talendJobs

Generated by Qlik Talend Cloud Enterprise Edition

Project Name	DemoPrep	GENERATION DATE	Mar 3, 2025, 4:08:10 PM
AUTHOR	Dae Wah, Daniel	Qlik Talend Cloud VERSION	8.0.1.20250126_0750-patch

Summary

Project Description

Properties	Values
Name	DemoPrep
Language	java
Description	demo

Description

Properties	Values
Name	ScrapeWeb
Author	Dae Wah, Daniel
Version	0.1
Purpose	
Status	
Description	
Creation	Feb 7, 2025, 2:35:29 PM
Modification	Feb 7, 2025, 3:14:38 PM

Preview Picture
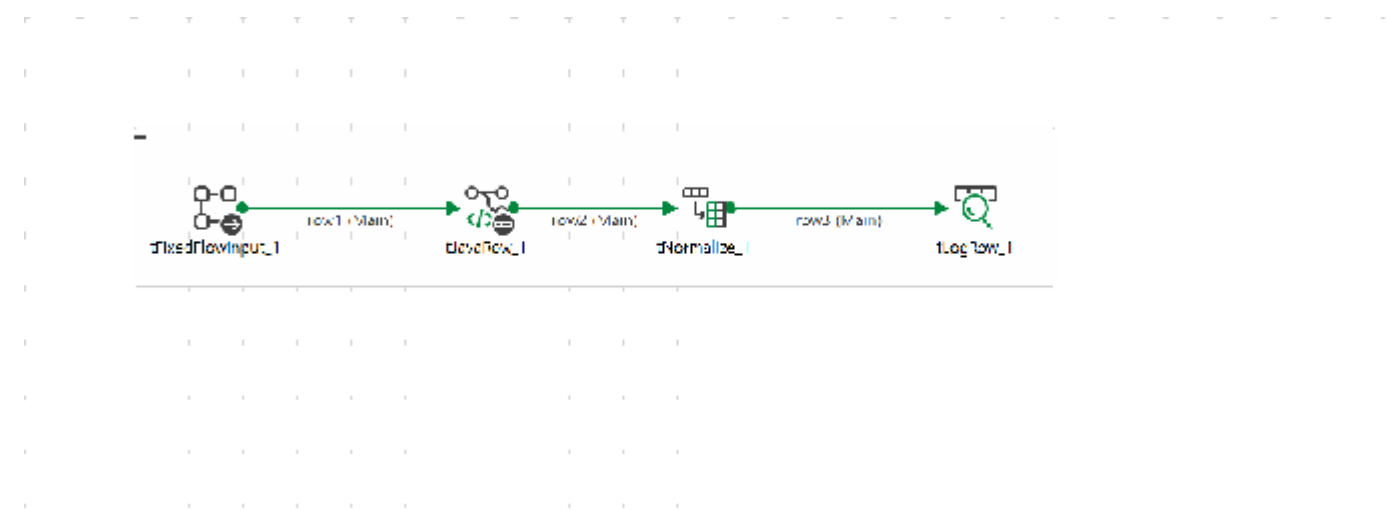
Settings

Extra settings

Name	Value
COMP_DEFAULT_FILE_DIR	
Multi thread execution	false
Implicit tContextLoad	false

Status & Logs

Name	Value
Use statistics (tStatCatcher)	false
Use logs (tLogCatcher)	false
Use volumetrics (tFlowMeterCatcher)	false
On Console	false
On Files	false
On Databases	false
Catch components statistics	false
Catch runtime errors	true
Catch user errors	true
Catch user warnings	true

Context List

ContextDefault

Name	Prompt	Need Prompt?	Type	Value	Source

Component List

Component Name	Component Type
tFixedFlowInput_1	tFixedFlowInput
tJavaRow_1	tJavaRow
tLogRow_1	tLogRow
tNormalize_1	tNormalize

Components Description

Component   tFixedFlowInput

	UNIQUE NAME	tFixedFlowInput_1	INPUT(S)	none
	LABEL	__UNIQUE_NAME__	OUTPUT(S)	tJavaRow_1

Component Parameters:

Properties	Values
Unique Name	tFixedFlowInput_1
Component Name	tFixedFlowInput
Version	0.102 (ALPHA)
Family	Misc
Start	true
Startable	true
SUBTREE_START	true
END_OF_FLOW	false
Activate	true
DUMMY	false
tStatCatcher Statistics	false
Help	org.talend.help.tFixedFlowInput
Update components	true
IREPORT_PATH	
JAVA_LIBRARY_PATH	C:\Talend\8.0.1\studio\configuration\lib\java
Subjob color	
Subjob header color	
Enable parallel execution	false
Number of parallel executions	2
PARALLELIZE_KEEP_EMPTY	false
REPOSITORY_ALLOW_AUTO_SWITCH	true
Schema	Built-In
Number of rows	1
Use Single Table	true
Values	[{SCHEMA_COLUMN=url, VALUE="https://www.bbc.com/news"}, {SCHEMA_COLUMN=searchfor, VALUE="Trump"}]
Use Inline Table	false
Inline Table	[]

Use Inline Content(delimited file)	false
Row Separator	"\n"
Field Separator	";"
Content	
Label format	__UNIQUE_NAME__
Hint format	__UNIQUE_NAME__ __COMMENT__
Connection format	row
Show Information	false
Comment	
Use an existing validation rule	false
Validation Rule Type	

