



Talend User Component tJsonNormalize



Purpose

<http://www.cimt-ag.de>

This component extracts all value attributes (attributes which have simple values and does not point to embedded documents) and creates for each attribute a output dataset. The output is an outer join with the input.

Talend-Integration

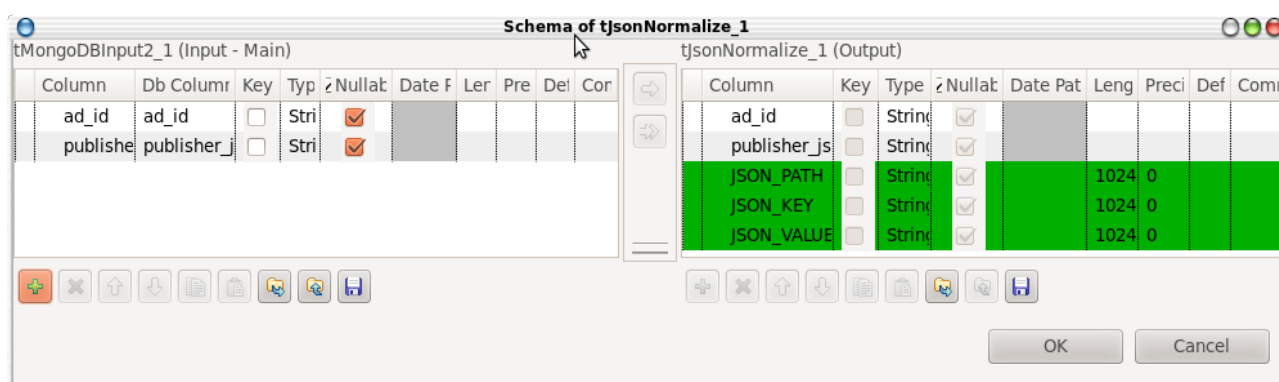
This component can be found in the palette under Processing->Fields
This component provides an output flow and several return values.

Parameters

Property	Content
JSON to normalize	This schema field must contain the JSON content as String or as BSONObject from which the attributes have to extract.
Allow empty output	In case the extraction method does not found attributes (can happens if attributes will be filtered) an output dataset will be created with null values in the extractor output.
Ignore attributes with empty values	Default=true All attributes, which do have an empty or null value, will be ignored.
Start path for extracting values	Specifies the path (dot delimited) from which the extraction of attributes has to start.
Delimiter for path representation	The path to the current found attribute can be delimited.
Exclude fields	Specifies fields, which should be ignored. A String will be expected.

Return values

Return value	Content
ERROR_MESSAGE	Last error message
NB_LINE	Total amount of output lines (number of returned attributes)
NB_PARSED	Number of parsed JSON documents (should be equal to the amount of input lines).

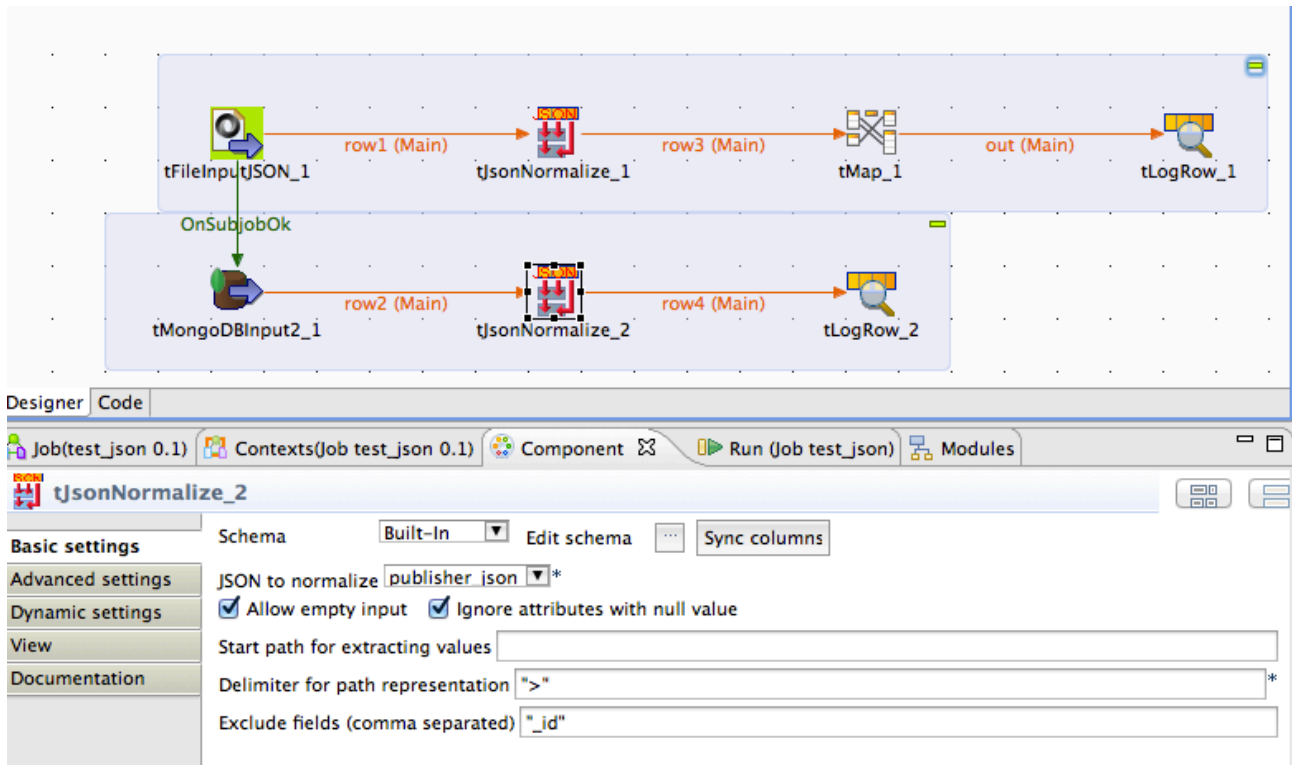


Schema output

Schema column	Content
JSON_PATH	Path of the current json attribute.
JSON_KEY	Key of the current attribute.
JSON_VALUE	Value (String) of the current attribute.

Scenario

Normalize the given JSON documents, read from a mongo db.



In this scenario, the JSON content in schema column `publisher_json` is the input for extraction. In this JSON document the root attributes will be ignored and the extraction starts with the embedded document referenced by the attribute `promotions`. The attribute `_id` will be ignored.