

DW y ETL

TALEND OPEN STUDIO
TRANSFORMACIONES



KEEPCODING

Tech School



ÍNDICE



1. GENERADOR DE DATOS
2. GENERADORES DE FLUJOS
3. DATA QUALITY
4. AGRUPACIÓN DE DATOS

GENERADOR DE DATOS

tRowGenerator: Generar datos automáticamente

Talend Open Studio for Data Integration - tRowGenerator - tRowGenerator_1

Schema		Functions	
Column	Type	Functions	Parameters
First_name	String	TalendDataGenerator.getFirstName()	
Last_name	String	TalendDataGenerator.getLastName()	
newColumn	String	TalendString.getAsciiRandomString(int) length = > 6 ;	

StringHandling.SQUOTE(String)
StringHandling.STR(String,int)
StringHandling.TRIM(String)
StringHandling.UPCASE(String)
TalendDataGenerator.getFirstName()
TalendDataGenerator.getLastName()
TalendDataGenerator.getUsCity()
TalendDataGenerator.getUsState()
TalendDataGenerator.getUsStateId()
TalendDataGenerator.getUsStreet()
TalendDate.addDate(String,String,int,Date)
TalendDate.formatDate(String,Date)
TalendDate.formatDateInUTC(String,Date)
TalendDate.formatDateLocale(String,Date,String)
TalendDate.getDate(String)
TalendString.checkCDATAForXML(String)


Function parameters Preview

getAsciiRandomString : Return a randomly generated String

Parameter	Value	Comment
length	6	length of the String to re...

GENERADOR DE DATOS

tFixedFlowInput: Generar datos manualmente

 **tFixedFlowInput_1**

Basic settings

Advanced settings

Dynamic settings

View

Documentation

Schema Type

Built-In

Edit schema

Number of rows

1

Mode

☐ Use Single Table

☒ Use Inline Table

Inline Table

Nombre	Puntos
"Pepe"	10
"Antonio"	25

+

×

↑

↓

📄

📅

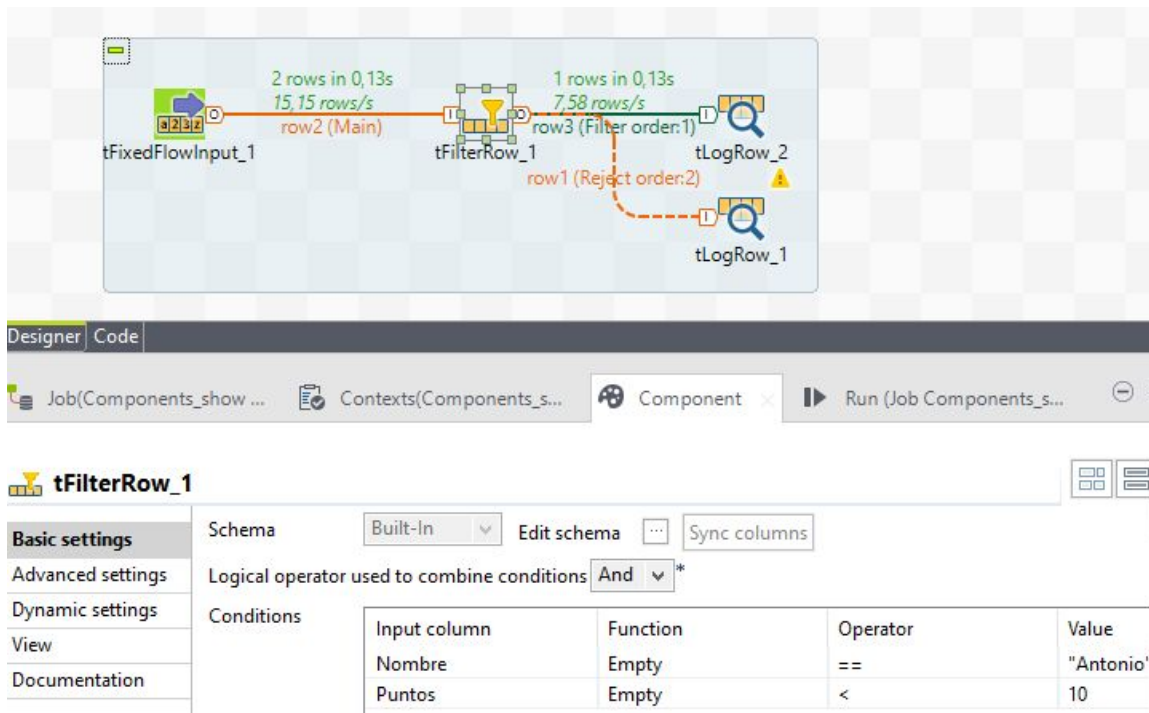
☐ Use Inline Content(delimited file)

GENERADOR DE FLUJOS

tFilterRow: Filtro de datos

Flujo de salida «Filter», es decir, el flujo que contiene los datos que cumplen la condición.

Flujo de salida «Reject», el flujo rechazado que no cumple la condición





GENERADOR DE FLUJOS

tReplicate: Replicar el flujo.

Este componente es utilizado cuando por motivo de proceso necesitamos que el flujo se replique, por ejemplo para hacer una misma salida de datos a una tabla de una base de datos y también a un fichero.



GENERADOR DE FLUJOS

tUnite: Unir dos flujos

Este componente hace lo contrario al tReplicate, ya que une dos flujos de datos que tengan un mismo esquema. No combina registros, sino que simplemente junta un flujo de datos debajo del otro, tipo UNION de SQL.

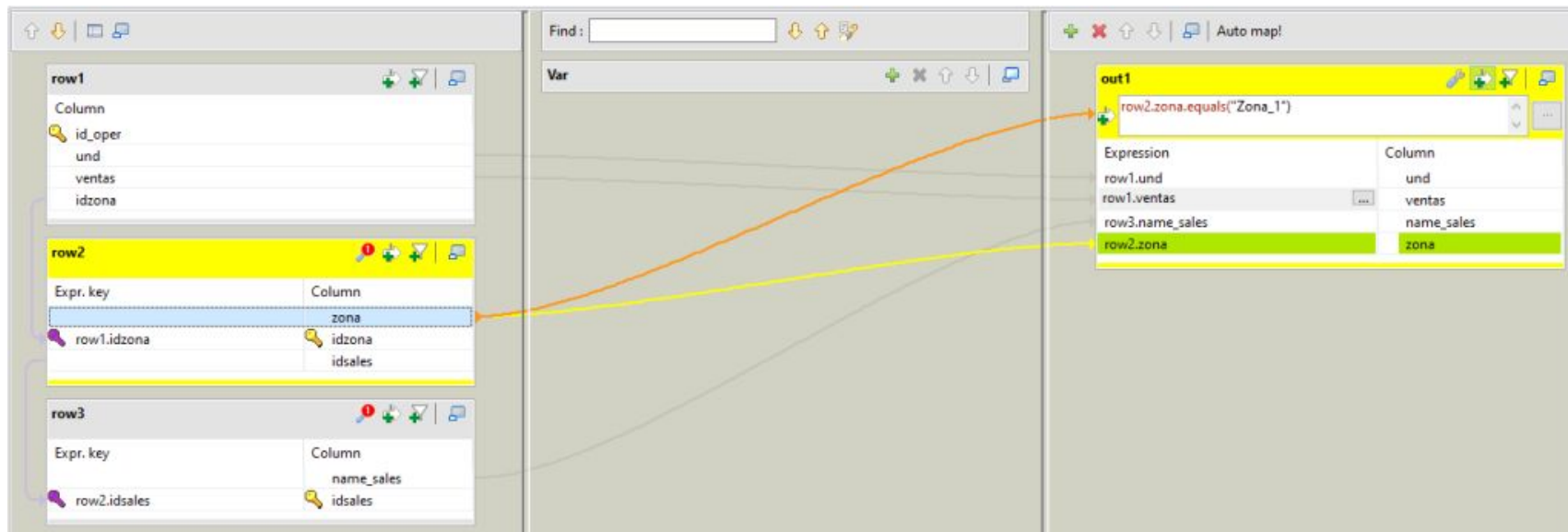
tJoin: Join de flujos.

