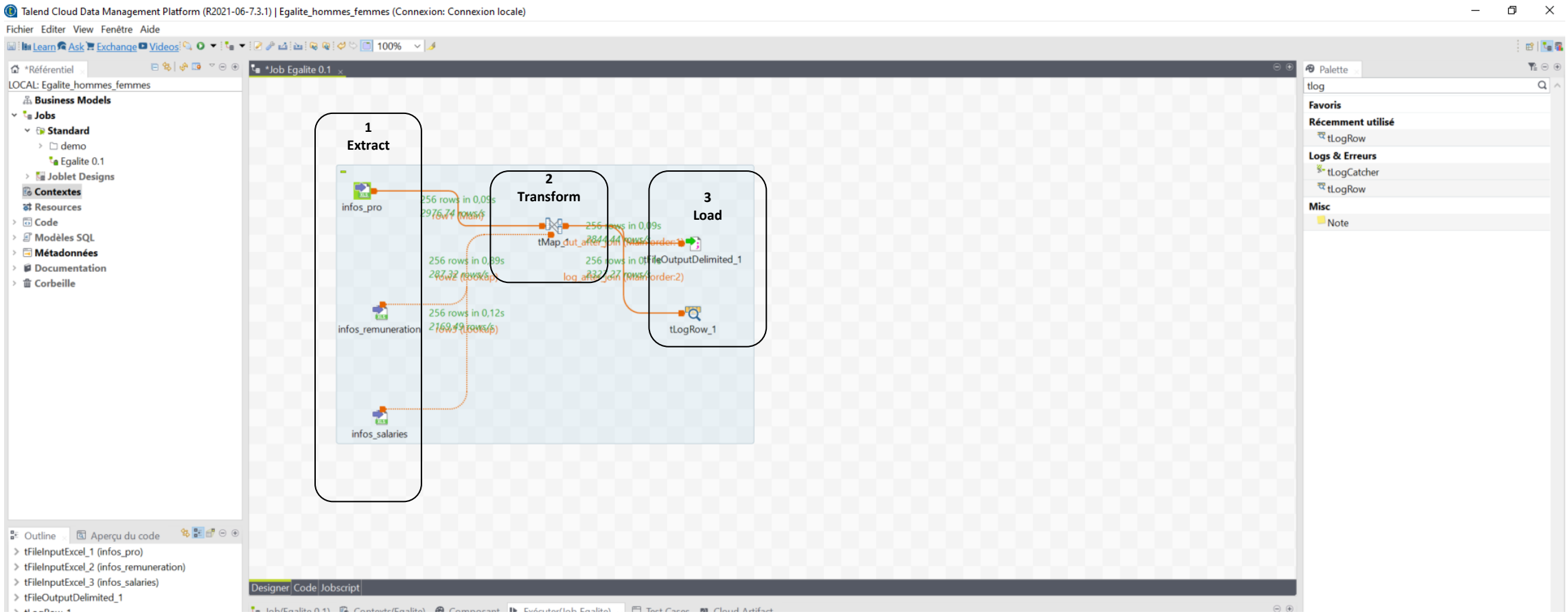
Utilisation de l'ETL TALEND® pour la création d'un fichier .CSV avec données anonymisées



1 Extract

Nos trois fichiers source présentent des données à caractère personnel sur les salariés d'une entreprise.

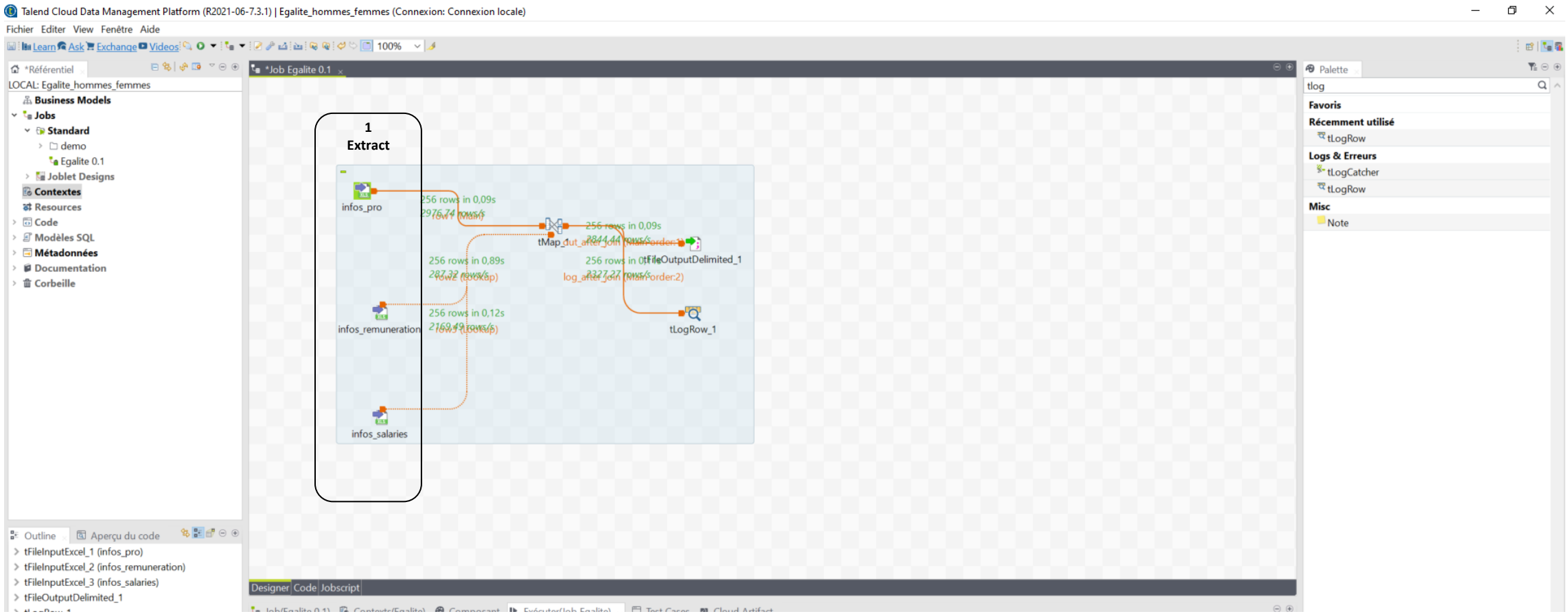
Objectif : créer un fichier CSV avec les données issues des trois fichiers source, dans le respect du RGPD. Ce fichier sera utilisé pour la réalisation d'un diagnostic égalité professionnelle hommes-femmes (embauche, qualification, promotion, conditions de travail , rémunération…)

