

tExtractKeyValue



tExtractKeyValue properties

Component family	Processing/Fields	
Function	tExtractKeyValue generates multiple columns from a given column containing key-value pairs.	
Purpose	tExtractKeyValue allows you to extract “values” from “keys” in a string. The “keys” are taken as fields and “values” as its contents. (This component allows empty key-value pairs in the input).	
Basic settings	<i>Field</i>	Select an incoming field from the Field list to extract.
	<i>Schema type and Edit Schema</i>	<p>A schema is a row description, i.e., it defines the number of fields that will be processed and passed on to the next component. The schema is either built-in or remote in the Repository.</p> <p>Click Edit Schema to make changes to the schema. Note that if you make changes, the schema automatically becomes built-in.</p> <p>Click Sync columns to retrieve the schema from the previous component connected to tExtractKeyValue.</p>
		Built-in: The schema will be created and stored locally for this component only.
		Repository: The schema already exists and is stored in the Repository, hence can be reused in various projects and Job flowcharts.
	<i>Field Separator</i>	It is used to separate key-value pairs in each row. It has to be a Regex Expression.
	<i>Key-Value Separator</i>	It separates a Value from a key. It has to be a Regex Expression.
	<i>Key Name</i>	Column: This field is automatically populated with the columns defined in the schema that you propagated.
		Key: This field should contain the “key” names as of in the selected <i>Field</i> .
		Trim Value: Select this check box to remove leading and trailing whitespaces from defined columns.
	<i>Die on error</i>	Select this check box to stop the execution of the Job when an error occurs. Clear the check box to skip the row on error and complete the process for error-free rows.
Advanced settings	<i>tStatCatcher Statistics</i>	Select this check box to gather the processing metadata at the Job level as well as at each component level.
Usage	This component handles flow of data therefore it requires input and output components. It allows you to extract data from a field containing key-value pairs , using a Row >	

	<p>Main link.Use this component to split a <i>Field</i> containing “key-value” pairs on the basis of “key”. It extracts the “values” from the “keys”.</p> <p>The “keys” are taken as fields and “values” as its contents.</p> <p>(This component even allows empty key-value pairs in the input).</p> <p>For further information, please see the section called “Scenario 2: Extracting values from a field containing key value pairs”.</p>
Author	Mahesh M.Pillai

Scenario 1: Extracting “values” from “keys” in a field.

The following scenario creates a three -component Job, which aims at reading a file wherein there is a field containing multiple key-value pairs .It extracts the “value” from the “keys ” from the column and displays the output in the Run log console.

The contents of the Input Text File is shown below:

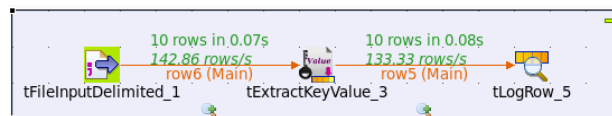
```

EmployeeID***EmployeeName***EmployeeAddress
554304***Abraham***Street=Bowles Avenue;State=Texas;StateID=TX;City=Oklahoma City
554085***Richard***Street=North Atherton Street;State=Maryland;StateID=VA;City=Charleston
559235***John***Street=San Ysidro Blvd;State=Vermont;StateID=ME;City=Denver
555807***Gerald***Street=Camelback Rd;State=Georgia;StateID=IN;City=Sacramento
551058***Martin***Street=Monroe Street;State=Virginia;StateID=GA;City=Albany
554040***Chester***Street=East Calle Primera;State=Michigan;StateID=IL;City=Juneau
550275***John***Street=Via Real;State=Wyoming;StateID=ME;City=Boston
555129***Lyndon***Street=Moreno Drive;State=Florida;StateID=UT;City=Honolulu
554812***Thomas***Street=Carpinteria North;State=Michigan;StateID=AR;City=Augusta
559263***Benjamin***Street=Bailard Avenue;State=Texas;StateID=WA;City=Bismarck

```

Field Containing Key Value Pairs

- The First Line is the Header.
- Here the *Field* “EmployeeAddress” contains the Key-Value Pairs.
- Here the keys are “Street”, “State”, “StateID”, “City”.



- Drop the following components from the Palette onto the design workspace: **tFileInputDelimited**, **tExtractKeyValue**, and **tLogRow**.
- Connect the three components using **Row Main** links.
- In the design workspace, select **tFileInputDelimited**.

- Click the **Component** tab to define the basic settings for **tFileInputDelimited**.
- In the **Basic settings** view, set **Property Type** to **Built-In**.
- Click the three-dot [...] button next to the **File Name** field to select the path to the input file.



