

TP1 — Installation & Hello PDI

Fait par :**Khadija Nachid Idrissi**

:**Rajae Fdili**

:**Aya Hamim**

7 janvier 2026

1 Introduction

Ce TP a pour objectif de découvrir et de se familiariser avec l'outil **Pentaho Data Integration (PDI)**. Nous allons créer une transformation simple pour lire un fichier CSV, sélectionner certaines colonnes et enregistrer le résultat dans un nouveau fichier.

2 Préparation des données

Le fichier `clients.csv` contient les données suivantes :

```
id,nom,prenom,email,age,ville,pays
1,Dupont,Jean,jean.dupont@example.com,34,Paris,France
2,Martin,Claire,claire.martin@example.com,28,Lyon,France
```

L'objectif est de conserver uniquement les colonnes : `id`, `nom`, `prenom`, `email`.

3 Création de la transformation

1. Créer une nouvelle transformation : **Fichier** → **Nouvelle** → **Transformation** et l'enregistrer sous `TP1_hello.ktr`.
2. Ajouter les composants et les configurer :
 - **Text file input** : lire le fichier `clients.csv`, cocher **Header row present**, séparateur `,`.
 - **Altération structure flux (Select values)** : garder uniquement les colonnes `id`, `nom`, `prenom`, `email`.
 - **Text file output** : écrire le fichier `clients_hello.csv`.
3. Relier les composants avec des **hops** (flèche verte) pour former le flux complet.
4. Exécuter la transformation.

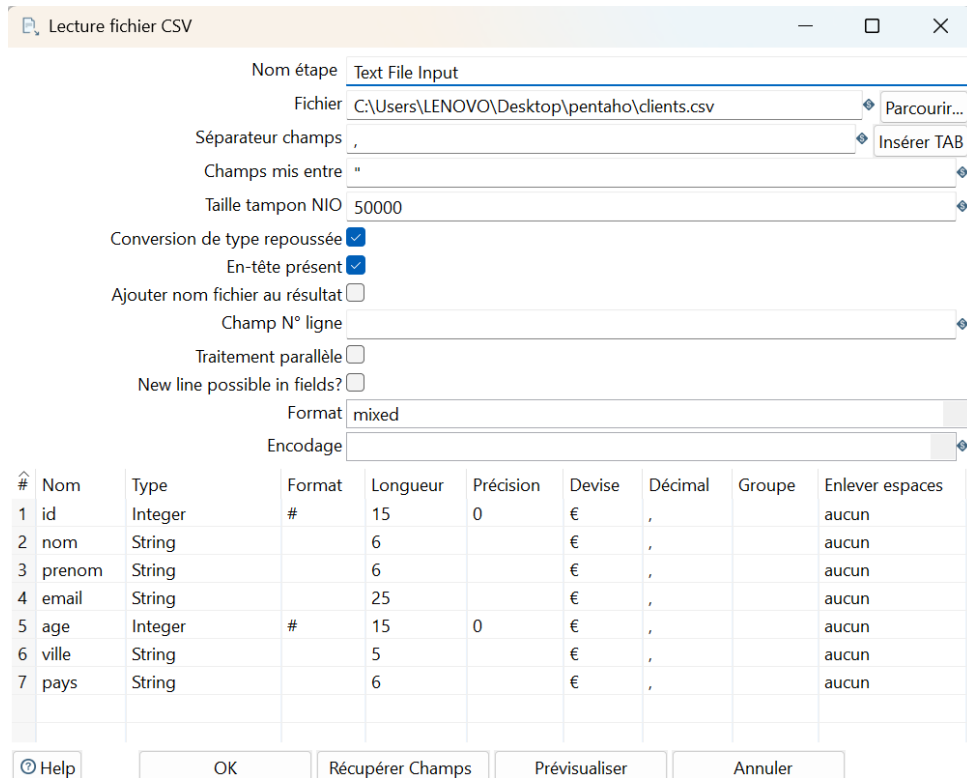


FIGURE 1 – Configuration du composant Text File Input

4 Paramétrage clé et pièges

- Dans **Text file input** : cocher **Header row present**, choisir l'encodage UTF-8, séparateur **,**, guillemets doubles si nécessaire.
- Cliquer sur **Get fields** pour détecter automatiquement les colonnes et leurs types.
- Vérifier que la flèche verte relie bien tous les composants ; sinon, aucune donnée ne sera transmise.