=> suppression de certaines variables non utiles (nom, prénom, état civil, nombre d'enfants…)

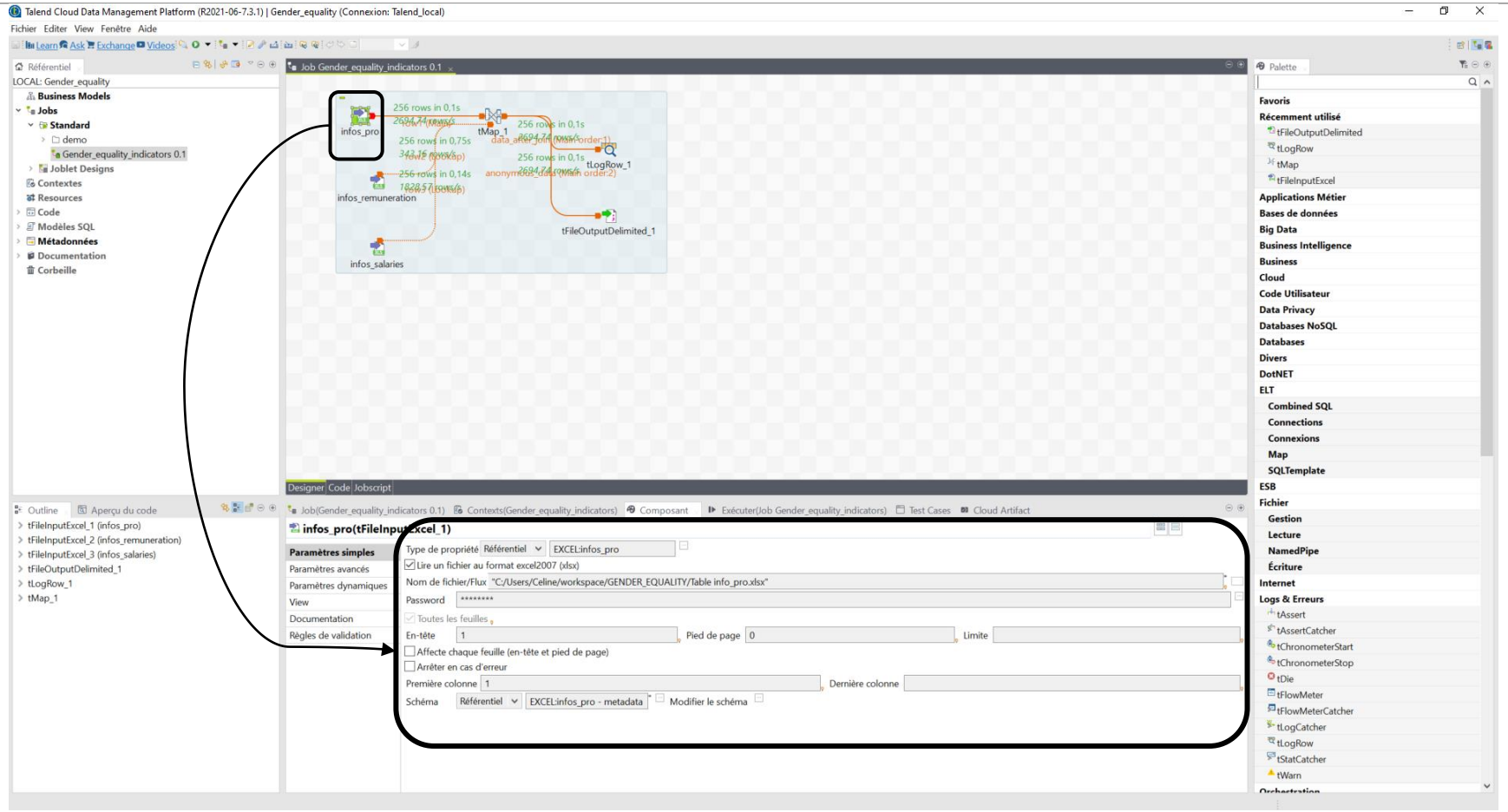
=> discrétisation de certaines variables numériques (âge, ancienneté…)

=> création de nouvelles variables (salaire total…)

