

A dark blue vertical bar runs down the left side of the page. A blue arrow points to the right, overlapping the bar, with the letters 'BI' in white inside it.

BI

ATELIER

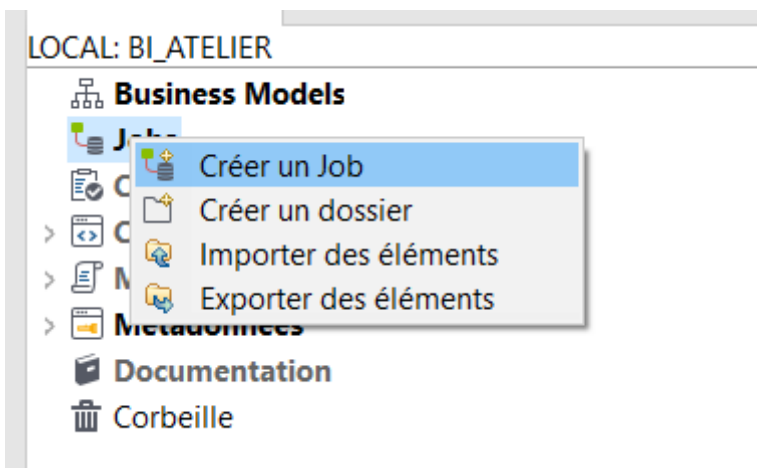
M2 MIND
2020 - 2021

Zakaria Belaribi

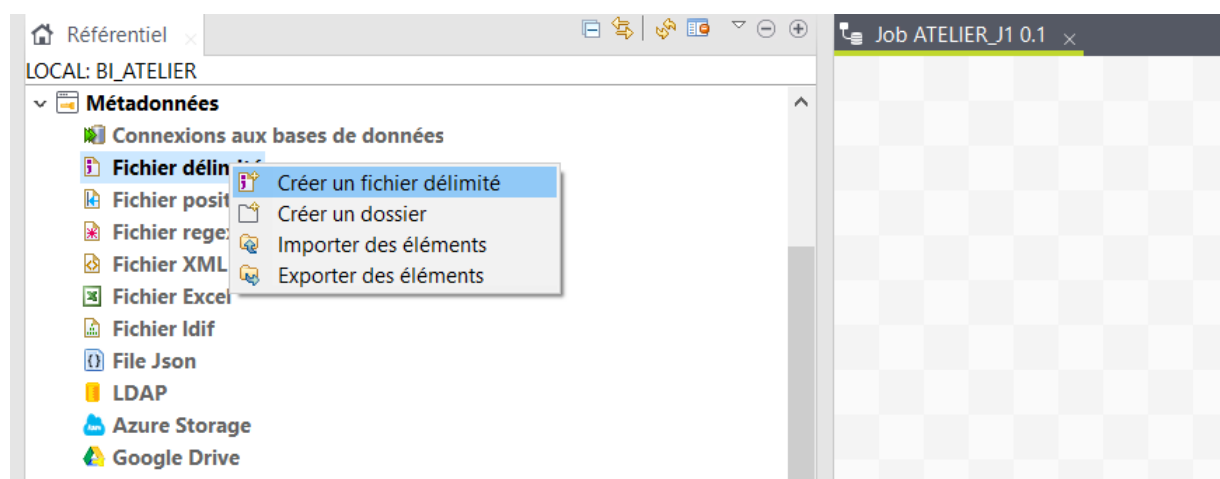
Chakib Bachounda

Several thin, curved lines in shades of blue and grey originate from the bottom left and sweep upwards and to the right, creating a dynamic, abstract design.

On commence d'abord par créer un job principal pour notre atelier :



On crée nos fichiers délimités pour nos inputs ; notre fichier CSV et le fichier texte :



Pour le premier fichier, on choisi le mode CSV et non délimité, et on choisit de définir l'entête comme nom de colonnes (comme le montre l'aperçu) :

Nouveau fichier délimité

Fichier - Etape 3 de 4

Add a Metadata File on repository
Define the setting of the parse job

Paramètres de fichier

Encodage

Séparateur de champs Caractère correspondant

Séparateur de lignes Caractère correspondant

Paramètres du caractère d'échappement

☒ CSV ☐ Délimité

Caractère d'échappement

Entourage du texte

☐ Découper la ligne avant le champ

Lignes à ignorer

Si des lignes doivent être ignorées, spécifiez les paramètres

En-tête ☒

Pied de page ☐

☐ Ignorer les lignes vides

Limite de lignes

Si le nombre de lignes doit être limité, spécifiez ce nombre

Limite ☐

Aperçu Sortie

☒ Définir les lignes d'en-tête comme nom de colonnes

NUMERO	NOM	PRENOM	AGE
1	NINO	mohamed	54
2	NANA	Douaa	36
3	NINA	Imene	49
4	NANI	Ali	25