The **File Name** field is mandatory.

Component **tFileInputDelimited_1**

Basic settings

Property Type: Built-In

Advanced settings: "When the input source is a stream or a zip file, footer and random shouldn't be bigger than 0."

Dynamic settings: File name/Stream: /home/550778/Desktop/new file1.txt

View: Row Separator: \n Field Separator: ***

Documentation: CSV options: Header: 1 Footer: 0 Limit:

Schema: Built-In Edit schema

☒ Skip empty rows ☐ Uncompress as zip file ☐ Die on error

- The input file used in this scenario is called new *file1.txt* . It is a text file that holds three columns: *EmployeeID*, *EmployeeName*, and *EmployeeAddress*.
- Here the *Field Separator* is "***".
- Fill in all other fields as needed. In this scenario, the **header** is 1 and the **footer** is not set and there is no **limit** for the number of processed rows
- Click Edit schema to describe the data structure of this input file. In this scenario, the schema is made of the three columns, *EmployeeID*, *EmployeeName*, and *EmployeeAddress* .
- In the design workspace, select **tExtractKeyValue**.
- Click the Component tab to define the basic settings for **tExtractKeyValue**.
- From the **Field to be processed** list, select the column to split, *EmployeeAddress* in this scenario.

Component **tExtractKeyValue_3**

Basic settings

Field to be processed: EmployeeAddress

Advanced settings: Schema: Built-In Edit schema Sync columns

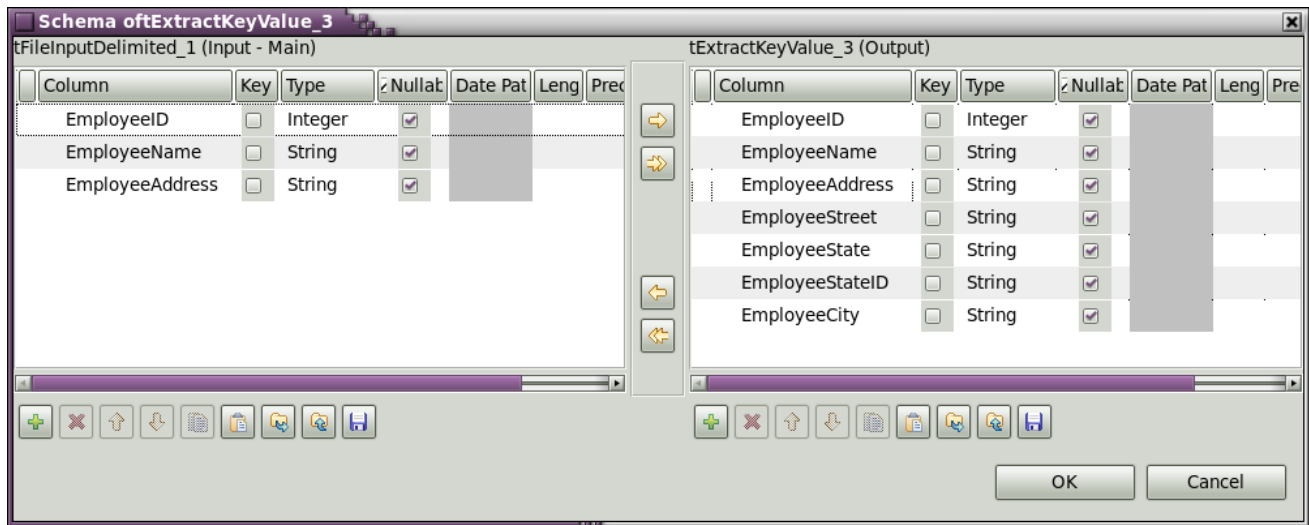
Dynamic settings: Field Separator(Regex): , Key-Value Separator(Regex): =