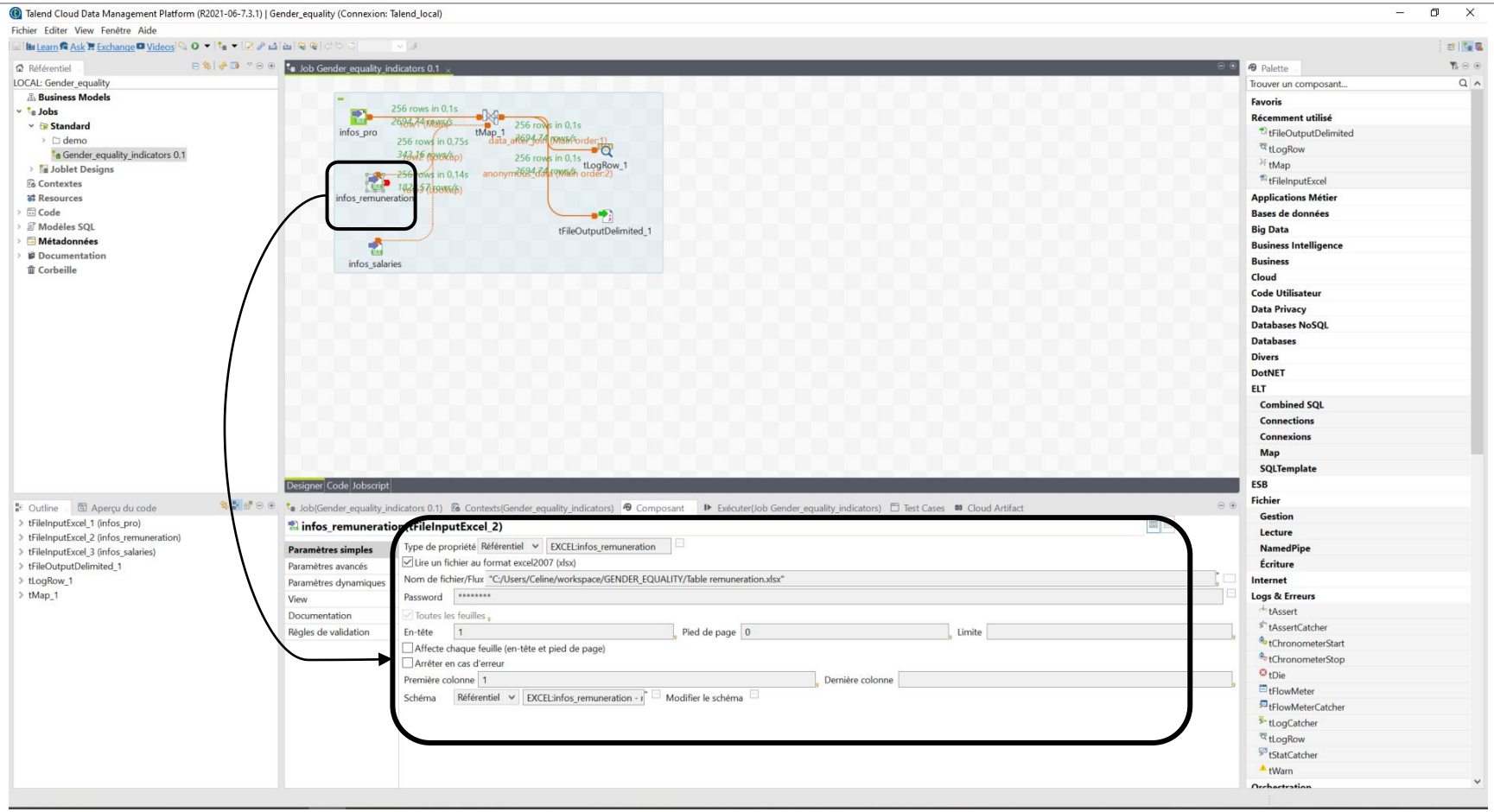
Utilisation de l'ETL TALEND® pour la création d'un fichier .CSV avec données anonymisées



