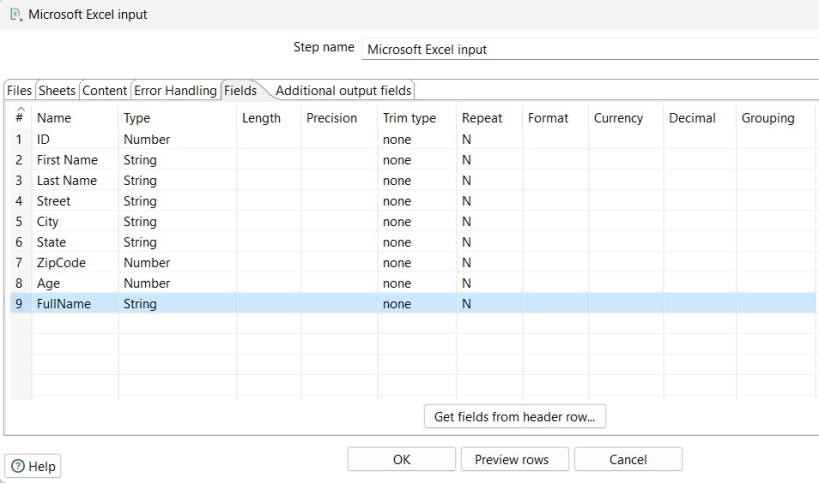


1. **Split fullName column into two columns Fname and Lname using staff table**

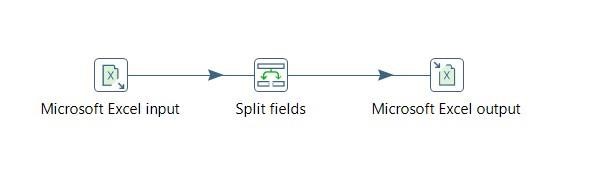
**Steps:**

* 1. Open Data integration folder and launch the Pentaho Data Integration software by clicking on Window Batch File.
  2. Click on File > New Transformations.
  3. An Input folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel input file into the transformation area.
  4. Add the concat excel file which contains FullName column added to it.

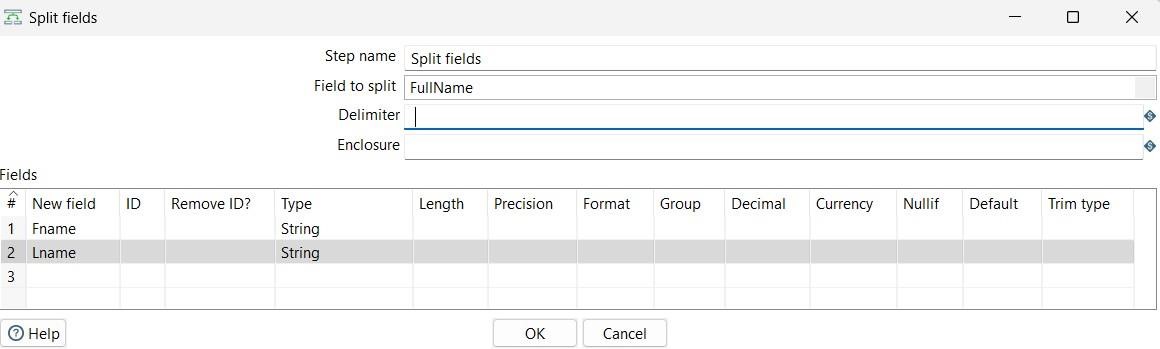
(performed in previous practical)



* 1. Similarly an Output folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel output file into the transformation area.
  2. From the Transform folder Drag and drop **Split fields** into the transformation area.
  3. Now create a new hop connection between all the three steps by clicking on the output connector from the input excel file.