5 Résultats et vérification

Après exécution, le fichier `clients_hello.csv` contient uniquement les colonnes sélectionnées :

```
id,nom,prenom,email
1,Dupont,Jean,jean.dupont@example.com
2,Martin,Claire,claire.martin@example.com
```

Le flux a été exécuté correctement, et les données ont été filtrées comme attendu.

6 Conclusion

Ce TP permet de prendre en main Pentaho Data Integration et de comprendre le principe de création de transformations simples. Nous avons appris à : lire un fichier CSV, sélectionner des colonnes et générer un fichier de sortie. Cette première transformation constitue la base pour des traitements ETL plus complexes.

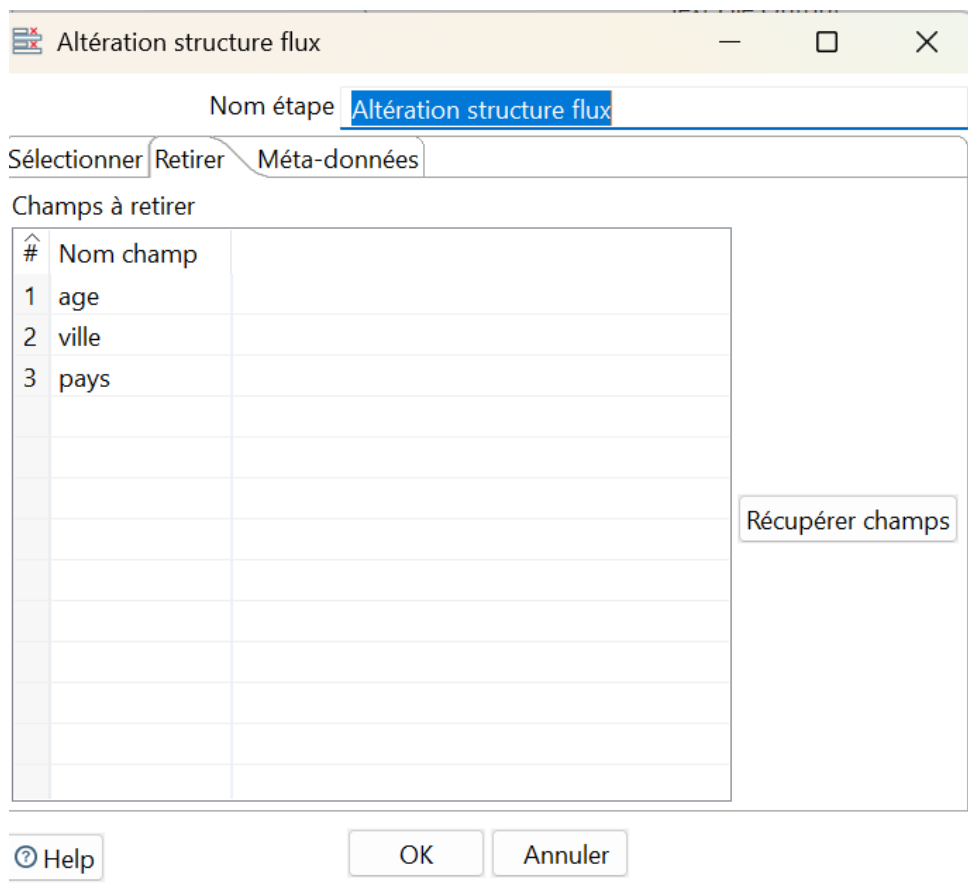


FIGURE 2 – Configuration du composant Altération structure flux pour sélectionner les colonnes

