This page intentionally left blank to ensure new chapters start on right (odd number) pages.

LESSON 4

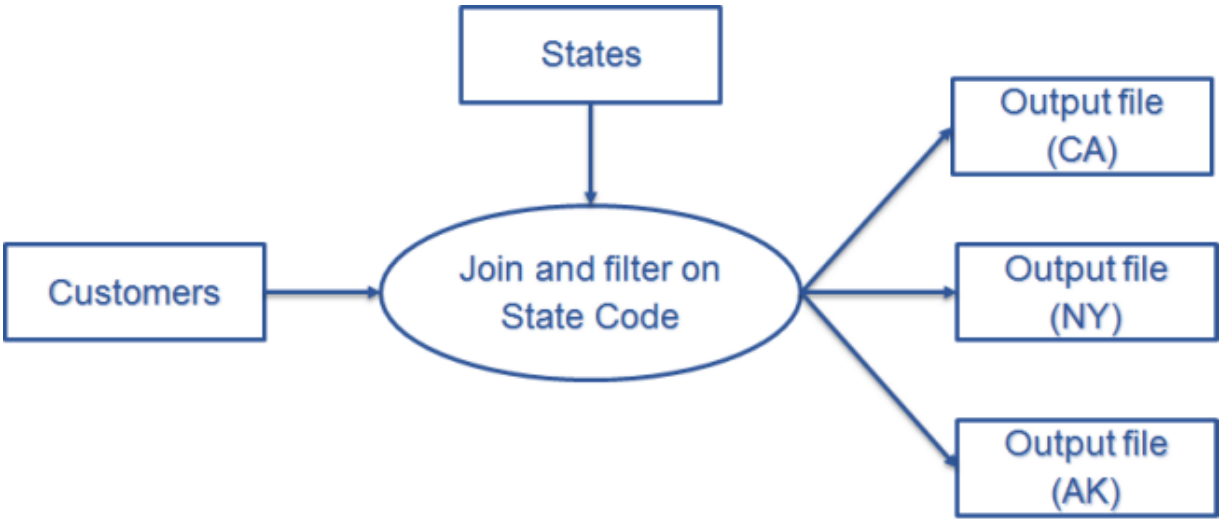
Filtering Data

This chapter discusses the following.

Filtering Data

Lesson Overview

A common task in data integration projects is to filter data rows based on content, for separate processing, storage, or reporting. In this lesson, you will build a Job that extends the previous Job to join the customer data with state data and then separate the results into different output flows based on the value of the column containing a State Code:



Objectives

After completing this lesson, you will be able to:

- Use the tMap component to filter data
- Execute Job sections conditionally
- Duplicate output flows

Next Step

The first step is to add a filter on the data so that only customers from one state are written to the output file.

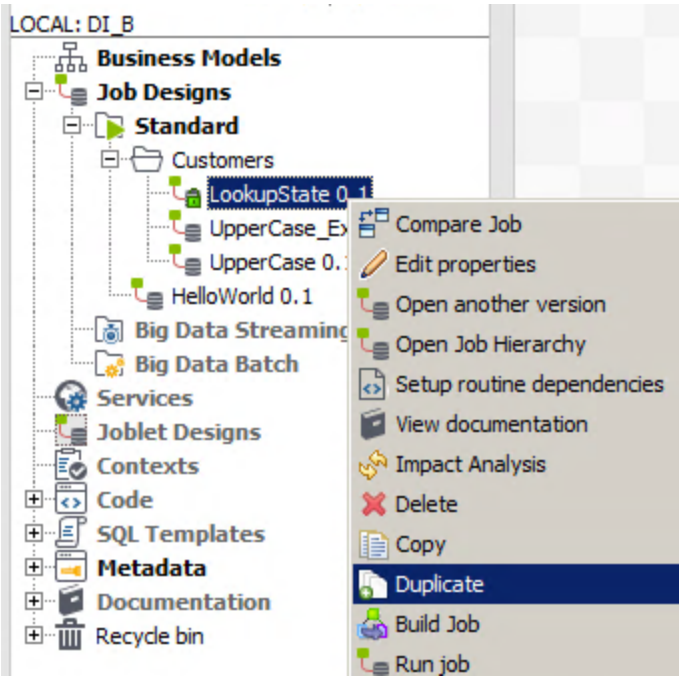
Filter output data

Overview

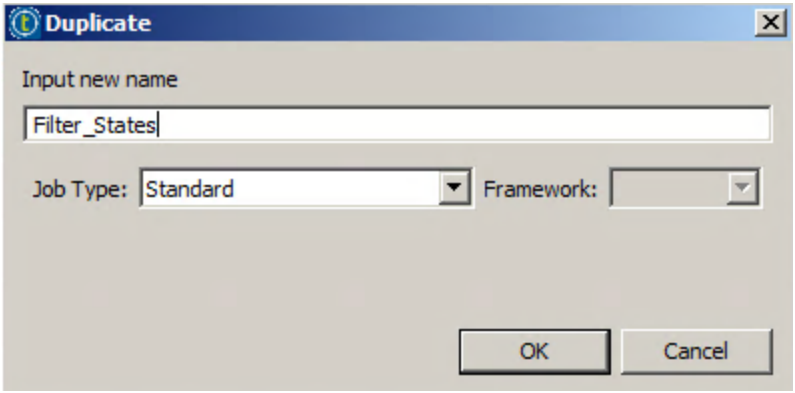
Your Job needs to filter the output based on the **State code** so that the resultant output files are isolated to specific states.

Duplicate Job

1. Right-click on the Job **LookupState** and click **Duplicate**:



2. Enter *Filter_States* in the **Input new name** field and then click **OK**:

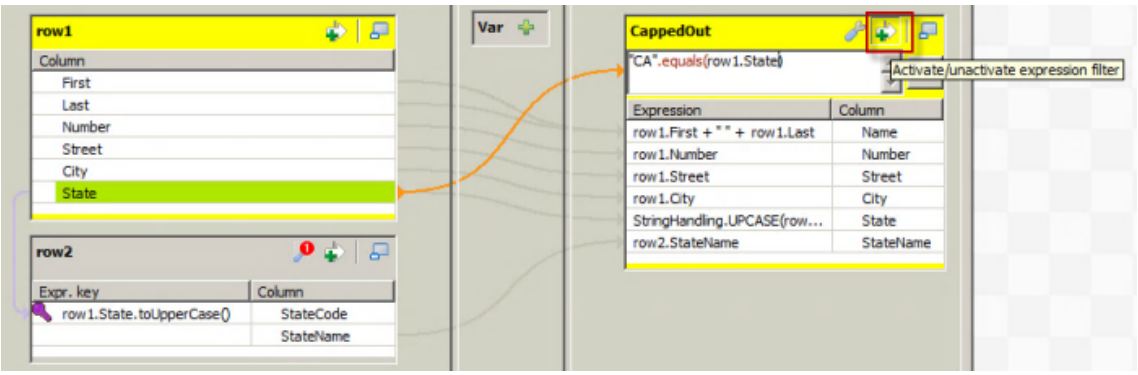


3. Double-click the new Job to open it in the design space.

Add Filter

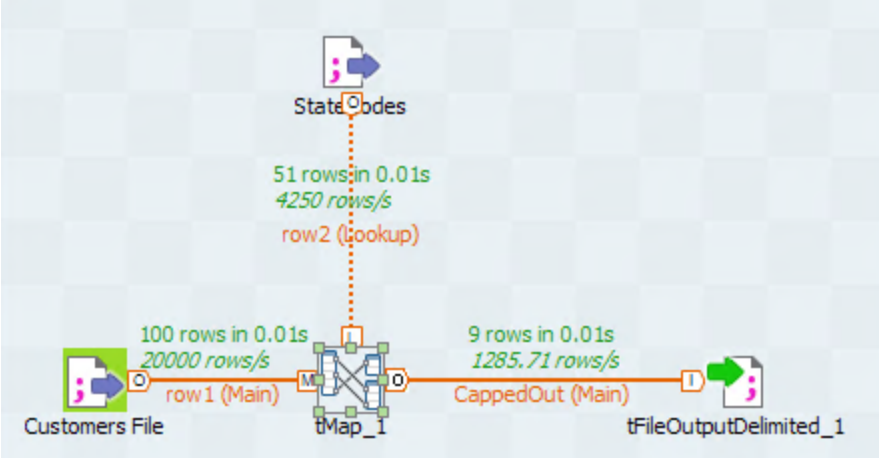
1. Double-click the **tMap** component.