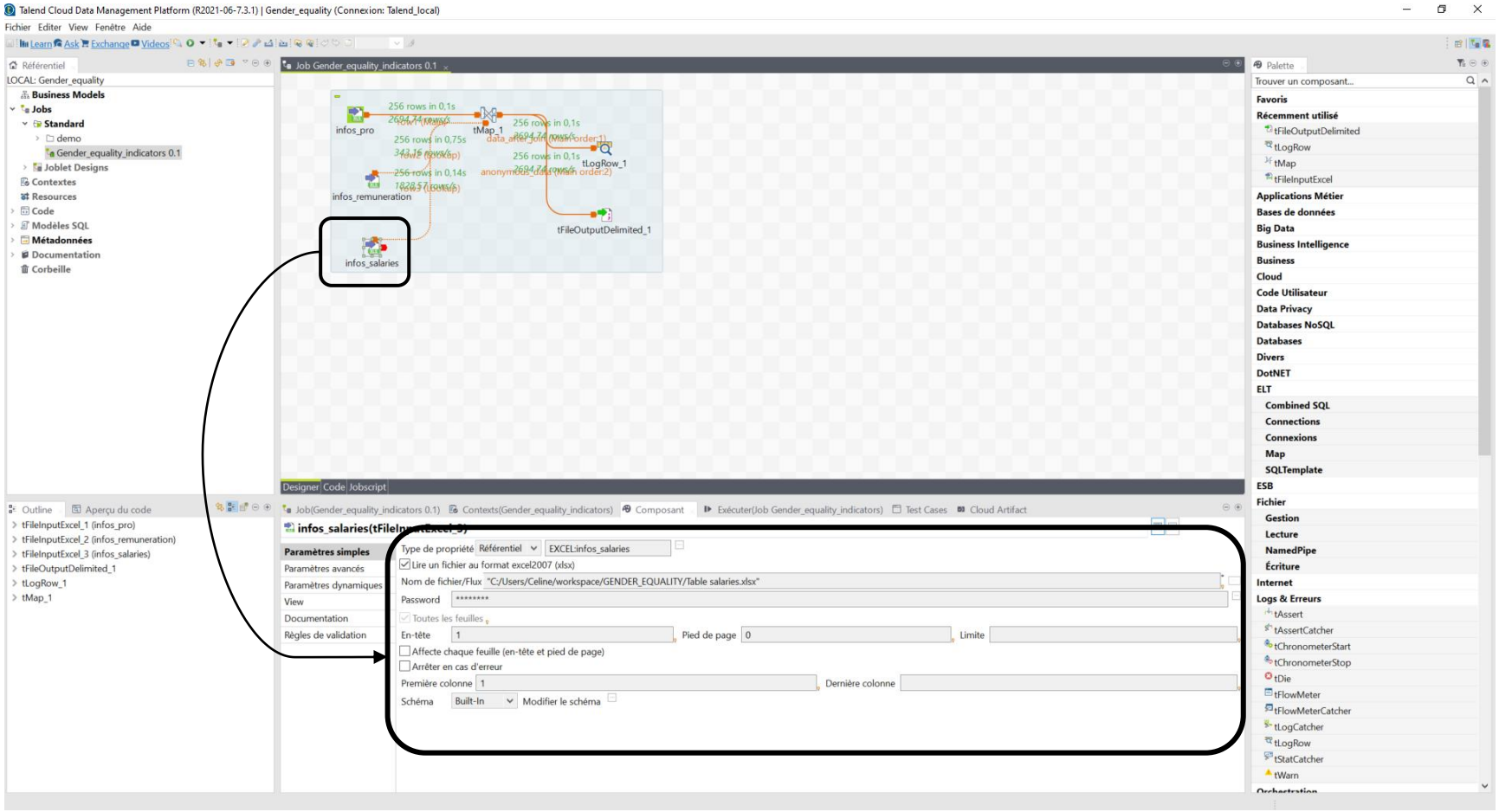
1.1 Extract : table_info_pro.xlsx



1.2 Extract : table_remuneration.xlsx

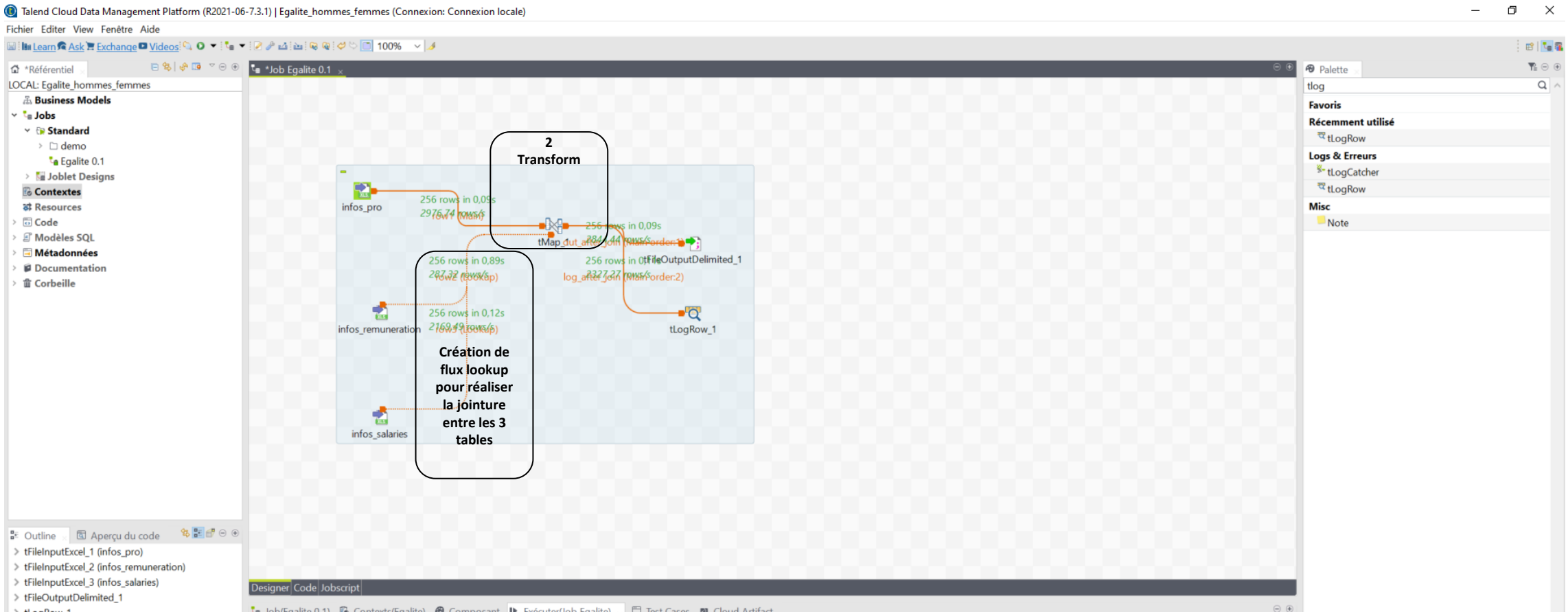


1.3 Extract : table_salaries.xlsx



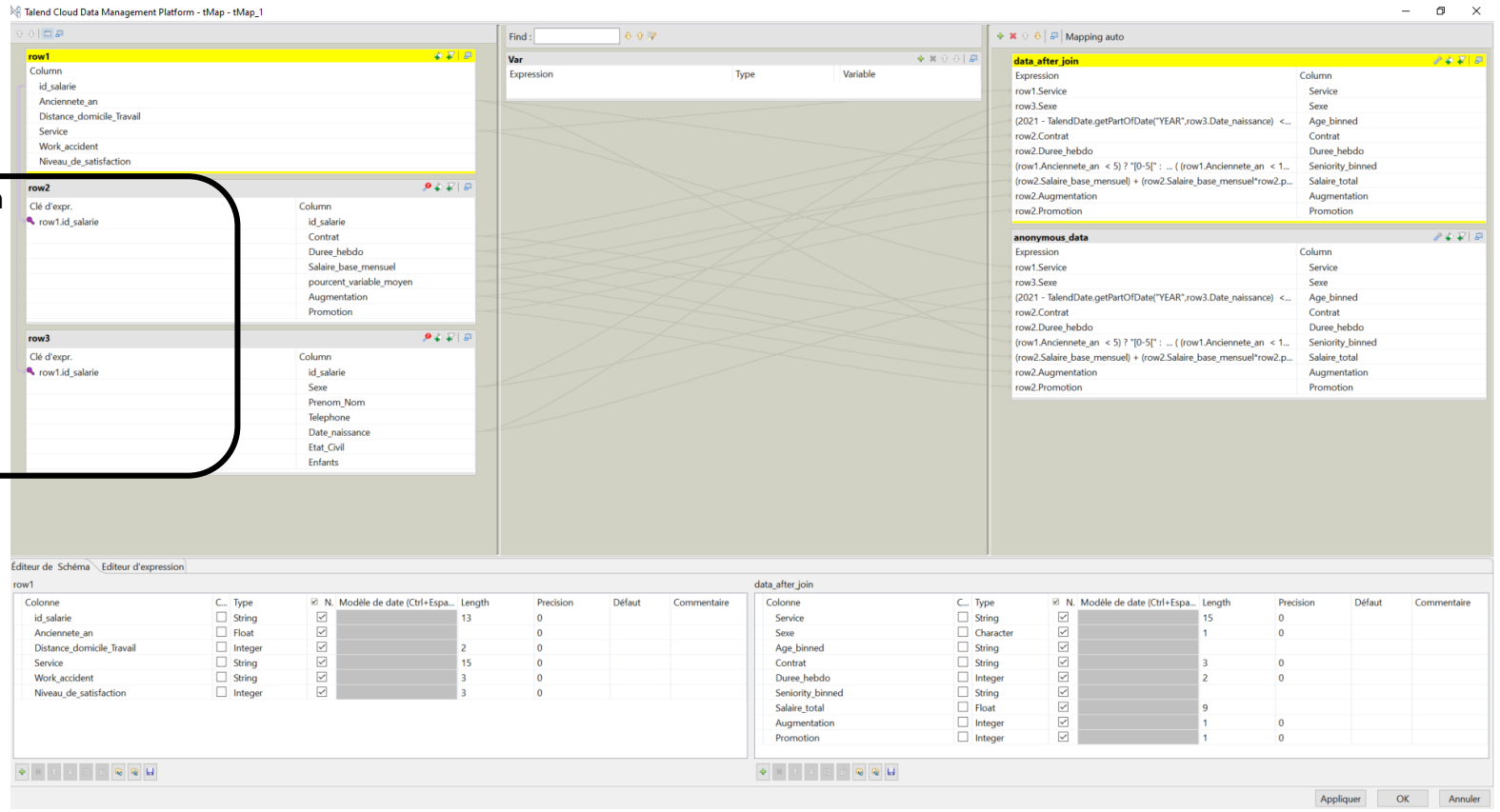
2 Transform

tMap


Talend Cloud Data Management Platform (R2021-06-7.3.1) | Egalite_hommes_femmes (Connexion: Connexion locale)
Fichier Editer View Fenêtre Aide
*Référentiel
LOCAL: Egalite_hommes_femmes
Business Models
Jobs
Standard
demo
Egalite 0.1
Joblet Designs
Contextes
Resources
Code
Modèles SQL
Métadonnées
Documentation
Corbeille
*Job Egalite 0.1
2
Transform
infos_pro
256 rows in 0,09s
tMap_1
256 rows in 0,09s
tFileOutputDelimited_1
256 rows in 0,89s
256 rows in 0,12s
infos_remuneration
tLogRow_1
infos_salaries
Création de flux lookup pour réaliser la jointure entre les 3 tables
Palette
tlog
Favoris
Récemment utilisé
tLogRow
Logs & Erreurs
tLogCatcher
tLogRow
Misc
Note
Outline
Aperçu du code
tFileInputExcel_1 (infos_pro)
tFileInputExcel_2 (infos_remuneration)
tFileInputExcel_3 (infos_salaries)
tFileOutputDelimited_1
Designer
Code
Jobscript