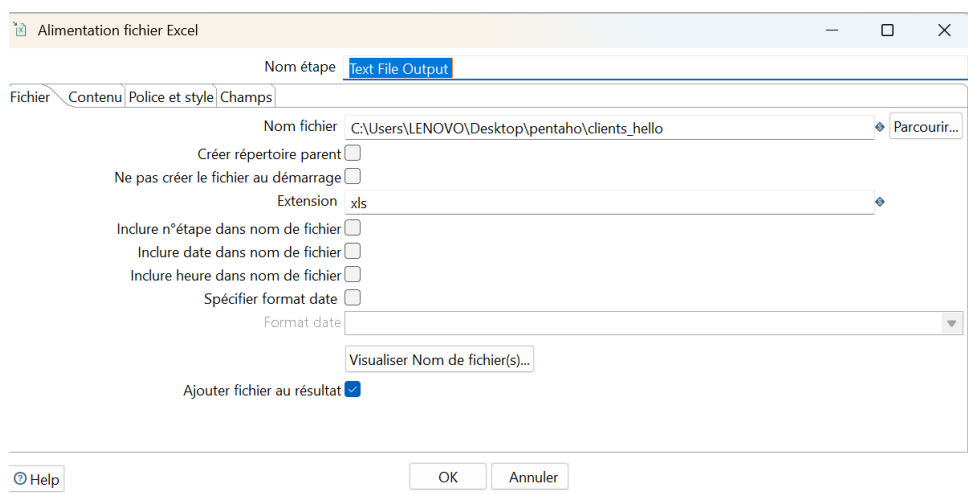


FIGURE 3 – Configuration du composant Text File Output

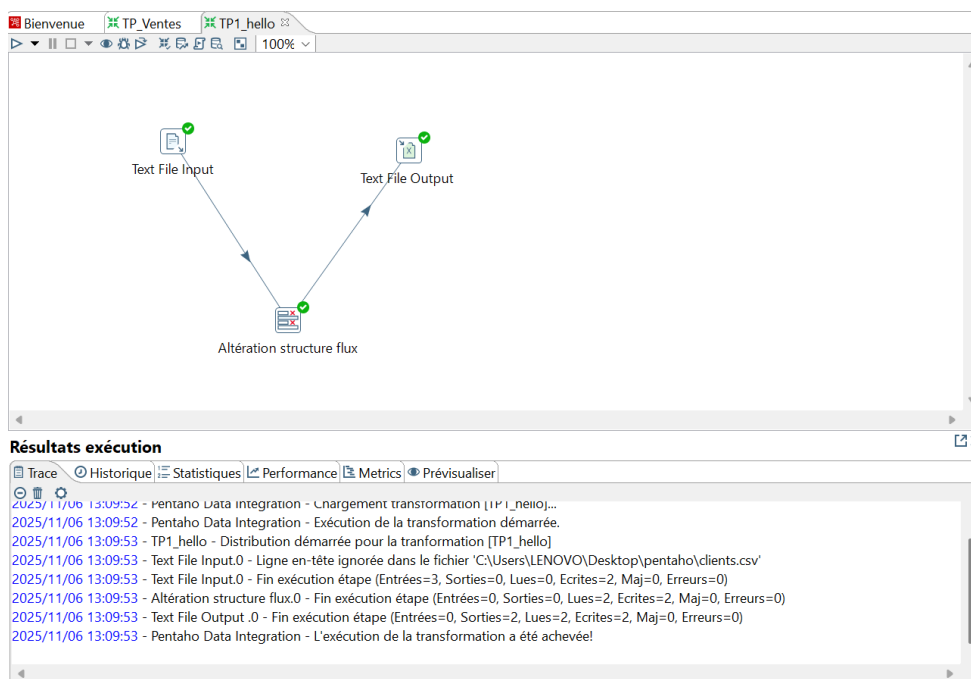


FIGURE 4 – Flux complet de la transformation dans Pentaho PDI