   Click the **Activate/ unactivate expression filter** Icon on top right of **CappedOut** and enter *"CA".equals(row1.State)* for the filter definition and click **Ok**:



Once the Expression filter is correct, an orange arrow is added for you that maps the flow from the appropriate place in the input table to the output table.

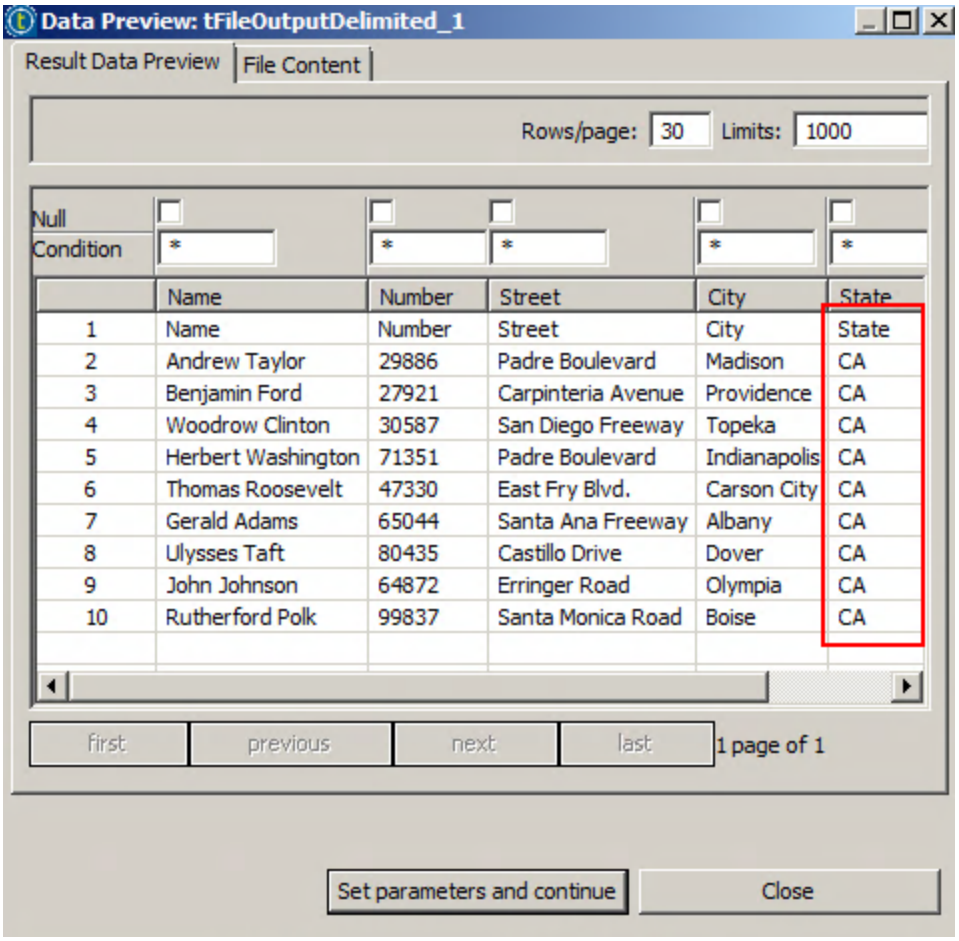
2. Run the Job.



Note that only 9 rows are now written to the output file.

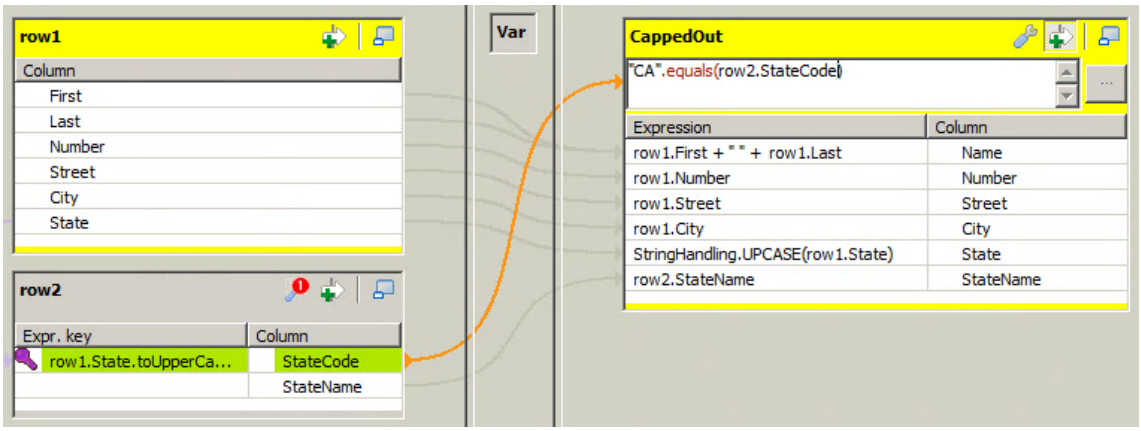
3. Right-click the output component **tFileOutputDelimited** and select **Data viewer.**

   Note the filtered data are only records having *CA* (in Upper Case) for the **State** column from the **Custs.csv** input file.



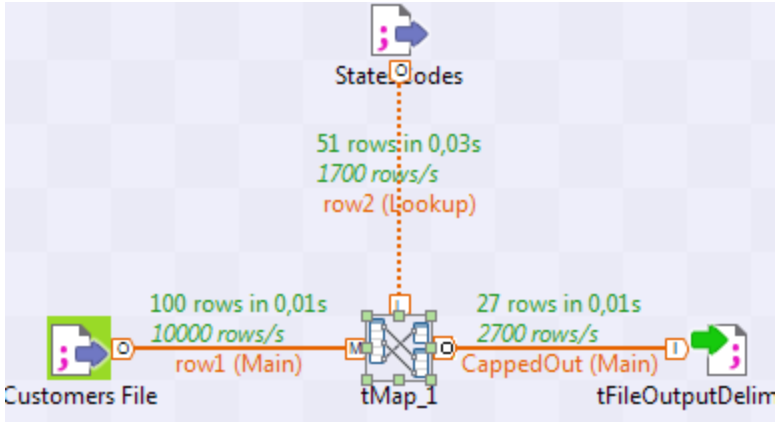
Click **Close**.