Lo cierto es que después de la potencia y optimización de proceso y otras capacidades que nos permite el componente tMap (que veremos a continuación), tJoin queda bastante canibalizado por el todo poderoso componente tMap y su uso es menor.

GENERADOR DE FLUJOS

tMap: Un comodín que une con JOIN, FILTER y GENERADOR DE EXPRESIONES.

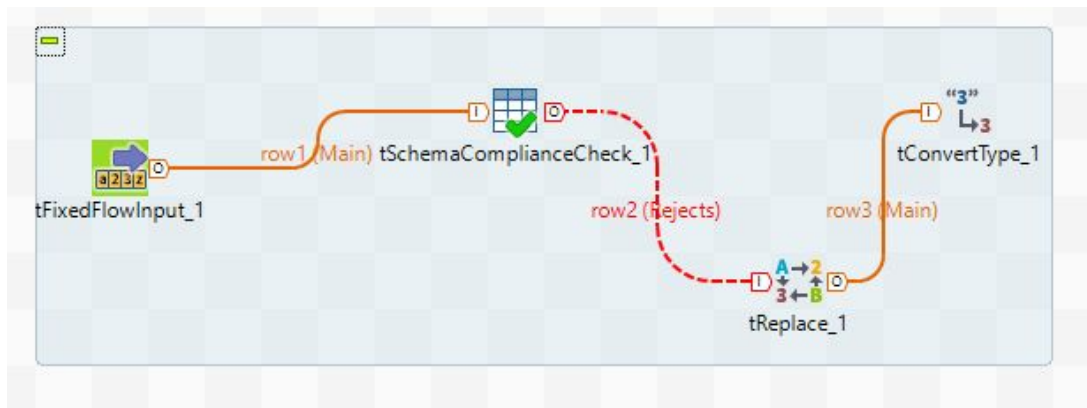
La joya de la corona de Talend. De manera muy intuitiva conecta cualquier número de flujos de entrada y genera cualquier número de flujos de salida. El tMap, tiene un asistente gráfico que permite una configuración muy sencilla.



DATA QUALITY

tReplace permite buscar una cadena de texto y sustituirla por otra.

tConvertType nos permite establecer un esquema y mediante la opción «**Autocast**» convertir los tipos de origen al «datatype» establecido en el esquema.



El ejemplo muestra cómo el componente de **tSchemaComplianceCheck** rechaza los registros que no cumplen. El componente **tReplace** sustituye los valores para que cumpla («dos» por «2»). Por último el componente **tConvertType**, cambia el datatype (En el ejemplo de «string» a «integer».)

AGRUPACIÓN DE DATOS

Componente tAggregate: Nos permite agrupar registros a partir de uno o varios campos y usar funciones de agregación de forma idéntica al **GROUP BY de SQL**. En la parte superior indicamos los campos que queremos agregar y en la inferior los campos con las funciones de agregación.

Job(aggregate 0.1) Contexts(aggregate) Component Run (Job aggregate)

tAggregateRow_1

Basic settings

Schema: Built-In Edit schema Sync columns

Group by

Output column	Input column position
Nombre	Nombre

Operations

Output column	Function	Input column position	Ignore null values
ventas	sum	ventas	<input type="checkbox"/>
uds	max	uds	<input type="checkbox"/>



KEEPCODING

Tech School

Madrid | Barcelona | Bogotá

Jon Pineda Galvez

jpinedagalvez@gmail.com

<https://www.linkedin.com/in/jonathanpg>