View: Key Name

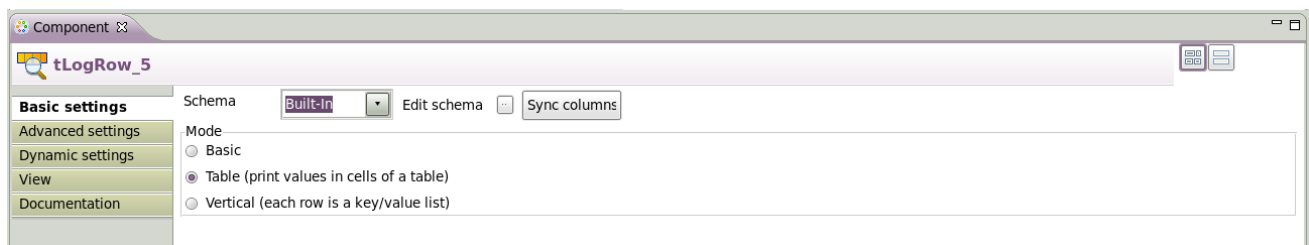
Column	Key	Trim Value
EmployeeID		<input checked="" type="checkbox"/>
EmployeeName		<input checked="" type="checkbox"/>
EmployeeAddress		<input checked="" type="checkbox"/>
EmployeeStreet	"Street"	<input checked="" type="checkbox"/>
EmployeeState	"State"	<input checked="" type="checkbox"/>
EmployeeStateID	"StateID"	<input checked="" type="checkbox"/>
EmployeeCity	"City"	<input checked="" type="checkbox"/>

Documentation: ☒ Die on error

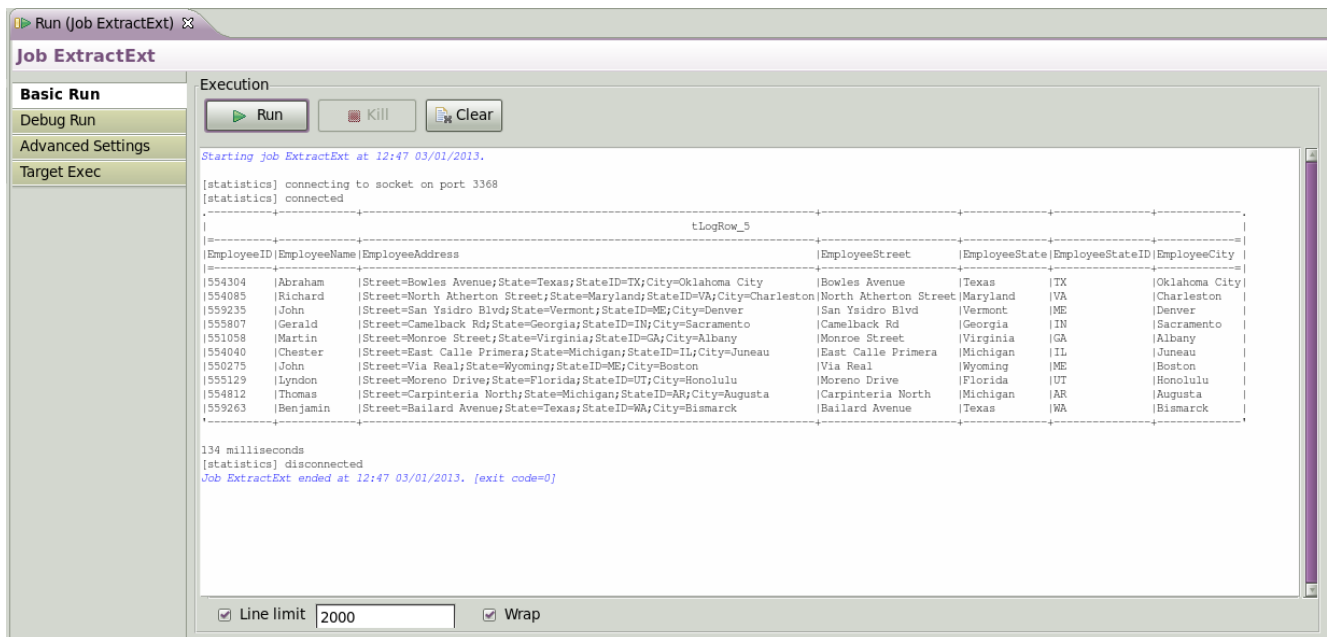
- Set the **Schema** as either a local (Built-in) or a remotely managed (Repository) to define the data to pass on to the **tLogRow** component.
- You can load and/or edit the schema via the **Edit Schema** function.
- Click Edit schema to describe the data structure of this processing component. In the present scenario we have in total a maximum of 4 key-value pairs in a field. You may define them as shown below.