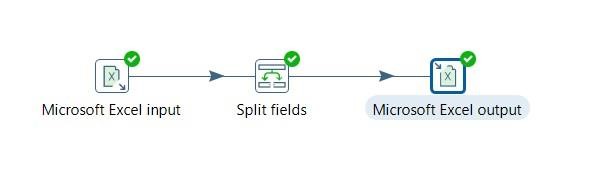


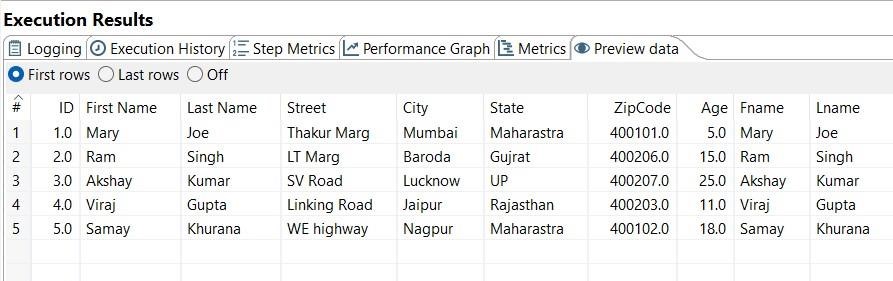
* 1. Double click on Split Fields and specify “Field to split” for the field name you would want to split and also assign the new names i.e “Fname” and “Lname” to columns after splitting.



* 1. Click on the Microsoft Excel Output file and save it by giving a suitable filename and click OK.
  2. Now run your project and you will see the following output.

**Output:**

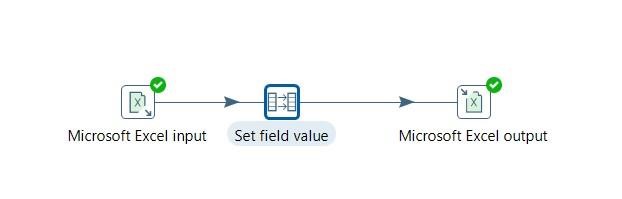




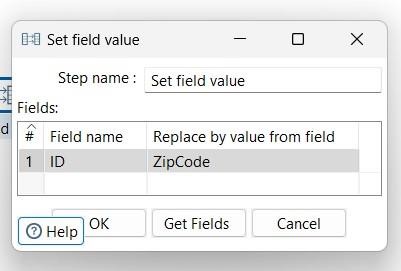
1. **Retrieve data from file staff.accdb and set value of zip column of staff table instaffid column**

**Steps:**

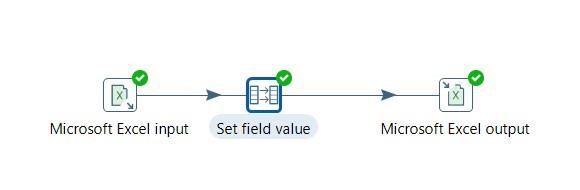
* 1. Open Data integration folder and launch the Pentaho Data Integration software by clicking on Window Batch File.
  2. Click on File > New Transformations.
  3. An Input folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel input file into the transformation area
  4. Similarly an Output folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel output file into the transformation area.
  5. From the Transform folder Drag and drop **Set Field Value** into the transformation area.
  6. Now create a new hop connection between all the three steps by clicking on the output connector from the input excel file.

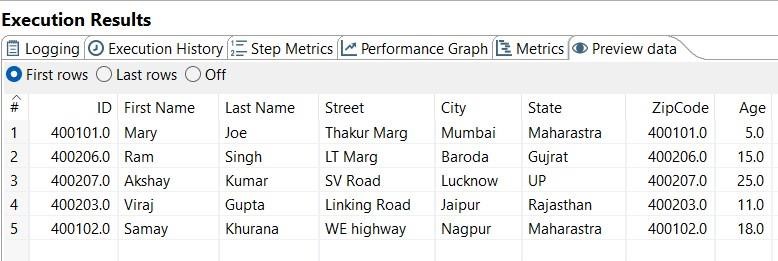


* 1. Double click on Set field value and specify the field name you want to set the value i.e Field name – ID , Replace by – ZIPcode column.