< Back

Next >

Finish

Cancel

Tout cela après avoir donné un nom à ce fichier lors de l'affichage de ce dernier dans la fenêtre des Jobs. On laisse le délimiteur par défaut ' ; '. Il ne reste plus qu'à préciser la clé primaire (vu que c'est fichier .csv) :

Fichier - Etape 4 de 4

Ajouter un schéma au référentiel

Définissez le schéma

Nom	metadata
Commentaire	

Schéma

Cliquez pour mettre à jour l'aperçu du schéma Détecter

Description du schéma

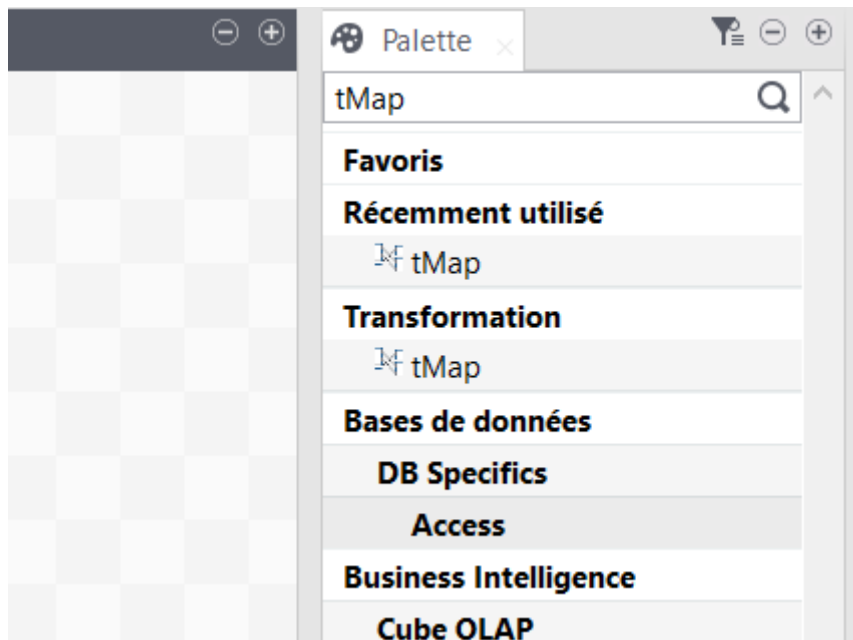
Colonne	Clé	Type	<input checked="" type="checkbox"/> N..	Modèle de date (...)	Length	Precision	Défaut	Co
NUMERO	<input checked="" type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		2	0		
NOM	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		6	0		
PRENOM	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		8	0		
AGE	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		2	0		

La prochaine étape sera de transformer ce fichier .csv en un fichier .xml avec l'outil dédié à ce process ; **tFileOutputXML**. Pour avoir ce dernier, il suffit de le chercher dans la fenêtre des palettes tout à droite (il figure en dernier de la liste). Il ne restera plus qu'à les lier entre eux avec un lien Main de type Row, et voilà, on aura en sortie de l'outil le fichier XML souhaité :

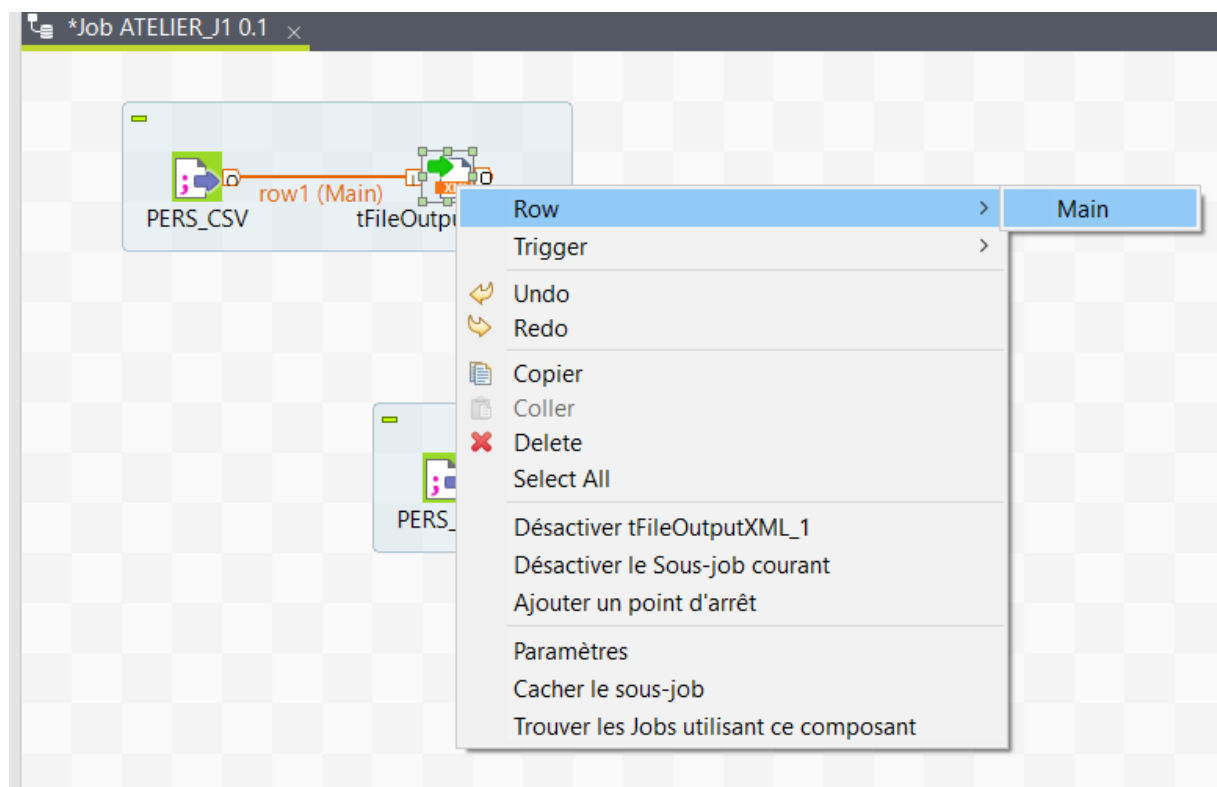


Idem pour le fichier texte, de la même façon on crée ce dernier.

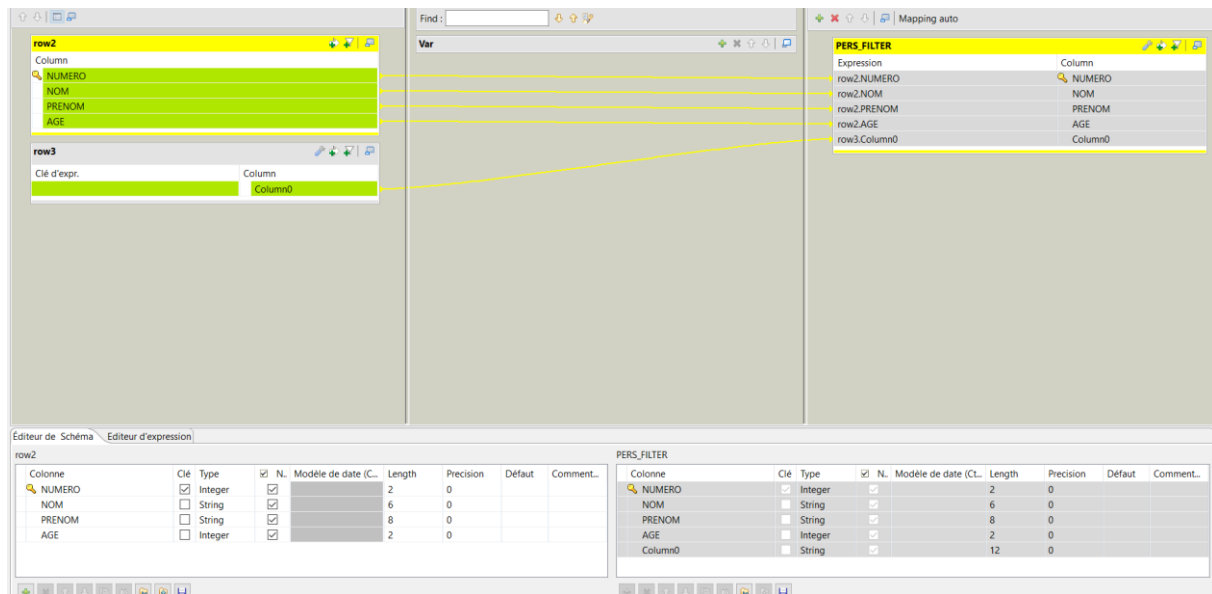
Pour joindre ces deux fichiers, on utilise l'outil **tMap**, et pour l'avoir lui aussi à son tour, on se redirige vers la palette :