4. Double-click **tMap** again and change the filter to *"CA".equals(row2.StateCode)* and click **Ok**:

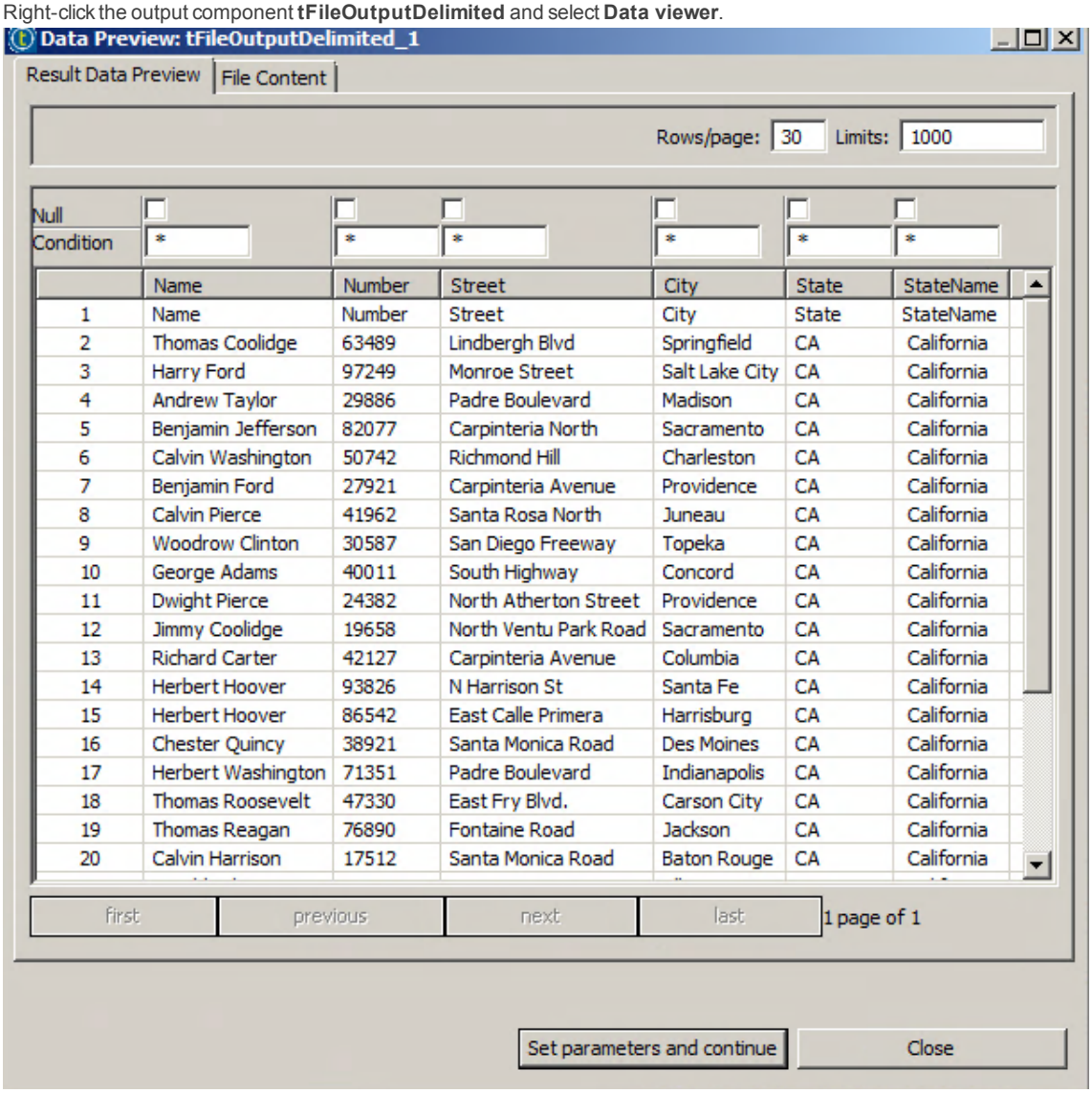


Notice that once again the orange arrow is placed for you and identifies the data flow for the changed expression. Recall that *row2.StateCode* contains the state code in uppercase.

5. Run the Job again and you will see that the output contains 27 rows:



6. Right-click the output component **tFileOutputDelimited** and select **Data viewer**.



	Name	Number	Street	City	State	StateName
1	Name	Number	Street	City	State	StateName
2	Thomas Coolidge	63489	Lindbergh Blvd	Springfield	CA	California
3	Harry Ford	97249	Monroe Street	Salt Lake City	CA	California
4	Andrew Taylor	29886	Padre Boulevard	Madison	CA	California
5	Benjamin Jefferson	82077	Carpinteria North	Sacramento	CA	California
6	Calvin Washington	50742	Richmond Hill	Charleston	CA	California
7	Benjamin Ford	27921	Carpinteria Avenue	Providence	CA	California
8	Calvin Pierce	41962	Santa Rosa North	Juneau	CA	California
9	Woodrow Clinton	30587	San Diego Freeway	Topeka	CA	California
10	George Adams	40011	South Highway	Concord	CA	California
11	Dwight Pierce	24382	North Atherton Street	Providence	CA	California
12	Jimmy Coolidge	19658	North Ventu Park Road	Sacramento	CA	California
13	Richard Carter	42127	Carpinteria Avenue	Columbia	CA	California
14	Herbert Hoover	93826	N Harrison St	Santa Fe	CA	California
15	Herbert Hoover	86542	East Calle Primera	Harrisburg	CA	California
16	Chester Quincy	38921	Santa Monica Road	Des Moines	CA	California
17	Herbert Washington	71351	Padre Boulevard	Indianapolis	CA	California
18	Thomas Roosevelt	47330	East Fry Blvd.	Carson City	CA	California
19	Thomas Reagan	76890	Fontaine Road	Jackson	CA	California
20	Calvin Harrison	17512	Santa Monica Road	Baton Rouge	CA	California

Next

Now that you have filtered data for one State Code, you can learn how to create multiple outputs, filtering different states in the **tMap** component.

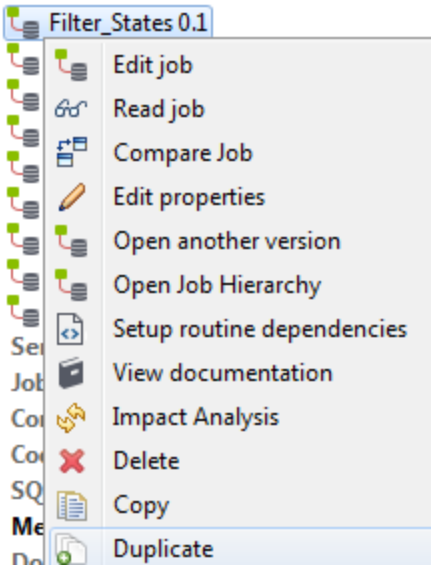
Using tMap for Multiple Filters

Overview

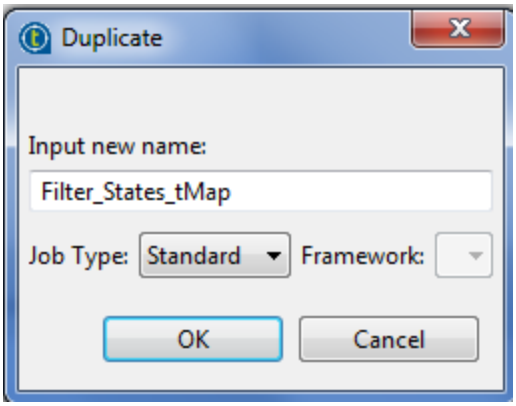
You will now learn how to filter on several states in one **tMap** component, following best practices. You will create 3 output files each containing a different state.

Duplicate Job

1. Right-click on the Job **Filter_States** and click **Duplicate**:



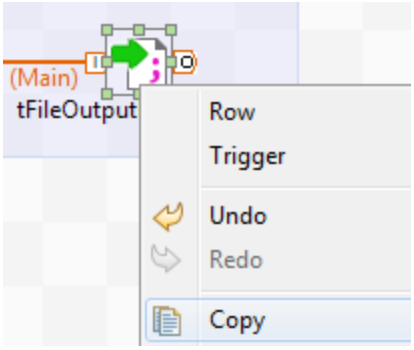
2. Name the new Job **Filter_States_tMap** and click **OK**:



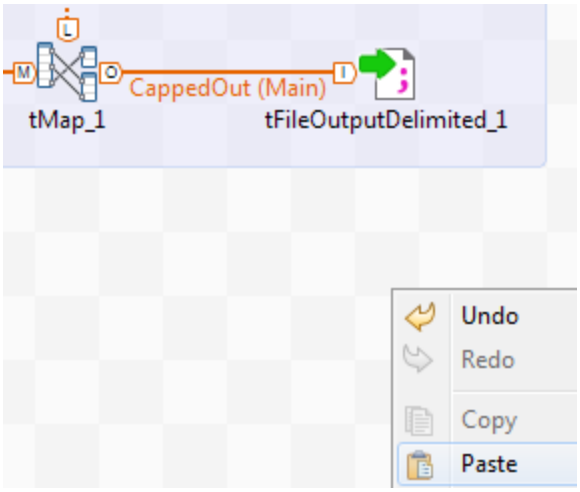
3. Double-click the newly created job to open it.