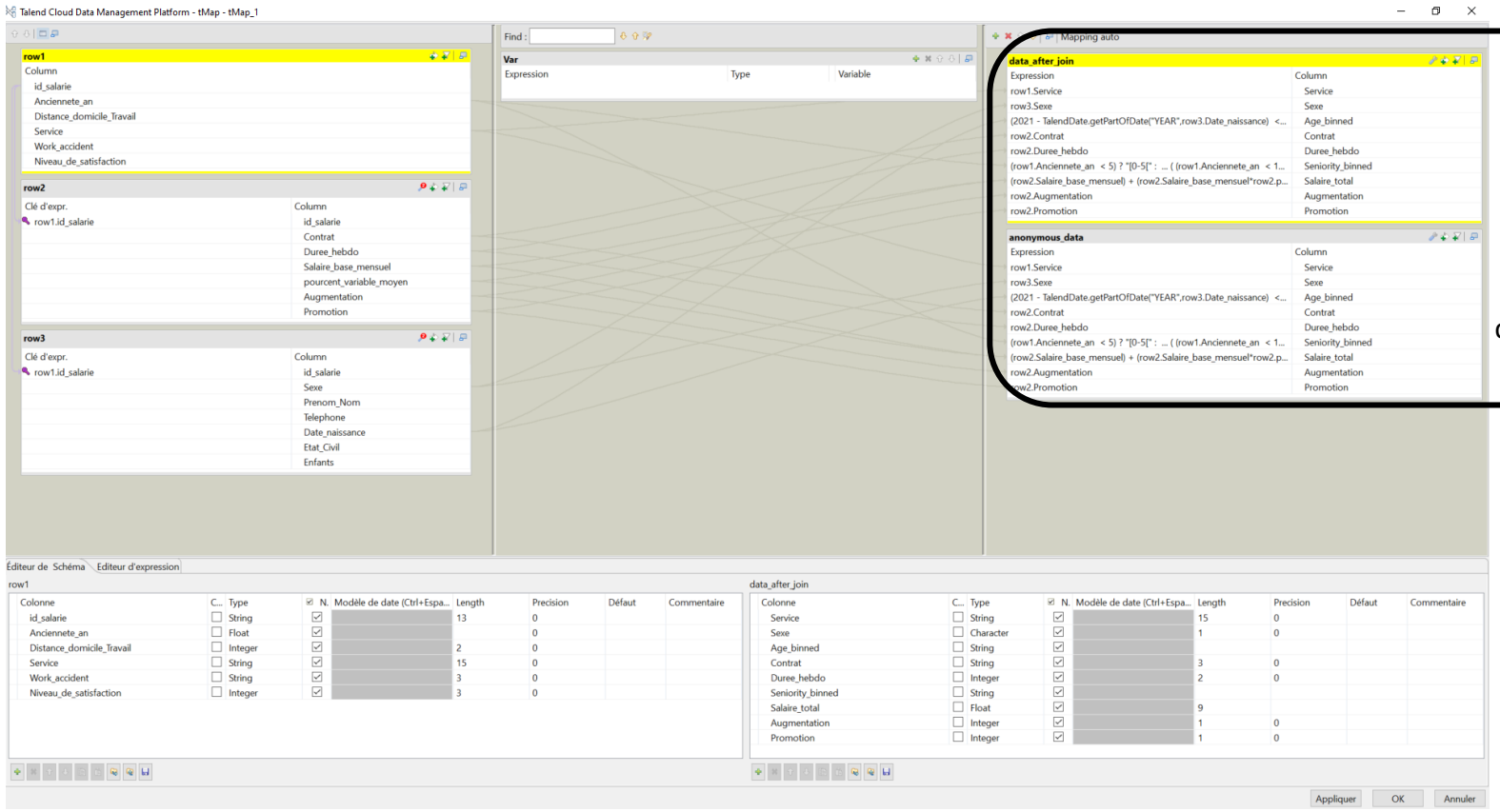
TMap : Jointure interne entre les 3 tables