* 1. Click on the Microsoft Excel Output file and save it by giving a suitable filename and click OK.
  2. Now run your project and you will see the following output. **Output:**



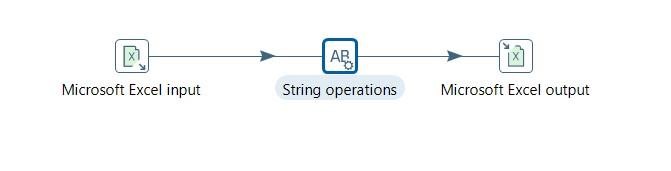


**8)Retrieve data from file staff.accdb and apply following string operation a. Convert city column in upper case**

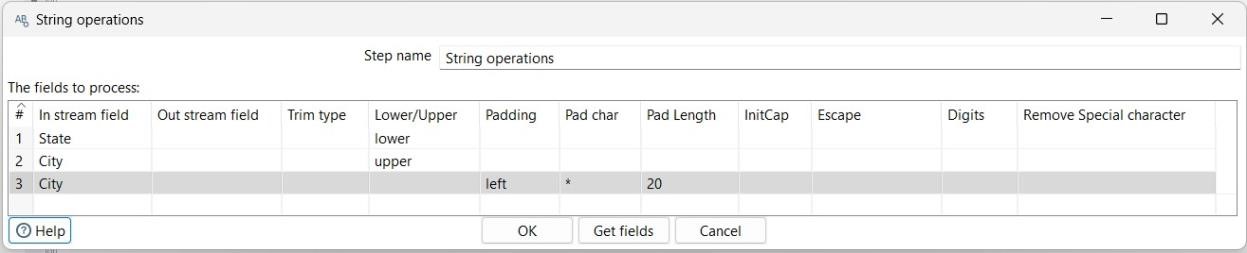
1. **Convert state column in lower case**
2. **Apply left padding in city and fix width 20 and padding character should be “\*”**
3. **Apply left padding in state and fix width 20 and padding character should b“#”**

**Steps:**

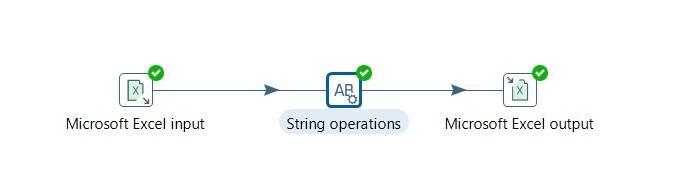
* 1. Open Data integration folder and launch the Pentaho Data Integration software by clicking on Window Batch File.
  2. Click on File > New Transformations.
  3. An Input folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel input file into the transformation area
  4. Similarly an Output folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel output file into the transformation area.
  5. From the Transform folder Drag and drop **String Operations** into the transformation area.
  6. Now create a new hop connection between all the three steps by clicking on the output connector from the input excel file.

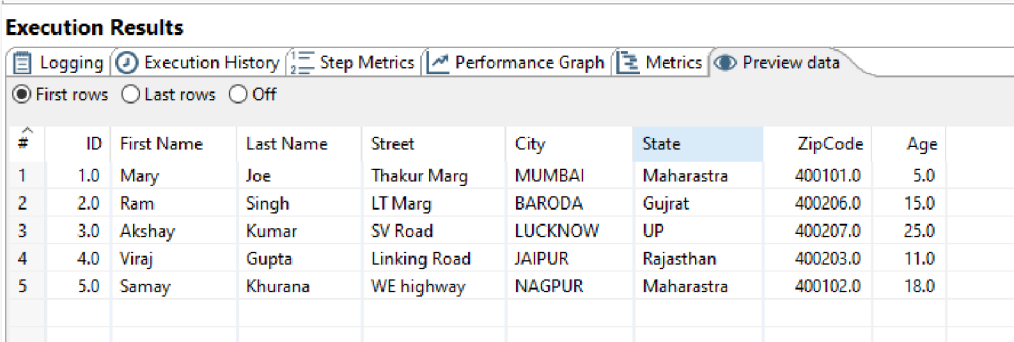


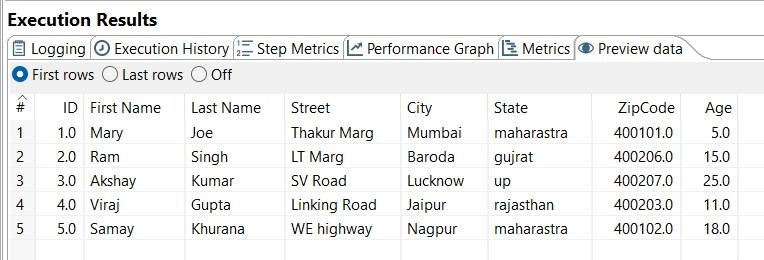
* 1. Double click on String operations and perform the required operations accordingly i.e convert to Uppercase or Lowercase or apply padding to some particular field



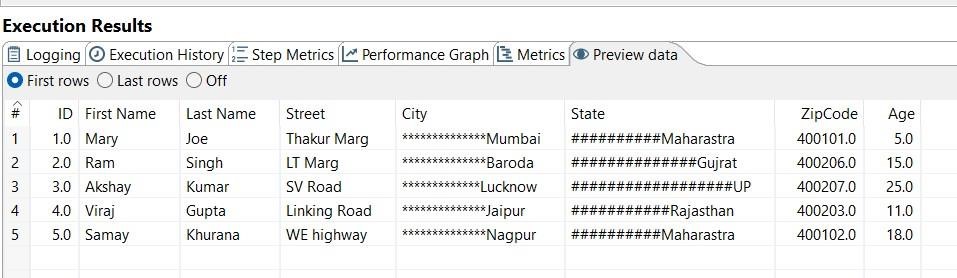
* 1. Click on the Microsoft Excel Output file and save it by giving a suitable filename and click OK.
  2. Now run your project and you will see the following output. **Output:**

 **a)**

 **b)**



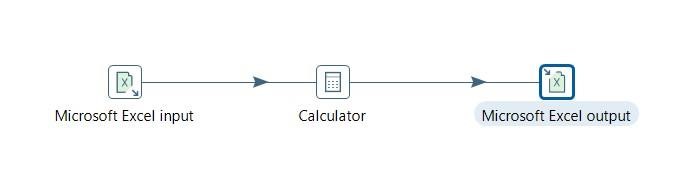
**(c,d)**



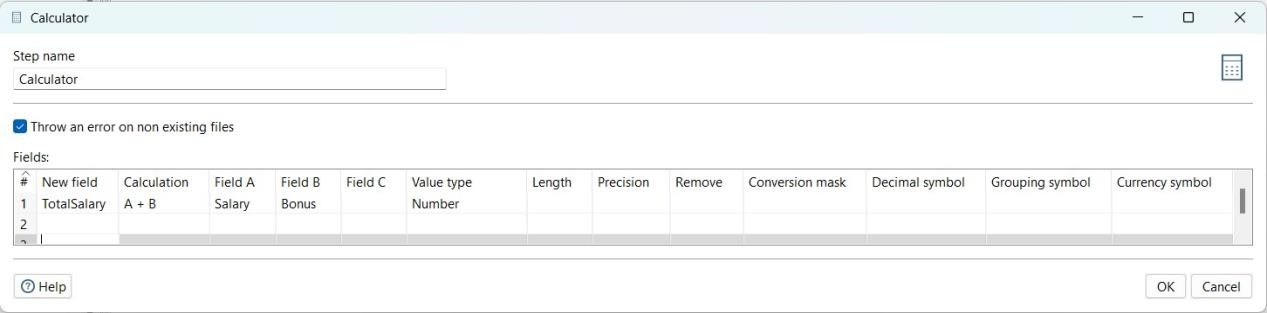
1. **Retrieve data from file staff.accdb and calculate total salary using which is sumof salary and bonus column**