On choisit le fichier XML comme flux de données principale, c'est pour ça qu'on choisit un lien Main de type Row avec le composant **tMap** (et l'autre fichier aura un lien secondaire – lookup) :

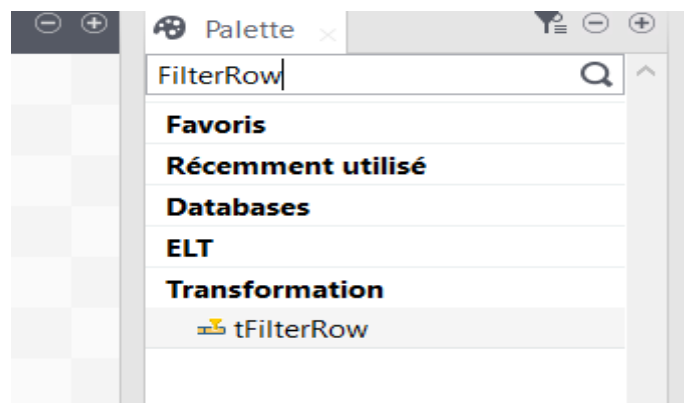


Voici le travail accompli à l'intérieur du composant **tMap** (jointure sans attributs communs) :

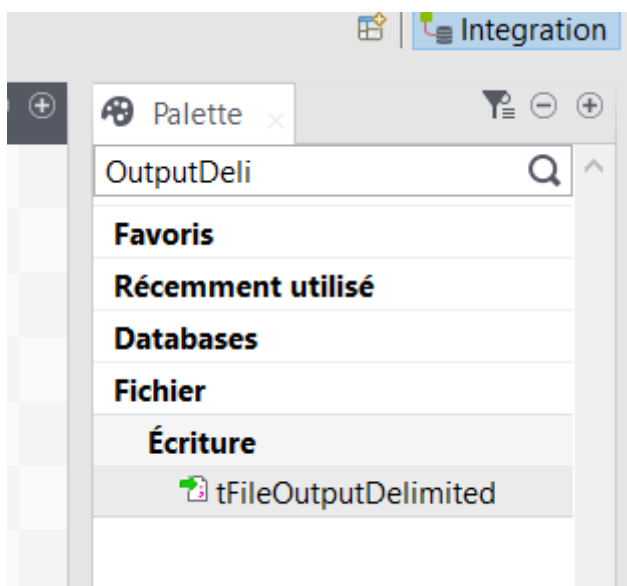


Après avoir joint les 2 fichiers, il ne reste plus qu'à filtrer le résultat de cette jointure ; à savoir les personnes ayant plus de 30 ans ou non.