Edit Output

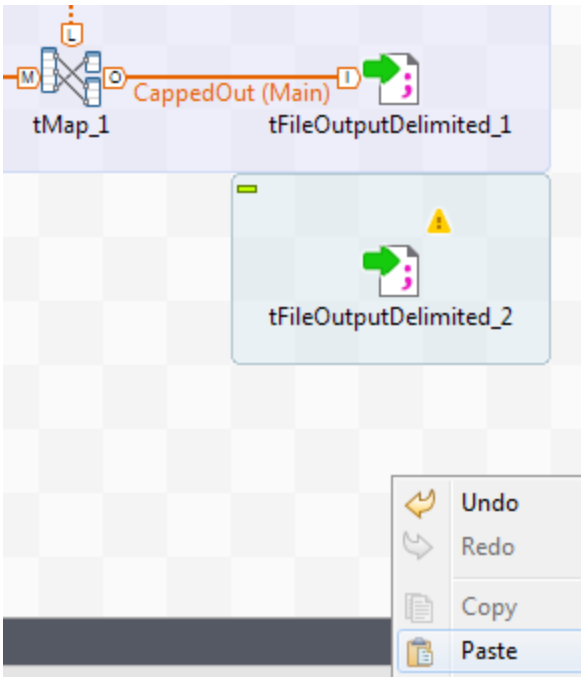
1. Enter a new output component to the Job. Right-click on **tFileOutputDelimited** and click **Copy**:



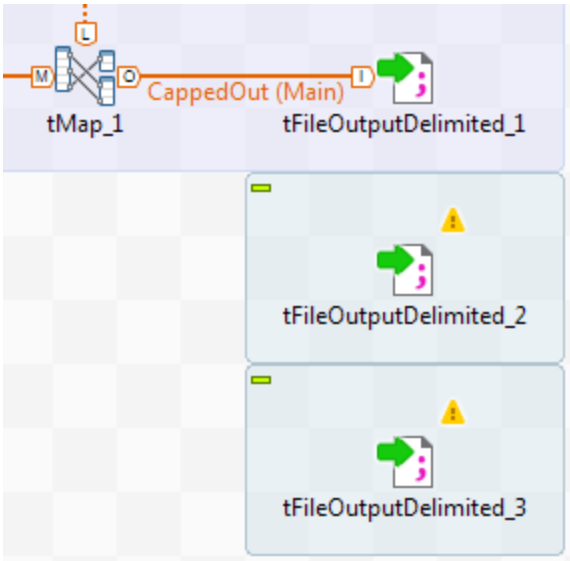
2. Right-click below the output component in the canvas and click **Paste** to place the new **tFileOutputDelimited** component:



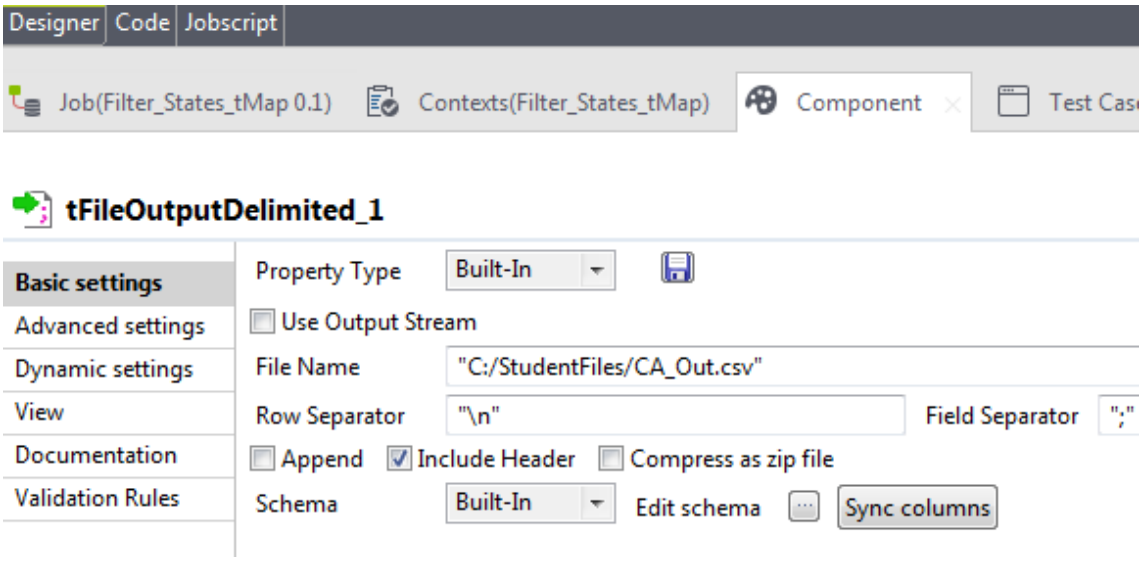
3. The component is placed below the existing **tFileOutputDelimited** component.
   Below the second **tFileOutputDelimited** component right-click and select **Paste** again to drop the third **tFileOutputDelimited** component to your Job:



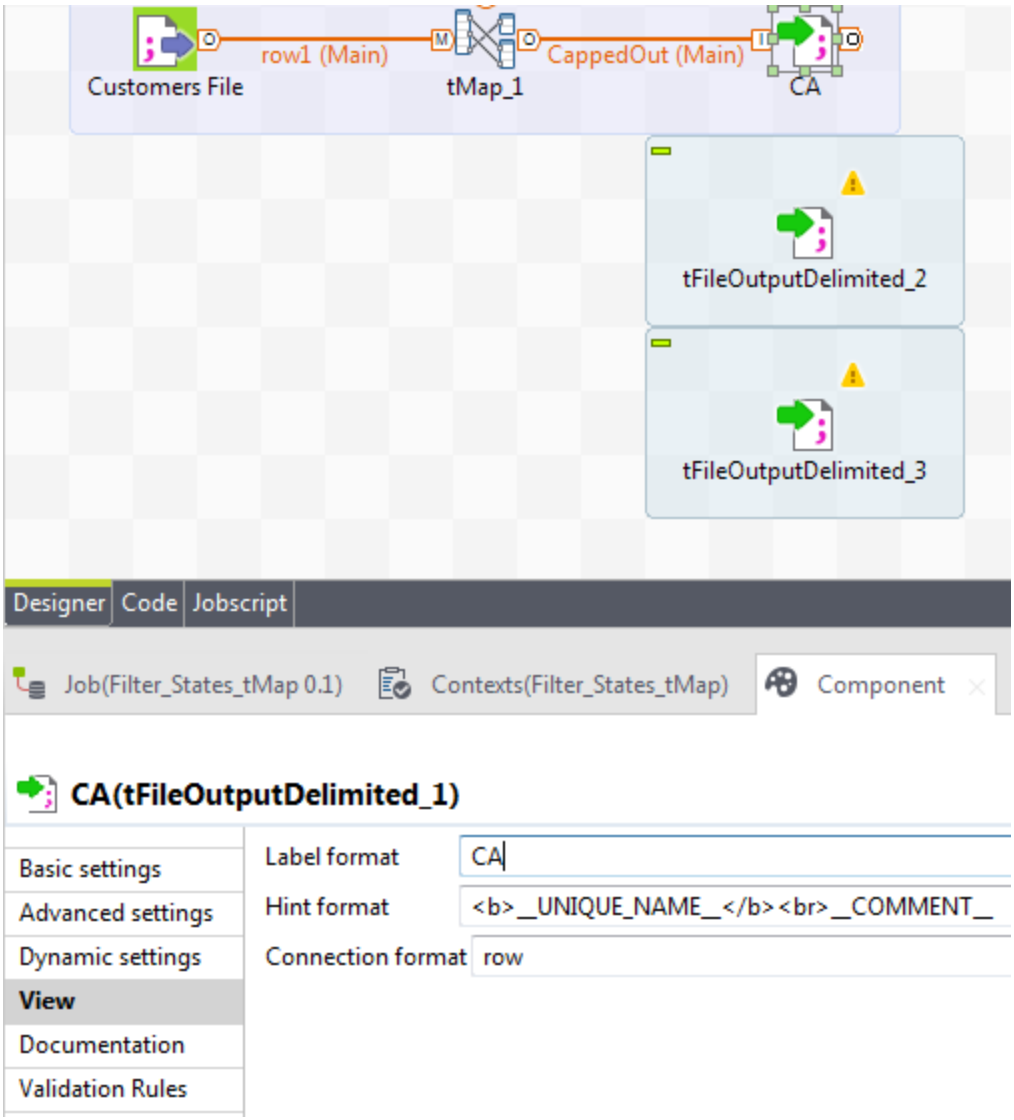
4. You have placed three outputs now in your Job:



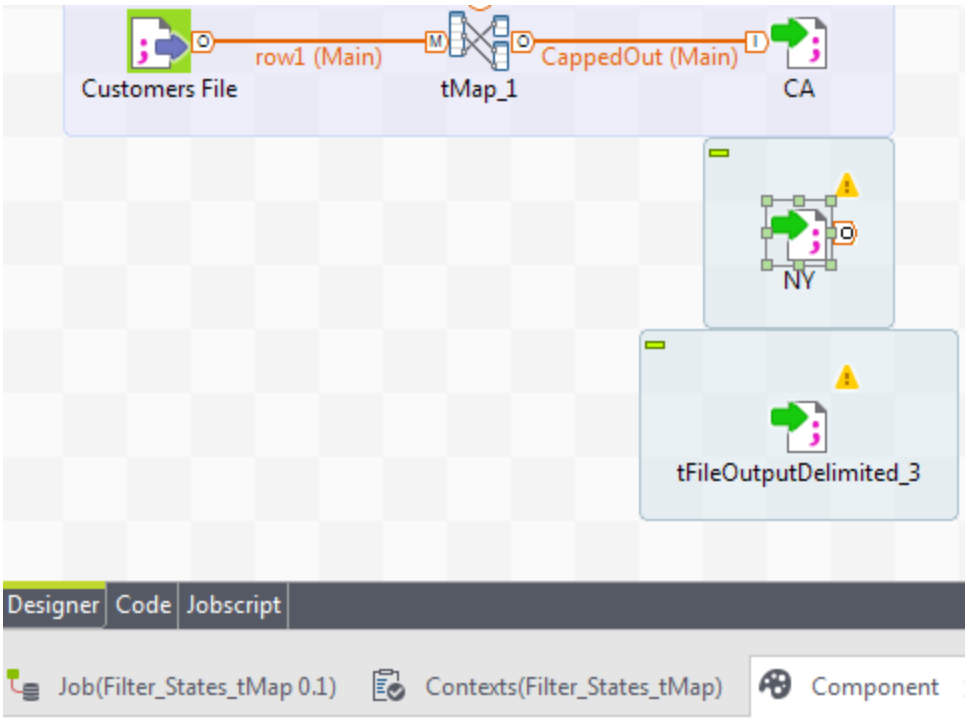
5. Remember the last setting in **tMap** was a filter, filtering entries for **CA** (California) and writing the output to a file. Now you will edit the settings for the first output component and change the output file name. Double-click **tFileOutputDelimited_1** and change the **File Name** to *"C:/StudentFiles/CA_Out.csv"*:

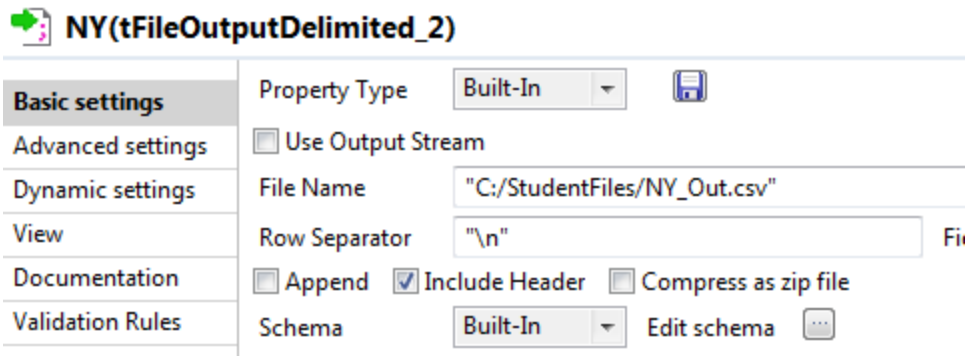


6. Click **View** to change the output name to **CA**. Enter **CA** into the field **Label format**:

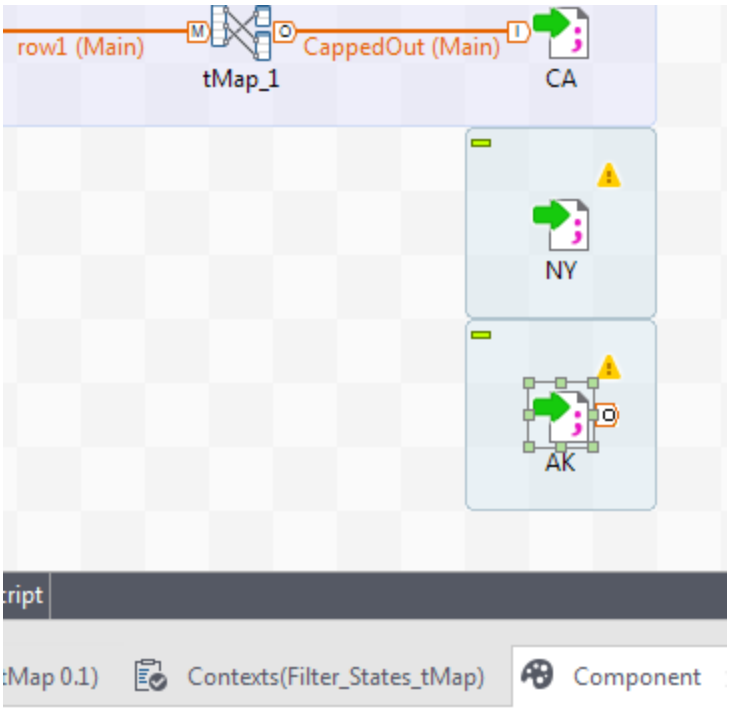


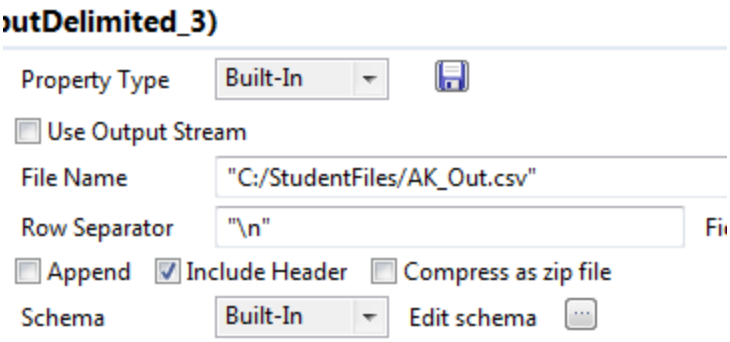
7. Now double-click the second **tFileOutputDelimited** component, change the file name to **NY_Out.csv** and enter **NY** for the **Label format**:



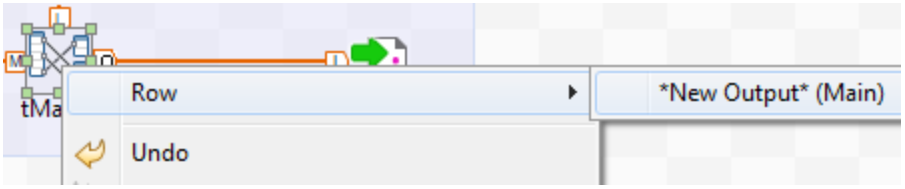


8. Enter the changes for the third **tFileOutputDelimited** component to **AK**:

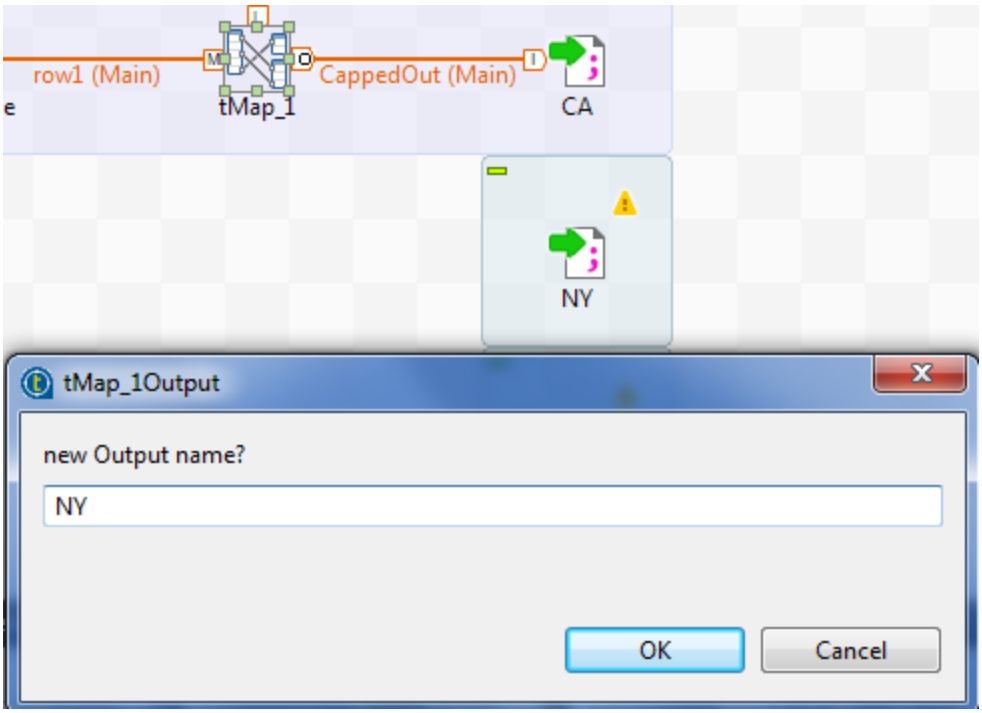




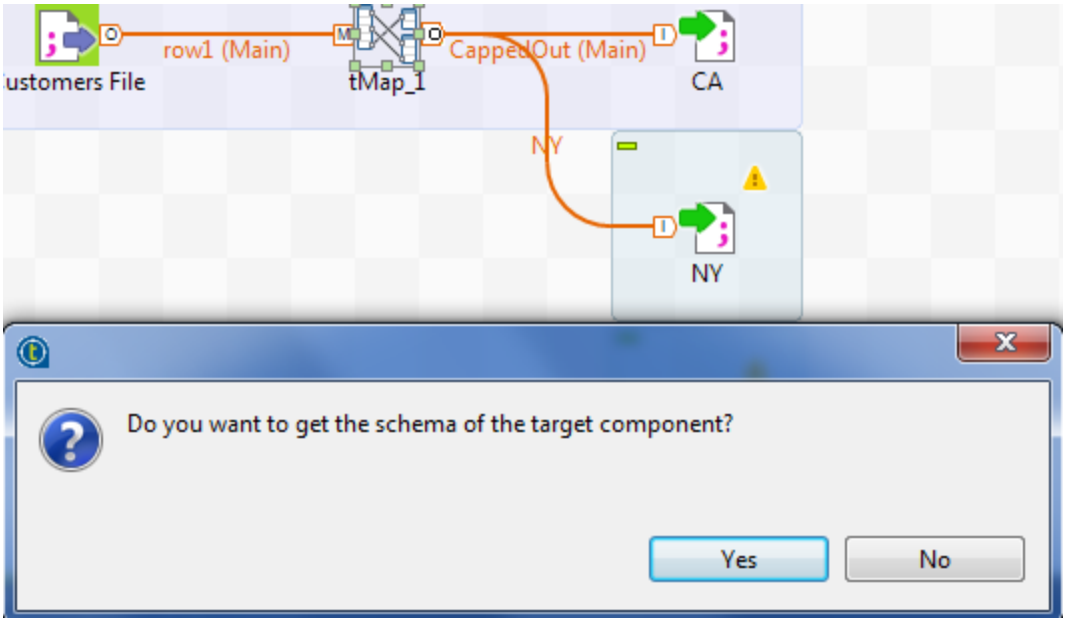
9. Right-click the **tMap** component and click **Row** then ***New Output* (Main)**:



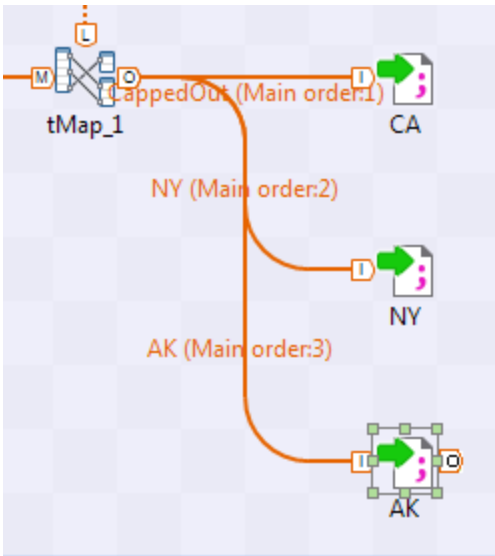
10. Connect it to the **NY** component, name it **NY** and then click **OK**:



11. Click **Yes** when prompted **Do you want to get the schema of the target component?**:

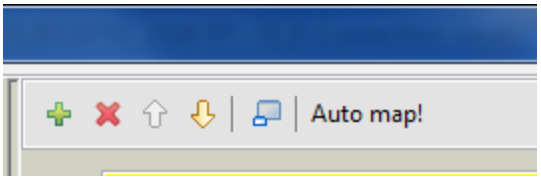


12. Repeat these steps for the last output component **AK**:

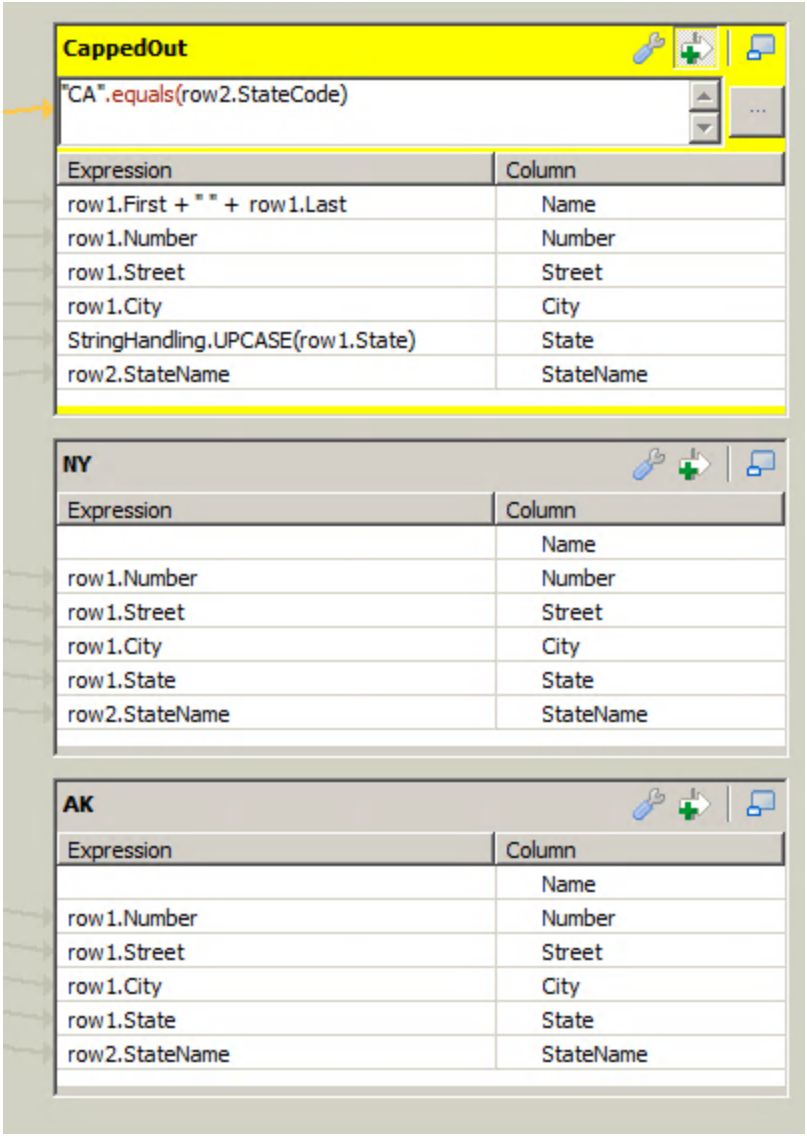


Edit tMap

1. Double-click the **tMap** component to open the Mapping Editor.
2. Click the **Auto map!** button on the top right:



3. The system will map the elements as follows:



CappedOut

"CA".equals(row2.StateCode)

Expression	Column
row1.First + " " + row1.Last	Name
row1.Number	Number
row1.Street	Street
row1.City	City
StringHandling.UPCASE(row1.State)	State
row2.StateName	StateName

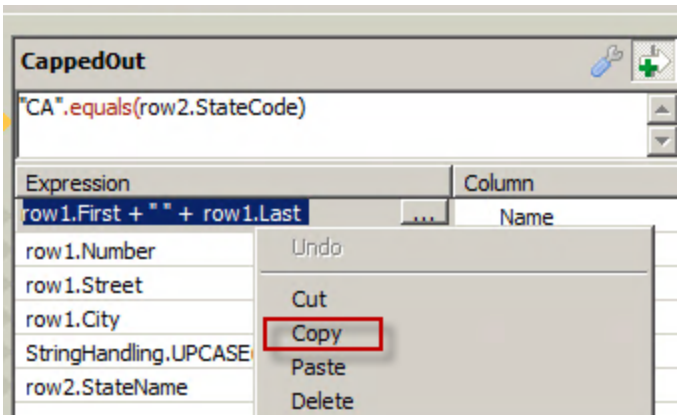
NY

Expression	Column
	Name
row1.Number	Number
row1.Street	Street
row1.City	City
row1.State	State
row2.StateName	StateName