**Steps:**

* 1. Open Data integration folder and launch the Pentaho Data Integration software by clicking on Window Batch File.
  2. Click on File > New Transformations.
  3. An Input folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel input file into the transformation area
  4. Similarly an Output folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel output file into the transformation area.
  5. From the Transform folder Drag and drop **Calculator** into the transformation area.
  6. Now create a new hop connection between all the three steps by clicking on the output connector from the input excel file.

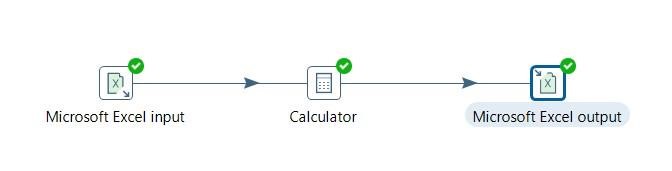


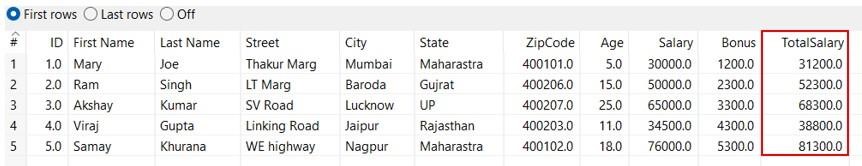
* 1. Double click on Calculator and add the constraints as shown in fig. below



* 1. Click on the Microsoft Excel Output file and save it by giving a suitable filename and click OK.
  2. Now run your project and you will see the following output.

