GROUP MEMBERS NAME

1. ANAND KARWA
2. MONIKA AVADHA

***MANUAL FOR TALEND WORKSSHOP***

***TOPIC:- 1)-tUniqRow***

***2)-tReplaceList***

***1)-tUniqRow:***

The Standard tUniqRow component belongs to the Data Quality family.

Ensures data quality of input or output flow in a Job.

tUniqRow compares entries and sorts out duplicate entries from the input flow.

***Basic Settings:***

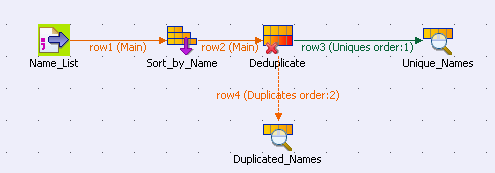
 Edit schema:- A schema is a row description. It defines the number of fields (columns) to be processed and passed on to the next component. The schema is either **Built-In** or stored remotely in the **Repository**.

Click **Edit schema** to make changes to the schema. If the current schema is of the **Repository** type, three options are available:

* **View schema**: choose this option to view the schema only.
* **Change to built-in property**: choose this option to change the schema to **Built-in** for local changes.

***tUniqRow For* Deduplicating entries:-**

In this five-component Job, we will sort entries on an input name list, find out duplicated names, and display the unique names and the duplicated names on the **Run** console.



#### Setting up the Job:-

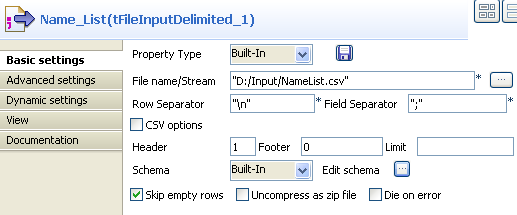
***Procedure:-***

1. Drop a **tFileInputDelimited**, a **tSortRow**, a **tUniqRow**, and two **tLogRow** components from the **Palette** to the design workspace, and name the components as shown above.
2. Connect the **tFileInputDelimited** component, the **tSortRow** component, and the **tUniqRow** component using **Row** > **Main** connections.
3. Connect the **tUniqRow** component and the first **tLogRow** component using a **Main** > **Uniques** connection.
4. Connect the **tUniqRow** component and the second **tLogRow** component using a **Main** > **Duplicates** connection.

#### Configuring the components:-

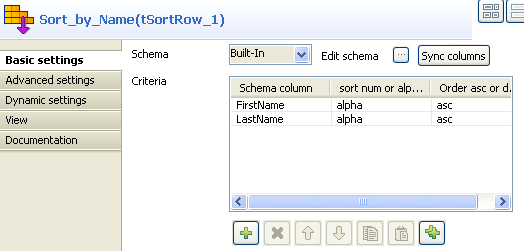
***Procedure:-***

1.component to display its **Basic settings** view. Double-click the **tFileInputDelimited**



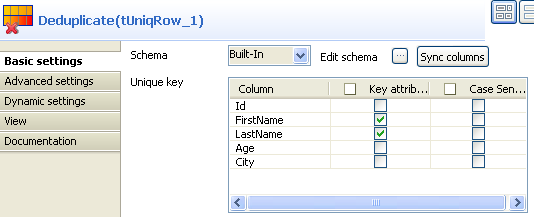
1. *2.Click the****[...]****button next to the****File Name****field to browse to your input file.*
2. *3.Define the header and footer rows. In this use case, the first row of the input file is the header row.*
3. *4.Click****Edit schema****to define the schema for this component. In this use case, the input file has five columns:*Id*,*FirstName*,*LastName*,*Age*, and*City*. Then click****OK****to propagate the schema and close the schema editor.*

*5.Double-click the****tSortRow****component to display its****Basic settings****view.*



1. ***6.To rearrange the entries in the alphabetic order of the names, add two rows in the Criteria table by clicking the plus button, select the*FirstName*and*LastName*columns under Schema column, select*alpha*as the sorting type, and select the sorting order.***

***7.Double-click the tUniqRow component to display its Basic settings view***.



***8.In the Unique key area, select the columns on which you want deduplication to be carried out. In this use case, you will sort out duplicated names.***

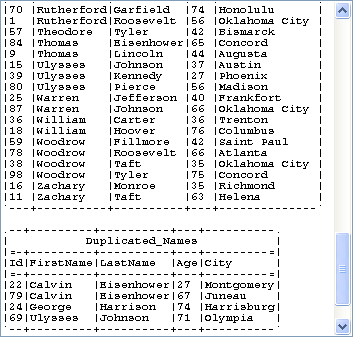
***9.In the Basic settings view of each of the tLogRow components, select the Table option to view the Job execution result in table mode.***

#### Saving and executing the Job:-

***Procedure:-***

1. Press **Ctrl+S** to save your Job.
2. Run the Job by pressing **F6** or clicking the **Run** button on the **Run** tab.

The unique names and duplicated names are displayed in different tables on the **Run** console.



***2)-tReplaceList:-***

***Carries out a Search and Replace operation in the input columns defined based on an external lookup.***

***Helps to cleanse all files before further processing.***

**Basic settings:-**

Edit schema:- A schema is a row description, it defines the number of fields to be processed and passed on to the next component. The schema is either **Built-in** or stored remotely in the**Repository**.

Since version 5.6, both the **Built-In** mode and the **Repository**mode are available in any of the Talend solutions.

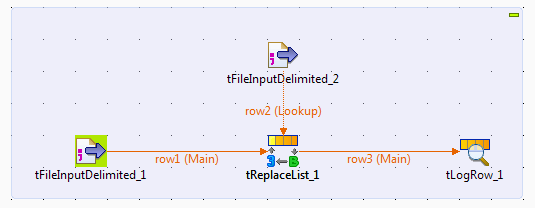
Two read-only columns, Value and Match are added to the output schema automatically.

### Warning

The data Type defined in the schemas must be consistent, ie., an integer can only be replaced by another integer using an integer as a look up field. Values of one type cannot be replaced by values of another type.

### Replacement from a reference file:-

### *The following Job searches and replaces a list of states with their corresponding two-letter codes. The relevant codes are taken from a reference file placed as lookup flow in the Job.*



#### Setting up the Job:-

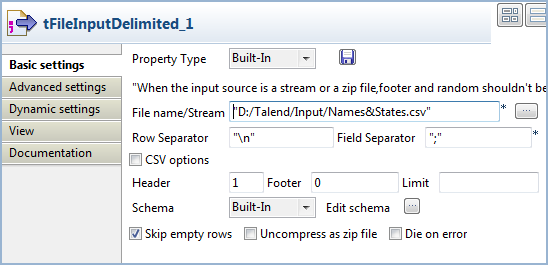
***Procedure:-***

1. Drop the following components from the **Palette** onto the design workspace: two **tFileInputDelimited**components, a **tReplaceList** and a **tLogRow**.
2. Connect the two **tFileInputDelimited** components to the **tReplaceList** component using **Row** > **Main**connections. Note that the link between the reference input component (the second **tFileInputDelimited**) and the **tReplaceList** component appears as a lookup row.
3. Connect the **tReplaceList** component to the **tLogRow** component using a **Row** > **Main** connection.

#### Configuring the components:-

***Procedure:-***

***1.Double-click the first tFileInputDelimited component to open its Basic settings view and set the parameters of the main input flow, including the path and name of the file to read and the number of header rows to skip.***



***In this example, the main input file provides a list of people names and US state names. The following shows an extract of the file content***:

name;state

Andrew Kennedy;Mississippi

Benjamin Carter;Louisiana

Benjamin Monroe;West Virginia

Bill Harrison;Tennessee

Calvin Grant;Virginia

Chester Harrison;Rhode Island

Chester Hoover;Kansas

Chester Kennedy;Maryland

Chester Polk;Indiana

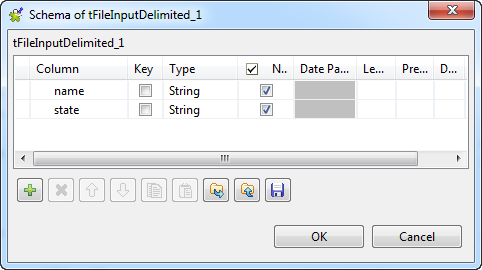
Dwight Nixon;Nevada

Dwight Roosevelt;Mississippi

Franklin Grant;Nebraska

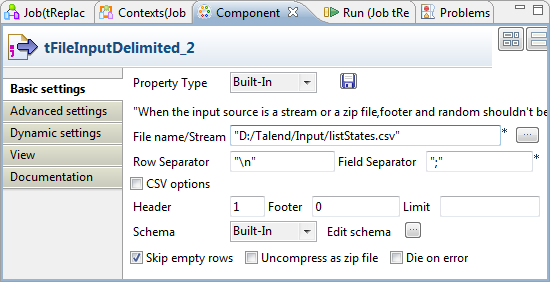
***2.Click the [...] button next to Edit schema to open the [Schema] dialog box and set the input schema.***

***According to the structure of the main input file, the input schema should contain two columns:*name*and* state*.***



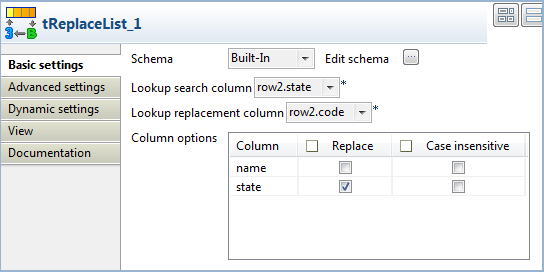
***When done, click OK to close the dialog box and propagate the changes to the next component.***

***3.Define the properties of the second tFileInputDelimited component similarly.***



***In this example, the reference input file provides a list of states and their two-letter codes. Accordingly, the reference input schema should have two columns:*state*and*code*.***

***4.Double-click the tReplaceList component to open its Basic settings view to set the operation to carry out.***



1. ***5.From the Lookup search column list, select the column to be searched. In this use case, we want to carry out a search on the*state*column.***
2. ***6.From the Lookup replacement column list, select the column containing the replacement values,*code*for the two-letter state codes in this example.***
3. ***7.In the Column options table, select Replace check box for the*states*column, to replace the state names with their corresponding codes.***

***8.In the tLogRow component, select the Table check box for a better readability of the output.***

#### Executing the Job

***-Save the Job and press F6 to execute it.***

1. 