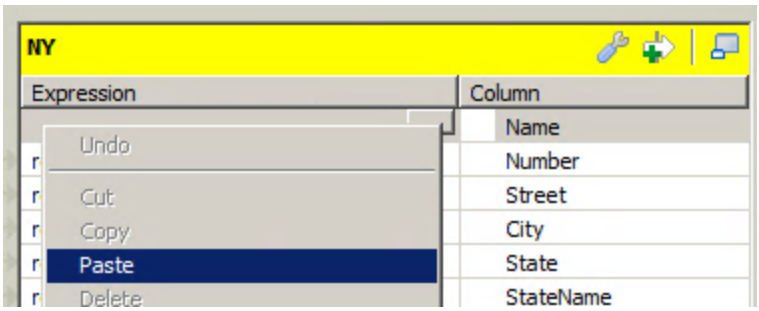
AK

Expression	Column
	Name
row1.Number	Number
row1.Street	Street
row1.City	City
row1.State	State
row2.StateName	StateName

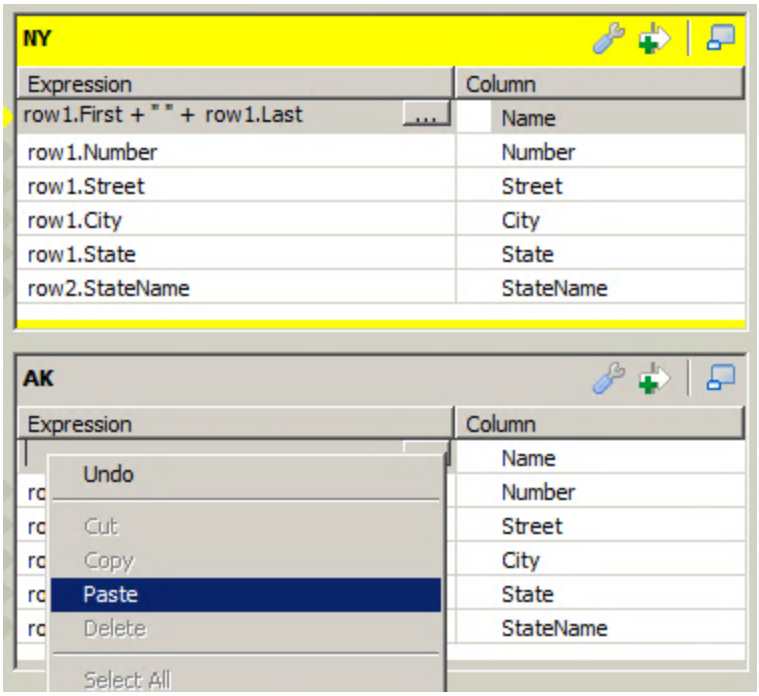
4. You can copy and paste Expressions. Right-click on the expression for the field **Name** in the **CappedOut** table and click **Copy**:



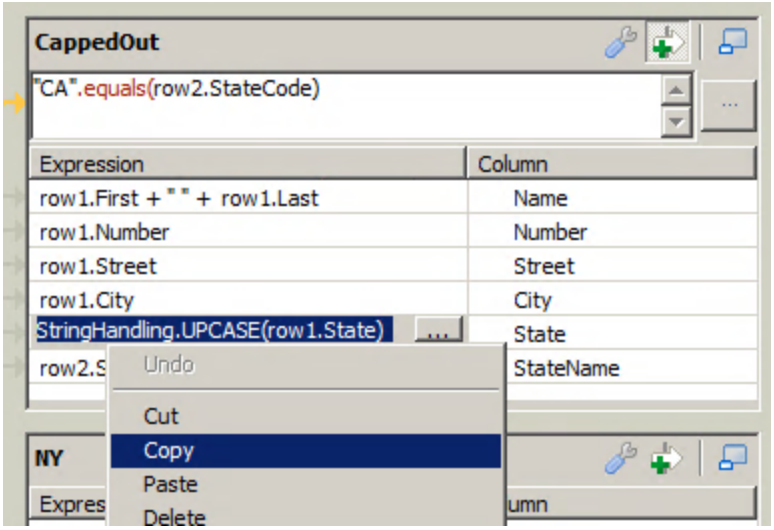
5. Right-click the expression for the field **Name** in the output **NY** and click **Paste**:



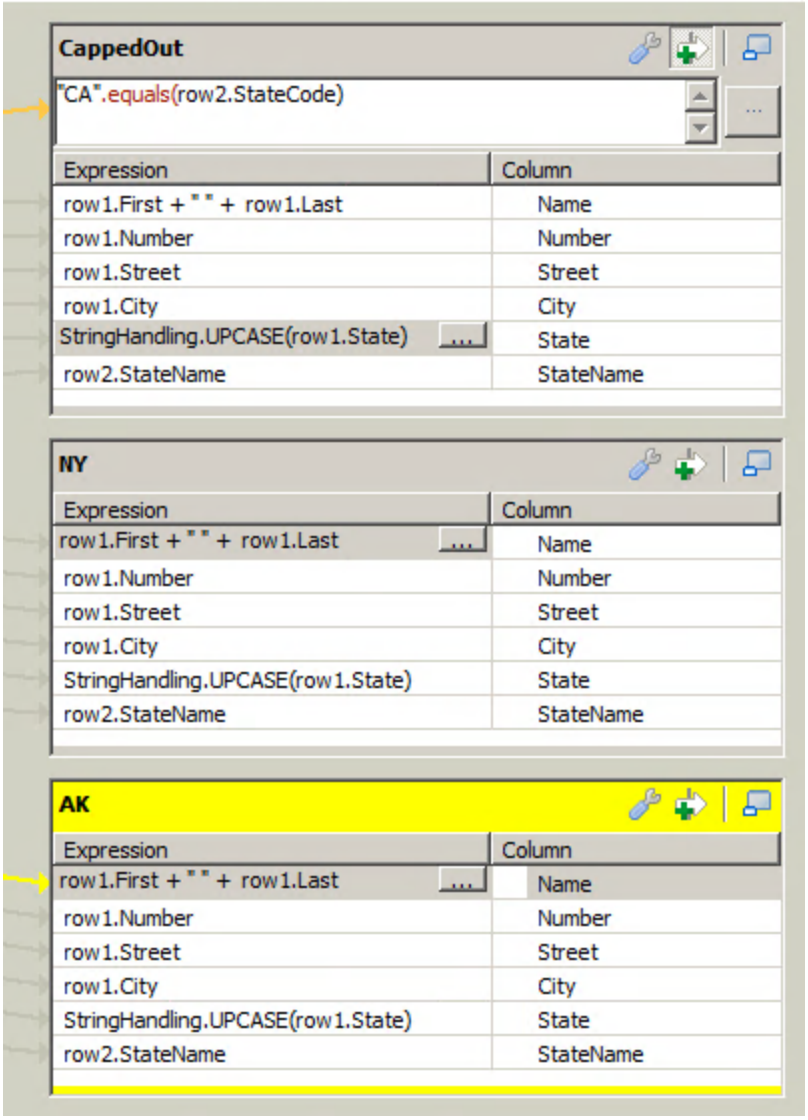
6. Paste it also in the **Name** column for the output **AK**:



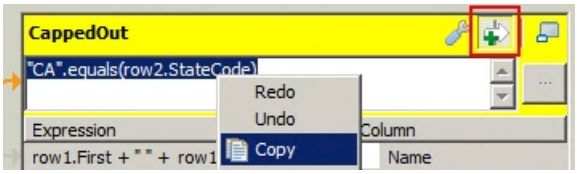
7. Copy and paste the expression for the field **State** also:



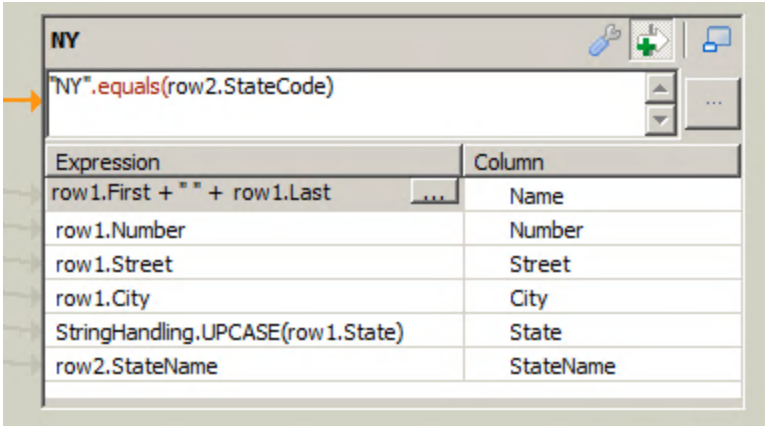
8. Your output tables should look as follows:



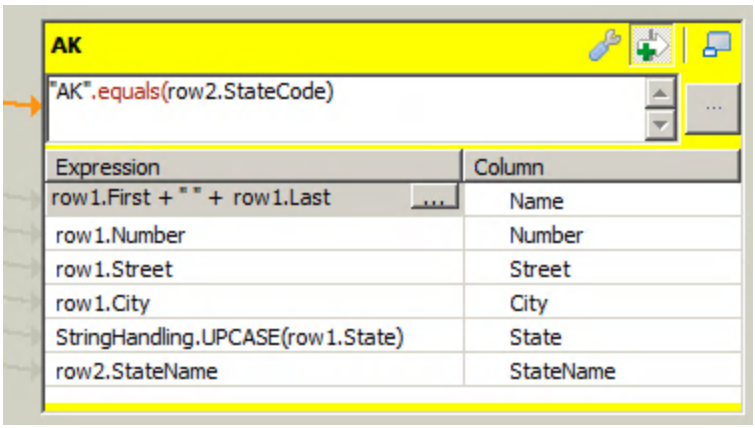
9. If it is not already opened, click the **Activate / unactivate the expression filter** Icon on the top right of the output schema **CappedOut** and copy the expression *"CA".equals(row2.StateCode):*



10. Click the **Activate / unactivate the expression filter** Icon on the top right of the output schema **NY** and paste the expression to it. Then change it to *"NY".equals(row2.StateCode)* :



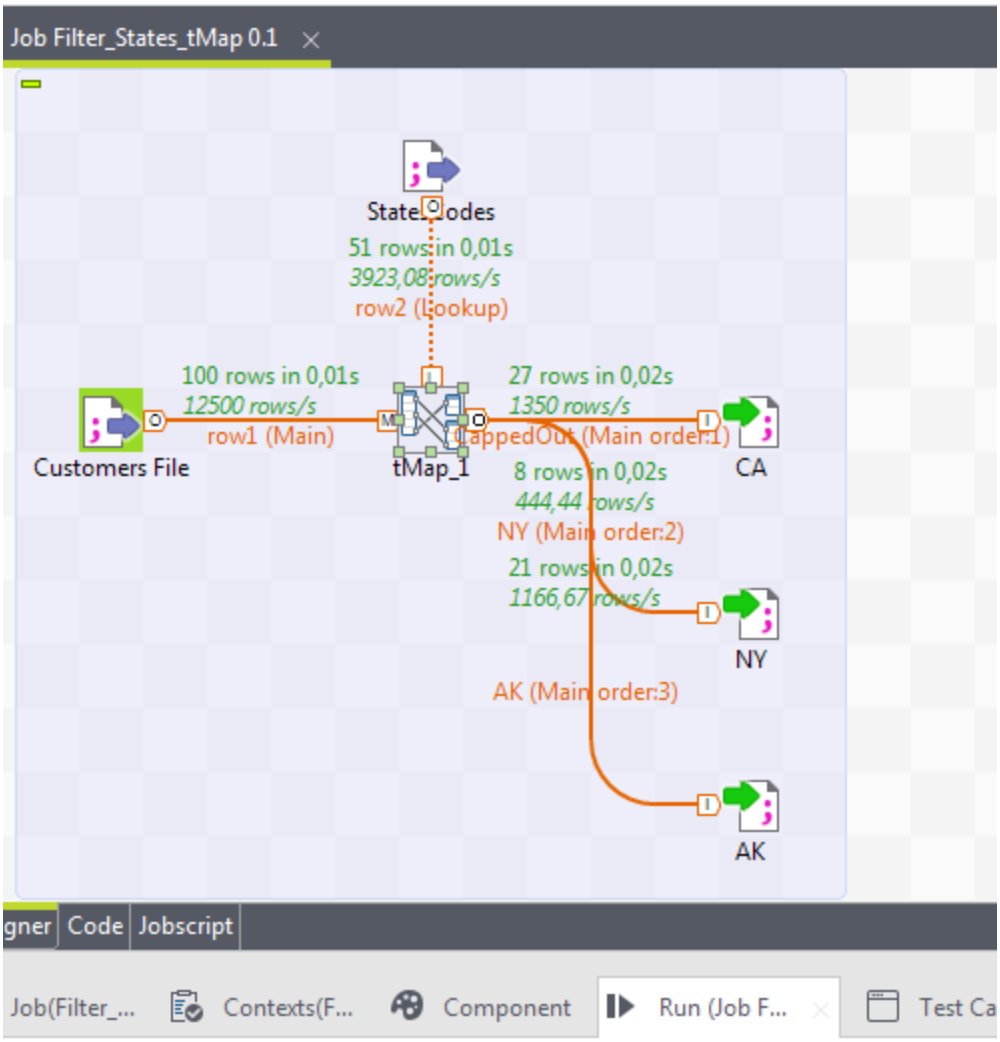
11. Apply the same changes for the output **AK**:

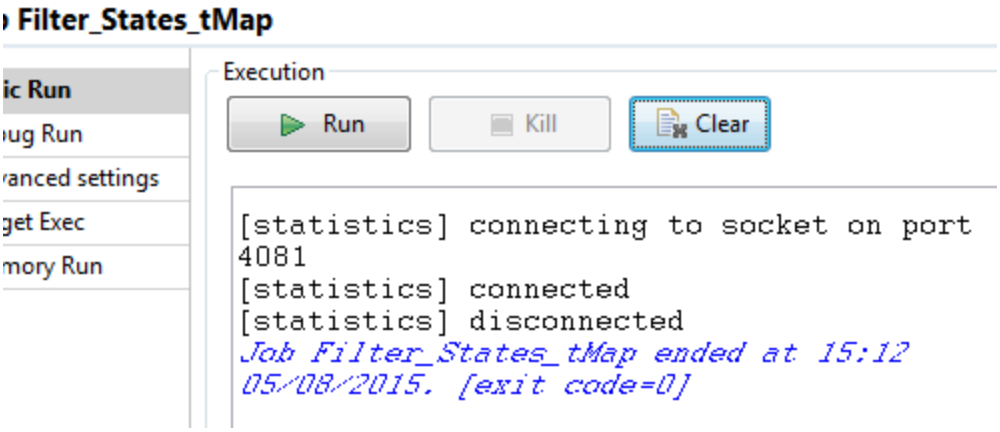


12. Click **OK** to save your **tMap** editor changes, and then save the Job.

Run Job

1. Run the Job and notice how the results are written to different output files based on the filters put in place:





2. Navigate to and display the output files in **C:/StudentFiles/** to verify the output. The **NY_Out.csv** file is displayed here:

NY_Out.csv | AK_Out.csv | CA_Out.csv

```
1 Name,Number,Street,City,State,StateName
2 Theodore Clinton,12292,San Marcos,Bismarck,NY, New York
3 Calvin Adams,52386,Lake Tahoe Blvd.,Montgomery,NY, New York
4 Martin Johnson,32782,Padre Boulevard,Springfield,NY, New York
5 Abraham Tyler,89260,Carpinteria North,Annapolis,NY, New York
6 Herbert Coolidge,65573,Cabrillo Highway,Salt Lake City,NY, New York
7 Gerald Eisenhower,12903,Jones Road,Albany,NY, New York
8 Abraham Wilson,83959,Hutchinson Rd,Nashville,NY, New York
9 Rutherford Harding,46743,Steele Lane,Boston,NY, New York
```

Next

You have now completed this lesson. It's now time to Wrap-Up.

Wrap-Up

In this lesson, you extended the previous Job to generate several output files depending on the State Code. You learned how to filter data based on the value of a column and you created several output tables while using the **tMap** component. You used the **Auto map!** function in the **tMap** component to map the new output tables. You copied and pasted filter expressions and changed the filter value. You defined two new output files for the new output tables. Multiple filters wrote to the different output files.

Next step

Congratulations! You have successfully completed this lesson. To save your progress, click **Check your status with this unit** below. To go to the next lesson, on the next screen, click **Completed. Let's continue >**.