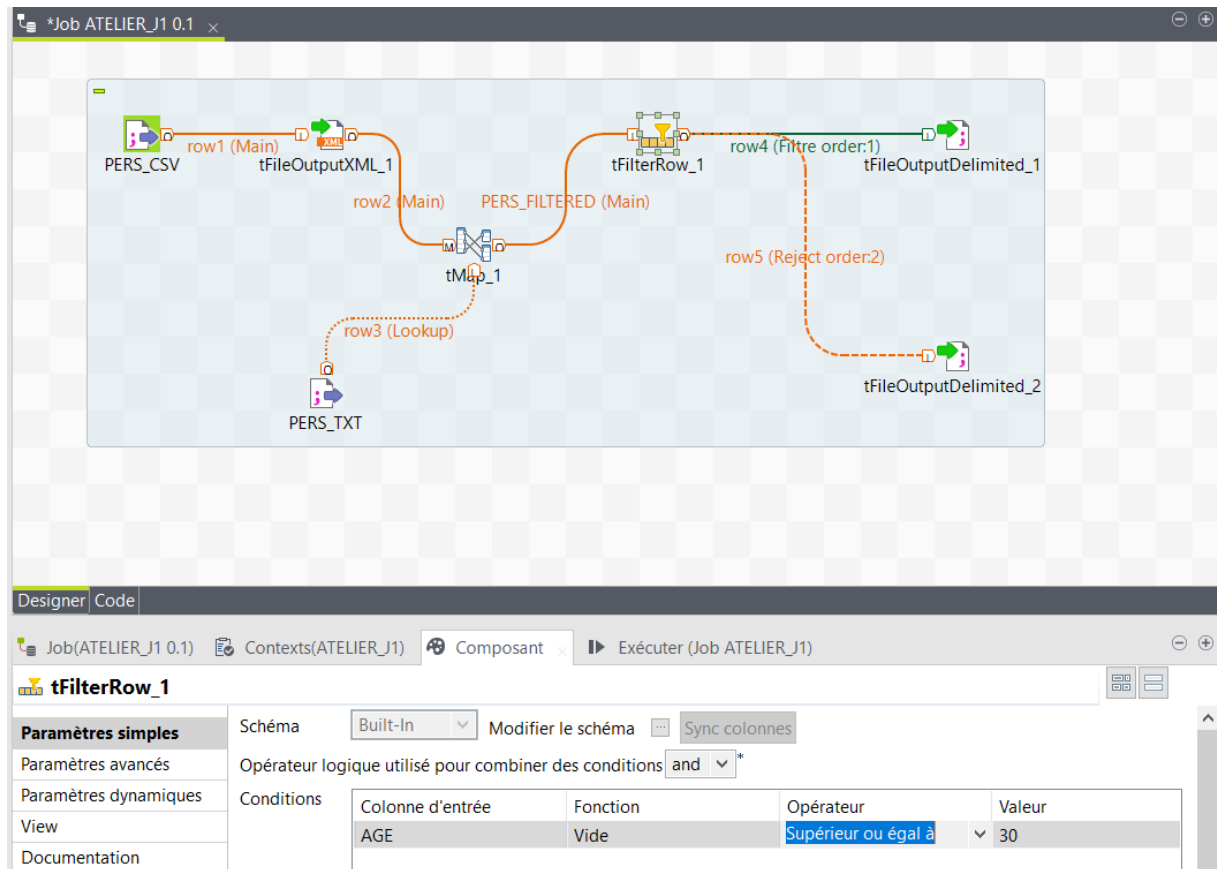
Afin de filtrer, on utilise le composant **tFilterRow** :



Après filtration (on verra dans ce qui suit les conditions de filtration), on crée nos fichiers de sorties via le composant **tFileOutputDelimited** :



Voici à quoi ressemble notre schéma final après avoir rajouté les fichiers de sorties (en mentionnant leurs paths). Voici la condition que porte le composant tFilterRow (tout en bas) :



En sortie du composant, on précise que les éléments qui répondent positif à la condition sont chargé (LOADED) dans le tFileOutputDelimited_1 (lien Filtre), et les autres dans le 2^e fichier tFileOutputDeilimited_2 (lien reject).

Voici à quoi ressemble le schéma après execution (code=0 en java signifie exécution sans erreur), et on a bien le nombre de ligne chargé dans chaque composant donc on peut affirmé que le job a été exécuter avec succès.

Job ATELIER_J1 0.1

Designer Code

Job(ATELIER_J1 0.1) Contexts(ATELIER_J1) Composant Exécuter (Job ATELIER_J1)

Job ATELIER_J1

Exécution simple

Exécution Debug

Paramètres avancés

Target Exec

Memory Run

Exécution

Exécuter Arrêter Effacer

Démarrage du job ATELIER_J1 à 15:27 27/01/2021.
 [statistics] connecting to socket on port 3884
 [statistics] connected
 [statistics] disconnected
 Job ATELIER_J1 terminé à 15:27 27/01/2021. [Code sortie=0]

Default

Nom

Voici un aperçu des fichiers en sorties :

pers_rejected.txt - Bloc-notes

Fichier Edition Format Affichage Aide

```
4;NANI;Ali;25;;AGE.compareTo(30) >= 0 failed
6;DODI;Omar;29;;AGE.compareTo(30) >= 0 failed
7;DIDO;Ahmed;18;;AGE.compareTo(30) >= 0 failed
9;DODA;Ikram;19;;AGE.compareTo(30) >= 0 failed
10;MINO;Nesrine;20;;AGE.compareTo(30) >= 0 failed
14;CHICHO;Adel;21;;AGE.compareTo(30) >= 0 failed
15;CHOCHI;Fatma;23;;AGE.compareTo(30) >= 0 failed
```

pers_filtered.txt - Bloc-notes

Fichier Edition Format Affichage Aide

```
1;NINO;mohamed;54;
2;NANA;Douaa;36;
3;NINA;Imene;49;
5;NONA;Yassmine;41;
8;DADO;Djamel;52;
11;MONI;Khadidja;32;
12;MINA;Iteb;55;
13;MANA;Soumia;40;
```