**Output:**

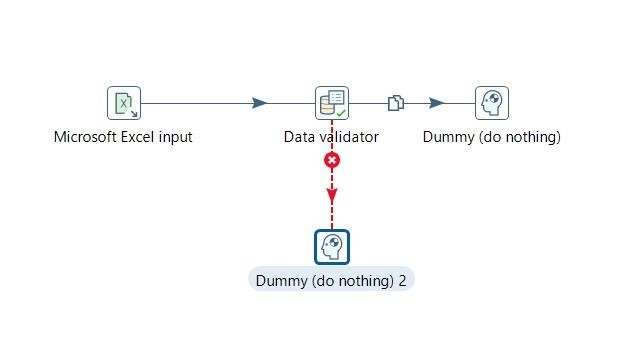




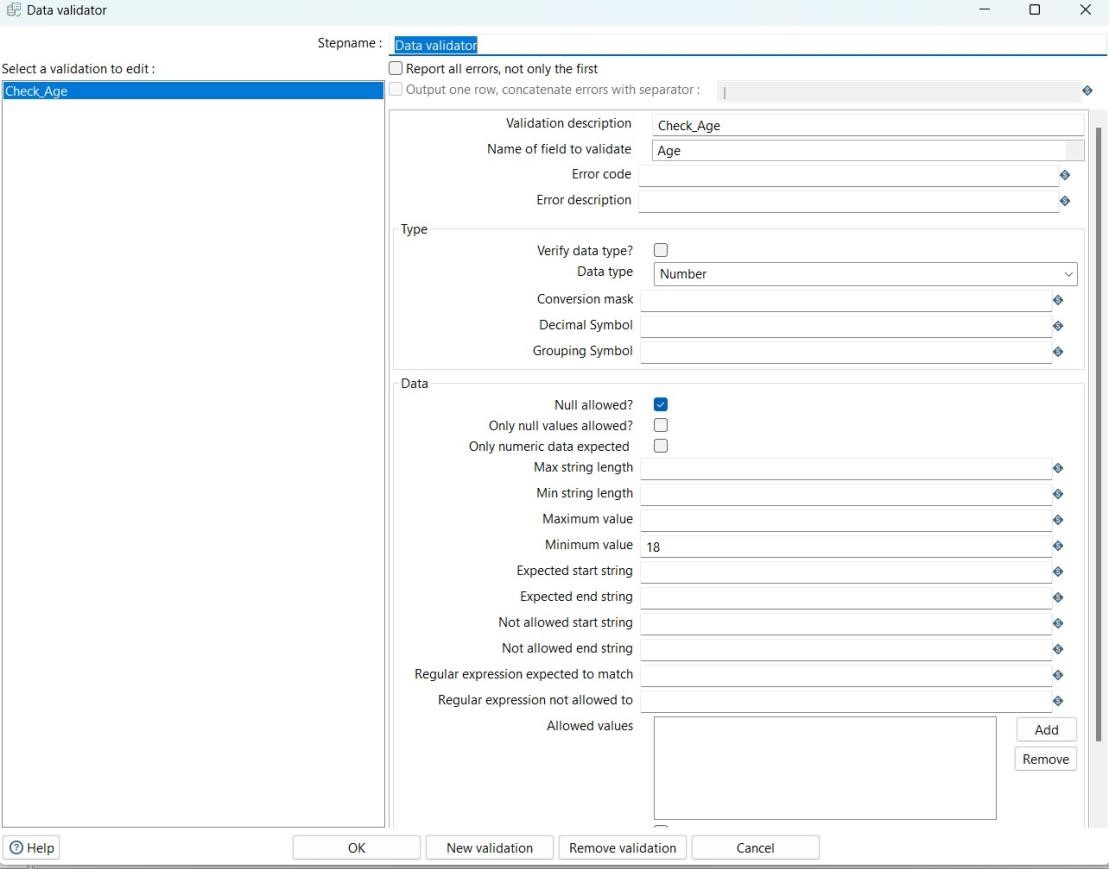
1. **Retrieve data from file staff.accdb and validate Age field where age should begreater or equal to 18**

**Steps:**

* 1. Open Data integration folder and launch the Pentaho Data Integration software by clicking on Window Batch File.
  2. Click on File > New Transformations.
  3. An Input folder can be seen in the Design Panel. From there Drag and Drop Microsoft Excel input file into the transformation area
  4. From the Validation folder Drag and drop **Data Validator** into the transformation area.
  5. Now drag two dummy data from Flow folder and connect Data Validator to one Dummy and select as a main output.
  6. Connect Data Validator to another Dummy and select as an error handling.

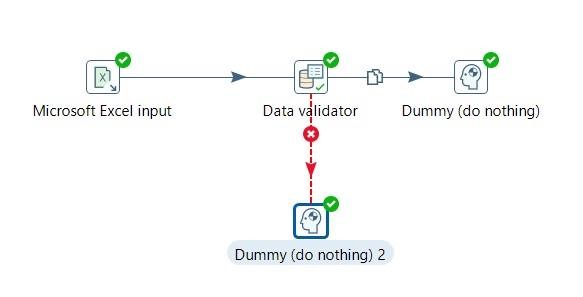


* 1. Double click on Data Validator and add New Validation and set the fields with conditions you want to validate

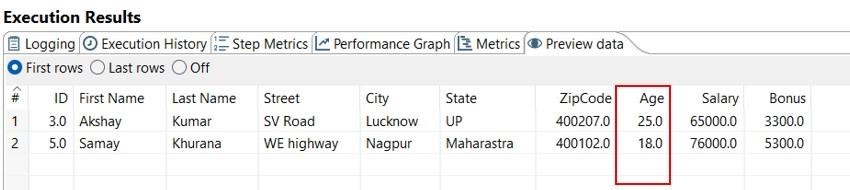


* 1. Now save your transformation and run your project.

**Output:**



**Main Output:**



**As error Handling:**