Schema for tFixedFlowInput_1 :

Column	Key	Type	Length	Precision	Nullable	Comment
url	false	String			true	
searchfor	false	String			true	

Original Function Parameters:
Component   tJavaRow

	UNIQUE NAME	tJavaRow_1	INPUT(S)	tFixedFlowInput_1
	LABEL	__UNIQUE_NAME__	OUTPUT(S)	tNormalize_1

Component Parameters:

Properties	Values
Unique Name	tJavaRow_1
Component Name	tJavaRow
Version	0.101 (ALPHA)
Family	Custom_Code
Startable	false
SUBTREE_START	false
END_OF_FLOW	false
Activate	true
DUMMY	false
tStatCatcher Statistics	false
Help	org.talend.help.tJavaRow
Update components	true
IREPORT_PATH	
JAVA_LIBRARY_PATH	C:\Talend\8.0.1\studio\configuration\lib\java
Subjob color	
Subjob header color	
Enable parallel execution	false
Number of parallel executions	2
PARALLELIZE_KEEP_EMPTY	false
REPOSITORY_ALLOW_AUTO_SWITCH	true
Schema	Built-In
Code	// code sample: // // multiply by 2 the row identifier // output_row.id = input_row.id * 2; // // lowercase the name // output_row.name = input_row.name.toLowerCase(); output_row.extraction = routines.webscrape.getweb(input_row.url, input_row.searchfor);
Import	//import java.util.List;
Label format	__UNIQUE_NAME__
Hint format	__UNIQUE_NAME__ __COMMENT__
Connection format	row
Show Information	false
Comment	
Use an existing validation rule	false
Validation Rule Type	

Schema for tJavaRow_1 :

Column	Key	Type	Length	Precision	Nullable	Comment
url	false	String			true	
extraction	false	String			true	

Original Function Parameters:
Component   tLogRow

	UNIQUE NAME	tLogRow_1	INPUT(S)	tNormalize_1
	LABEL	__UNIQUE_NAME__	OUTPUT(S)	none

Component Parameters:

Properties	Values
Unique Name	tLogRow_1
Component Name	tLogRow
Version	0.101 (ALPHA)
Family	Logs_Errors
Startable	false
SUBTREE_START	false
END_OF_FLOW	true
Activate	true
DUMMY	false
tStatCatcher Statistics	false
Help	org.talend.help.tLogRow
Update components	true
IREPORT_PATH	
JAVA_LIBRARY_PATH	C:\Talend\8.0.1\studio\configuration\lib\java
Subjob color	
Subjob header color	
Enable parallel execution	false
Number of parallel executions	2
PARALLELIZE_KEEP_EMPTY	false
REPOSITORY_ALLOW_AUTO_SWITCH	true
Schema	Built-In
Basic	false
Table (print values in cells of a table)	true
Vertical (each row is a key/value list)	false
Print unique name	true
Print label	false
Print unique name and label	false
Field Separator	"|"
Print header	false
Print component unique name in front of each output row	false
Print schema column name in front of each value	false
Use fixed length for values	false
Lengths	[{SCHEMA_COLUMN=url, LENGTH=10}, {SCHEMA_COLUMN=extraction, LENGTH=10}]
Print content with log4j	true
Min column number of optimize code	100
Label format	__UNIQUE_NAME__
Hint format	__UNIQUE_NAME__ __COMMENT__
Connection format	row
Show Information	false
Comment	

Use an existing validation rule	false
Validation Rule Type	

Schema for tLogRow_1 :

Column	Key	Type	Length	Precision	Nullable	Comment
url	false	String			true	
extraction	false	String			true	

Original Function Parameters:
Component   tNormalize

	UNIQUE NAME	tNormalize_1	INPUT(S)	tJavaRow_1
	LABEL	__UNIQUE_NAME__	OUTPUT(S)	tLogRow_1

Component Parameters:

Properties	Values
Unique Name	tNormalize_1
Component Name	tNormalize
Version	0.101 (ALPHA)
Family	Processing/Fields
Startable	false
SUBTREE_START	false
END_OF_FLOW	false
Activate	true
DUMMY	false
tStatCatcher Statistics	false
Help	org.talend.help.tNormalize
Update components	true
IREPORT_PATH	
JAVA_LIBRARY_PATH	C:\Talend\8.0.1\studio\configuration\lib\java
Subjob color	
Subjob header color	
Enable parallel execution	false
Number of parallel executions	2
PARALLELIZE_KEEP_EMPTY	false
REPOSITORY_ALLOW_AUTO_SWITCH	true
Schema	Built-In
Column to normalize	extraction
Item separator	";"
Get rid of duplicate rows from output	false
Use CSV parameters	false
Escape mode	ESCAPE_MODE_DOUBLED
Text enclosure	"\""
Discard the trailing empty strings	false
Trim resulting values	false
Label format	__UNIQUE_NAME__
Hint format	__UNIQUE_NAME__ __COMMENT__
Connection format	row
Show Information	false
Comment	
Use an existing validation rule	false
Validation Rule Type	

Schema for tNormalize_1 :

Column	Key	Type	Length	Precision	Nullable	Comment
url	false	String			true	

extraction	false	String			true	

Original Function Parameters: