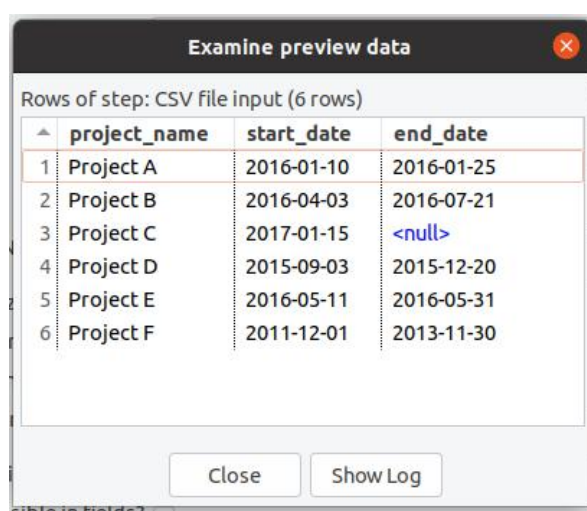


ETL TP : 2

1. Creation d'une nouvelle transformation.
2. Ajouter une step «CSV file input».
3. Double-cliquez sur l'icône d'entrée du fichier CSV et entrez "Projects" dans le champ "Step name".
4. Sous «Filename», tapez le chemin d'accès complet au fichier de votre projet Alternativement, cliquez sur «Browse» et cherchez le fichier sur votre disque.
5. cliquez sur "Get Fields" pour récupérer les champs du fichier.
6. Changer les type des date et ajouter la forme des date «yyy-MM-dd».

▲	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim
1	project_name	String							
2	start_date	Date	yyy-MM-dd						
3	end_date	Date	yyy-MM-dd						

7. Cliquez sur "Preview", et dans la petite fenêtre qui apparaît, cliquez sur OK pour voir l'ensemble de données défini.



Examine preview data

Rows of step: CSV file input (6 rows)

▲	project_name	start_date	end_date
1	Project A	2016-01-10	2016-01-25
2	Project B	2016-04-03	2016-07-21
3	Project C	2017-01-15	<null>
4	Project D	2015-09-03	2015-12-20
5	Project E	2016-05-11	2016-05-31
6	Project F	2011-12-01	2013-11-30

Close Show Log

8. Cliquez sur "Preview", et dans la petite fenêtre qui apparaît, cliquez sur OK pour voir l'ensemble de données défini.
9. Ajouter la step «Calculator».
10. Créez un hop de l'étape d'entrée du fichier CSV vers l'étape de la Calculator.
11. Un petit menu apparaîtra pour le type de houblon. Parmi les options, sélectionnez "Main output of step"
12. Double-cliquez sur la step "Calculator" et remplissez la première ligne de la grille avec les champs suivants :

Step name: Calculator

Fields:

	New field	Calculation	Field A	Field B	Field C	Value type	Length
1	diff_date	Date A - Date B (in days)	end_date	start_date		Integer	

Buttons: ? Help, OK, Cancel

Enfin, nous évaluerons la performance du projet :

13. Ajoutez une nouvelle step, "Number Range", et reliez la step "Calculator" à "Number Range" par un nouveau hop.
14. Avec "Number ranges", vous allez créer un nouveau champ, "performance", basé sur la valeur le champ entrant, "diff_dates". sur la valeur d'un champ entrant, diff_dates. Double-cliquez sur l'étape et remplissez la grille comme indiqué dans la capture d'écran suivante :

Step name: Number range

Input field: diff_date

Output field: performance

Default value(if no range matches): unknown

Ranges (min <= x < max):