- The column name can be any valid name.
- Define the **Field separator** used to delimit key-value pairs in a row (Here it is “;”). Also define the **Key-Value Separator** (Here it is “=”).
- These fields should be in Regex Patterns.
- Now the **Key Name** Field should be filled.
Column field is automatically populated with the columns defined in the schema that you propagated.
The **Key** field corresponds to the keys in the *Field* that we have selected.
In the current scenario the keys are “Street”, “State”, “StateID”, “City”.
- The **Trim Value** field can be used to remove leading and trailing whitespaces from the “values” corresponding to the fields.
- Die on Error can be set as per need.(Select this check box to stop the execution of the Job when an error occurs. Clear the check box to skip the row on error and complete the process for error-free rows.)
- Select the tLogRow and define the Field separator to use for the output display.



- Go to **Run** tab, and click on Run to execute the Job.
The extracted fields are displayed on the Run log as defined in both components .
The data from the input schema will be automatically propagated into the Output by default.



Scenario 2: Extracting “values” from “keys” in a field with

- Key-Value Pairs randomly distributed within the field
- Key-Value Pairs missing in the field

The following scenario describes the Flexibility this component offers:

- It allows missing key value pairs.
- It allows missing values.
- It allows random distribution of key-value pairs within the field.

The following scenario depicts a two-component Job, which aims at extracting the “value” from the “keys ” from a field (containing multiple key-value pair) and displays the output in the Run log console.

The above mentioned(*in italics*) conditions are created in the input file.

The Input File (with Annotations) is shown below:

```

EmployeeID***EmployeeName***EmployeeAddress
554304***Abraham***Street=Bowles Avenue;State=Texas;StateID=TX;City=Oklahoma City
554085***Richard***Street=North Atherton Street;State=Maryland;StateID=VA;City=Charleston
559235***John***Street=San Ysidro Blvd;State=Vermont;City=Denver
555807***Gerald***Street=Camelback Rd;State=Georgia;StateID=IN;City=Sacramento
551058***Martin***Street=;State=Virginia;StateID=GA;City=Albany
554040***Chester***Street=East Calle Primera;State=Michigan;StateID=IL;City=Juneau
550275***John***Street=Via Real;State=Wyoming;StateID=ME;City=Boston
555129***Lyndon***Street=Moreno Drive;State=Florida;StateID=UT;City=Honolulu
554812***Thomas***Street=Carpinteria North;State=Michigan;StateID=AR;City=Augusta
559263***Benjamin***City=Bismarck;State=Texas;Street=Bailard Avenue;StateID=WA

```

Missing Value

Missing Key-Value Pair

Randomly arranged Key-Value Pairs

The Job is the same as in Scenario1.
The output with annotations is shown below:

Run (Job ExtractExt)

Job ExtractExt

Basic Run

Debug Run

Advanced Settings

Target Exec

Execution

Run Kill Clear

Starting job Job ExtractExt at 14:21 03/01/2013.

[statistics] connecting to socket on port 3468

[statistics] connected

EmployeeID	EmployeeName	EmployeeAddress	EmployeeStreet	EmployeeState	EmployeeStateID	EmployeeCity
554304	Abraham	Street=Bowles Avenue;State=Texas;StateID=TX;City=Oklahoma City	Bowles Avenue	Texas	TX	Oklahoma City
554085	Richard	Street=North Atherton Street;State=Maryland;StateID=VA;City=Charleston	North Atherton Street	Maryland	VA	Charleston
559235	John	Street=San Ysidro Blvd;State=Vermont;City=Denver	San Ysidro Blvd	Vermont		Denver
555807	Gerald	Street=Camelback Rd;State=Georgia;StateID=IN;City=Sacramento	Camelback Rd	Georgia	IN	Sacramento
551058	Martin	Street=;State=Virginia;StateID=GA;City=Albany		Virginia	GA	Albany
554040	Chester	Street=East Calle Primera;State=Michigan;StateID=IL;City=Juneau	East Calle Primera	Michigan	IL	Juneau
550275	John	Street=Via Real;State=Wyoming;StateID=ME;City=Boston	Via Real	Wyoming	ME	Boston
555129	Lyndon	Street=Moreno Drive;State=Florida;StateID=UT;City=Honolulu	Moreno Drive	Florida	UT	Honolulu
554812	Thomas	Street=Carpinteria North;State=Michigan;StateID=AR;City=Augusta	Carpinteria North	Michigan	AR	Augusta
559263	Benjamin	City=Bismarck;State=Texas;Street=Bailard Avenue;StateID=WA	Bailard Avenue	Texas	WA	Bismarck

193 milliseconds

[statistics] disconnected

Job ExtractExt ended at 14:21 03/01/2013. [exit code=0]

☒ Line limit 2000 ☒ Wrap

Inspite of the input being in an unordered format, the output is in the correct format.