

Handling tasks in Talend Data Stewardship

Overview

In this module, you learned how to create a merging campaign in Talend Data Stewardship. Using a Talend Job, you created tasks for the campaign. Then, acting as a data steward, you resolved and validated the merging tasks in Talend Data Stewardship.

Key steps

1 Creating a merging campaign

A campaign contains a list of tasks that have the structure defined by a data model and allows you to merge the duplicate tasks.

To create a campaign, you connect Talend Cloud Portal using your cloud account credentials, launch Talend Data Stewardship application and provide the following:

- Campaign Name
- Campaign Type (merging)
- **Roles** defining groups of data stewards
- **Data model** (a data model can be imported)
- Workflow defining the steps for resolving the campaign tasks and roles dedicated to these steps.

This is a sample campaign defined with a DATA STEWARD role (containing the Adam Brown steward), and the data model Customers with its attributes.

The screenshot displays the 'EDIT CAMPAIGN' interface in Talend Data Stewardship. The sidebar on the left indicates the current step is '2/ Roles'. The main content area shows the 'ROLES' section with a 'DATA STEWARD' field containing 'Adam Brown' and an 'ADD A ROLE' button. Below this, the '3/ Data Model' section is visible, showing a table for 'Customers' with columns for 'ALL ATTRIBUTES', 'IDENTIFIER', 'DATA STEWARD', 'FIRST NAME', 'LAST NAME', and 'DATA STEWARD'. Each row has a 'SURVIVORSHIP RULE' dropdown set to 'First valid'.

2

Creating merging tasks in a Talend Job

In a Talend Job, you can use the **tMatchGroup** component for matching records, and **tDataStewardshipTaskOutput** component to create tasks.

The **tMatchGroup** component allows you to define matching rules that will be applied to match records and group them by similarity.

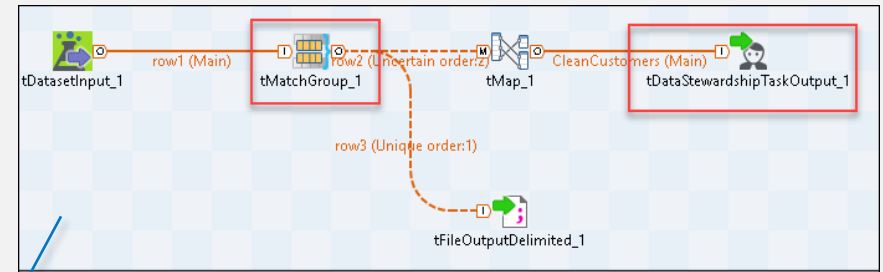
The output rows of **tMatchGroup** component can be used as input rows of the **tDataStewardshipTaskOutput** component.

For each of input rows, the **tDataStewardshipTaskOutput** component will create tasks for a campaign in Data Stewardship cloud application.

In this Job, the cleaned **tMatchGroup** output rows are used as input rows of the **tDataStewardshipTaskOutput** component to create tasks in Data Stewardship.

Finally, each row handled by the **tDataStewardshipTaskOutput** component corresponds to a task in Talend Data Stewardship. Tasks can be assigned to data stewards.

In this screenshot, three tasks in a campaign have been created in Data Stewardship after executing the **tDataStewardshipTaskOutput** component.



Brown							State: New	Unassigned	↩
3/3									
Search and filter									
	Identifier *	First Name *	Last Name *	Email	Job Title	Company			
	text	text	text	email	text	text			
1	687811	Denise	REYNOLDS	dreynoldsqs@...	Engineer	Blogpad			
2	908517	Roy	TURNER	rturnerka@wik...	Senior Sales ...	Thoughtstorm			
3	656555	Martha	PEREZ	mpereznr@mai...	VP Sales	Omba			

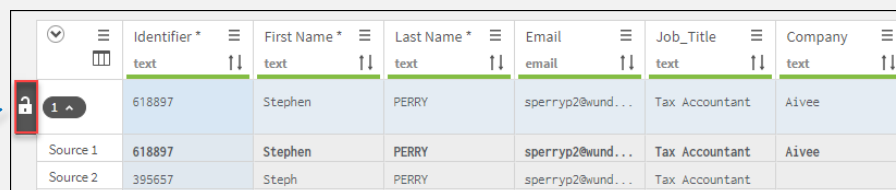
3

Resolving and validating merging tasks

Tasks are assigned to data stewards to resolve and validate merging tasks.

When resolving merging tasks, the data steward can manually update the golden record by choosing one of the values from the source records or by defining a different value.

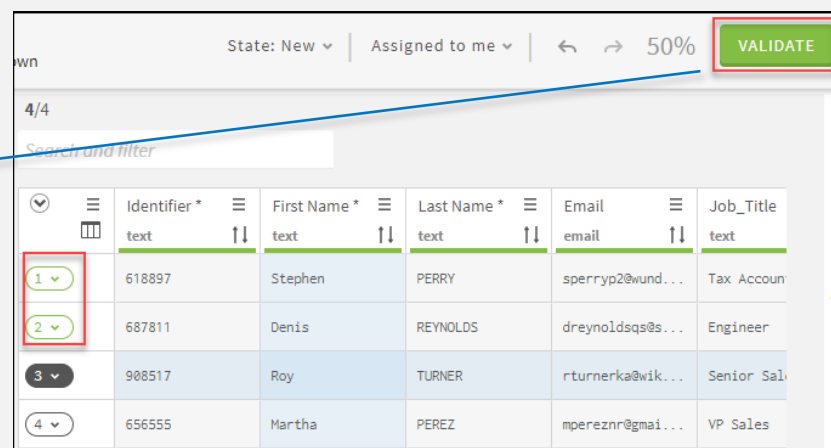
After resolving tasks, as a data steward, click the **lock** icon to mark the task ready for validation.



	Identifier *	First Name *	Last Name *	Email	Job Title	Company
	text	text	text	email	text	text
1	618897	Stephen	PERRY	sperry2@wund...	Tax Accountant	Aivee
Source 1	618897	Stephen	PERRY	sperry2@wund...	Tax Accountant	Aivee
Source 2	395657	Steph	PERRY	sperry2@wund...	Tax Accountant	

Finally, a data steward validates the tasks that are ready for validation. After validation, the task is resolved.

Click the **VALIDATE** button to change the tasks' state to **RESOLVED**.



State: New | Assigned to me | 50% **VALIDATE**

4/4

Search and filter

	Identifier *	First Name *	Last Name *	Email	Job Title
	text	text	text	email	text
1	618897	Stephen	PERRY	sperry2@wund...	Tax Accountant
2	687811	Denis	REYNOLDS	dreynoldsqs@...	Engineer
3	908517	Roy	TURNER	rturnerka@wik...	Senior Sal
4	656555	Martha	PEREZ	mpereznr@gmail...	VP Sales