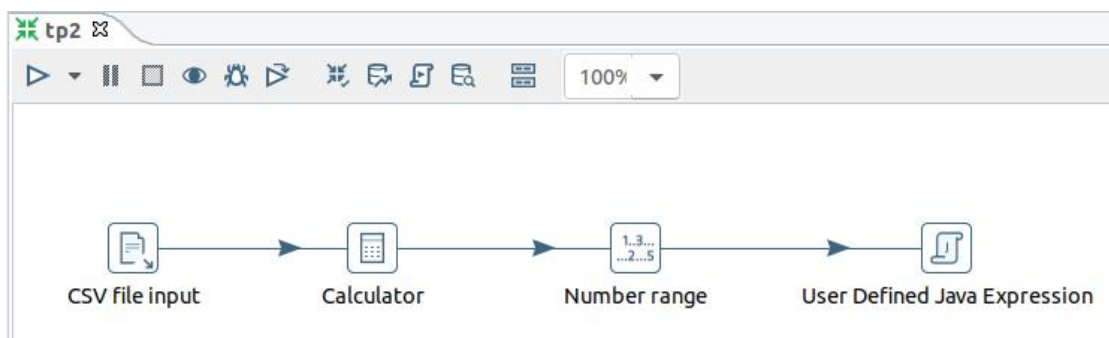
	Lower Bound	Upper Bound	Value
1		30.0	Excellent
2	30.0	80.0	Très bien
3	80.0	160.0	Bien
4	160.0		Mauvais

Buttons: ? Help, OK, Cancel

15. Avec la step "User Defined Java Expression", vous allez créer deux messages informatifs : "duration" et "message". Comme dans l'étape "Calculator", cette étape vous permet également de créer un nouveau champ par ligne. Double-cliquez sur l'étape et remplissez la grille comme indiqué dans la capture d'écran suivante :

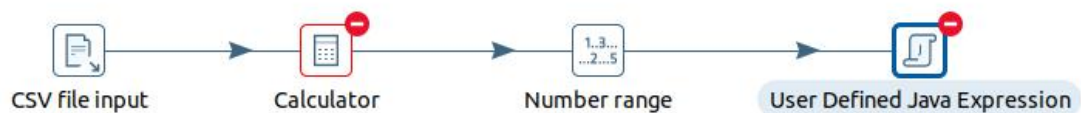


16. Votre transformation finale Transformation devrait ressembler à la capture d'écran suivante :



17. Prévisualisation et correction des erreurs au fur et à mesure qu'elles apparaissent

18. Sélectionnez l'étape User Defined Java Expression «UDJE» et lancez un aperçu.



19. En lisant le journal, on peut voir la source des l'erreurs :

- - Calculator.0 - ERROR (version 7.1.0.0-12, build 1 from 2017-05-16 17:18.02 by buildguy) : !Calculator.ErrorInStepRunning: - Calculator.0 - end_date String: couldn't convert string [??] to a date using format [yyy-MM-dd] on offset location 0
- - User Defined Java Expression.0 - org.codehaus.janino.CompileException: Line 1, Column 45: Unknown variable or type "diff_dates"
- - User Defined Java Expression.0 - Line 1, Column 45: Unknown variable or type "diff_dates"
- - User Defined Java Expression.0 -

20. Modifiez le fichier projects.txt et supprimez la ligne qui provoque l'erreur.

Aperçu final d'une Transformation :

The image shows a data transformation workflow with four steps: CSV file input, Calculator, Number range, and User Defined Java Expression. Below the workflow is a window titled "Examine preview data" showing the output of the User Defined Java Expression step. The window displays a table with 6 rows of data.

Workflow steps:

- CSV file input
- Calculator
- Number range
- User Defined Java Expression

Examine preview data (6 rows):

	project_name	start_date	end_date	diff_date	performance	duration	message
1	Project A	2016-01-10	2016-01-25	15	Excellent	15 jours	Les performances sont Excellent
2	Project B	2016-04-03	2016-07-21	109	Bien	109 jours	Les performances sont Bien
3	Project C	2017-01-15	2018-07-21	552	Mauvais	552 jours	Les performances sont Mauvais
4	Project D	2015-09-03	2015-12-20	108	Bien	108 jours	Les performances sont Bien
5	Project E	2016-05-11	2016-05-31	20	Excellent	20 jours	Les performances sont Excellent
6	Project F	2011-12-01	2013-11-30	730	Mauvais	730 jours	Les performances sont Mauvais

Close