Inner Join on *id_salarie*



talend

TMap : constructeur d'expression

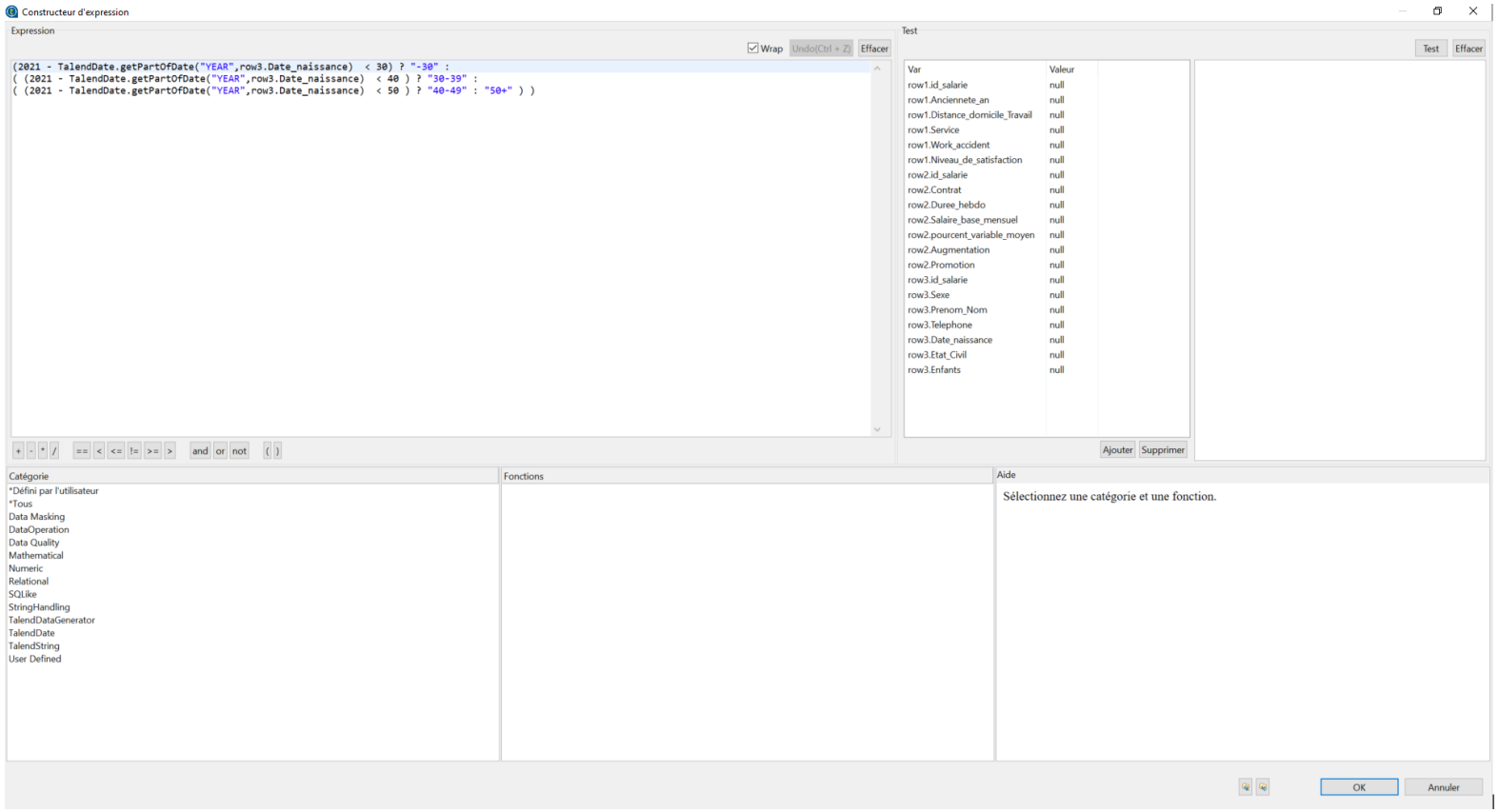


- Seules les données utiles aux analyses sont conservées

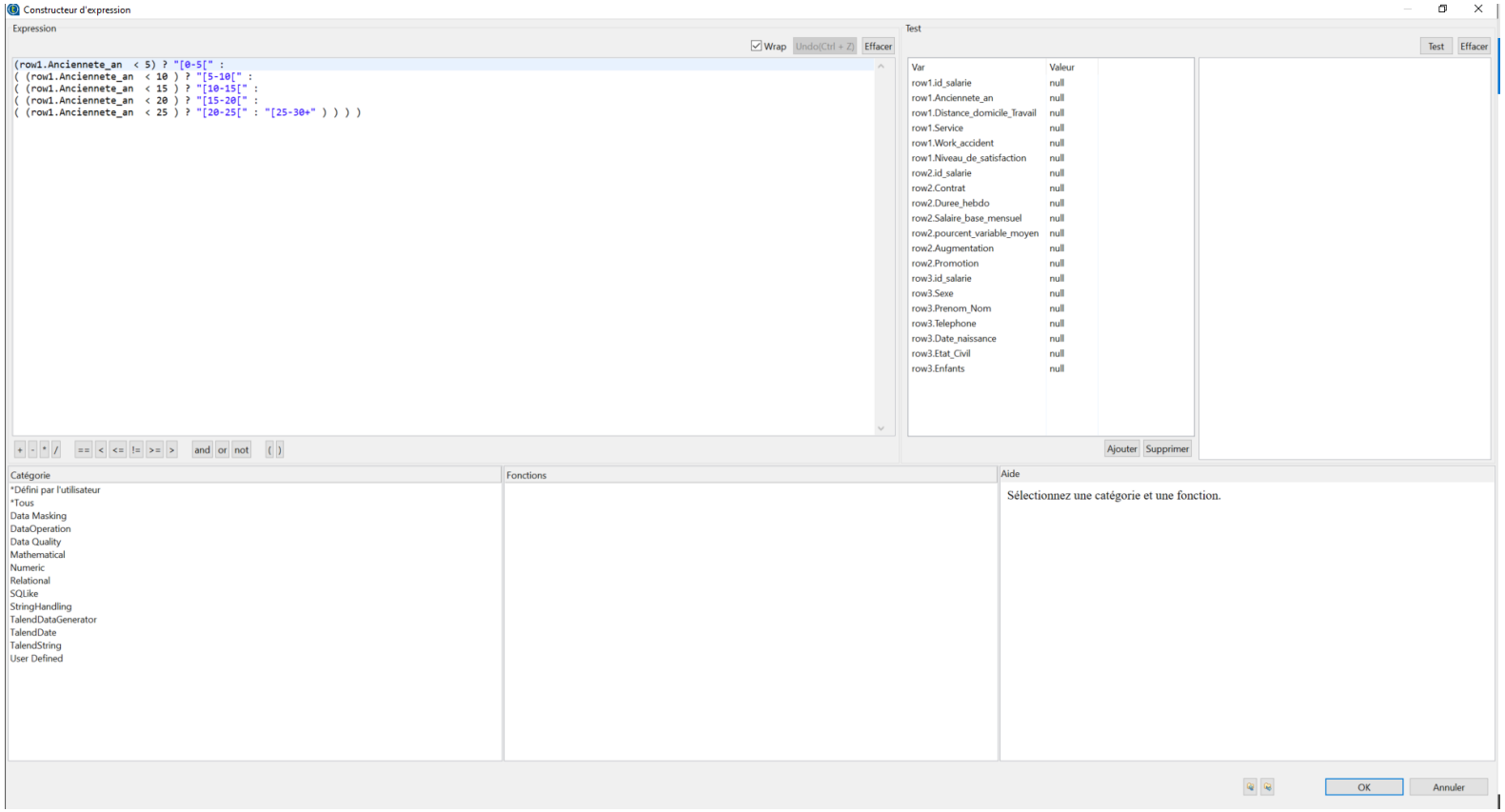
- Création de 3 nouvelles variables au moyen du constructeur d'expression

talend

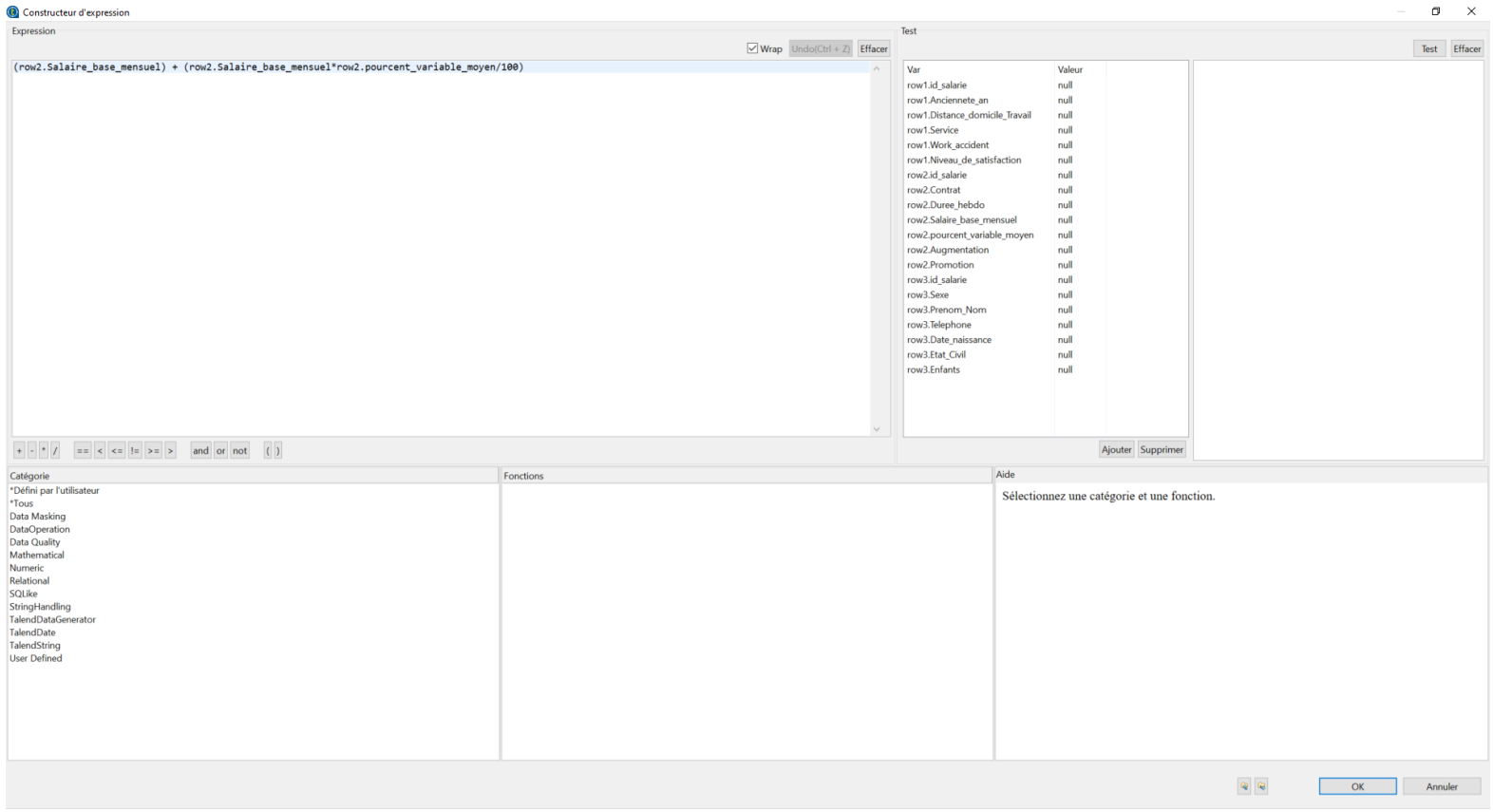
tMap constructeur d'expression : calcul de l'âge et discrétisation



tMap constructeur d'expression : discrétisation de la variable ancienneté

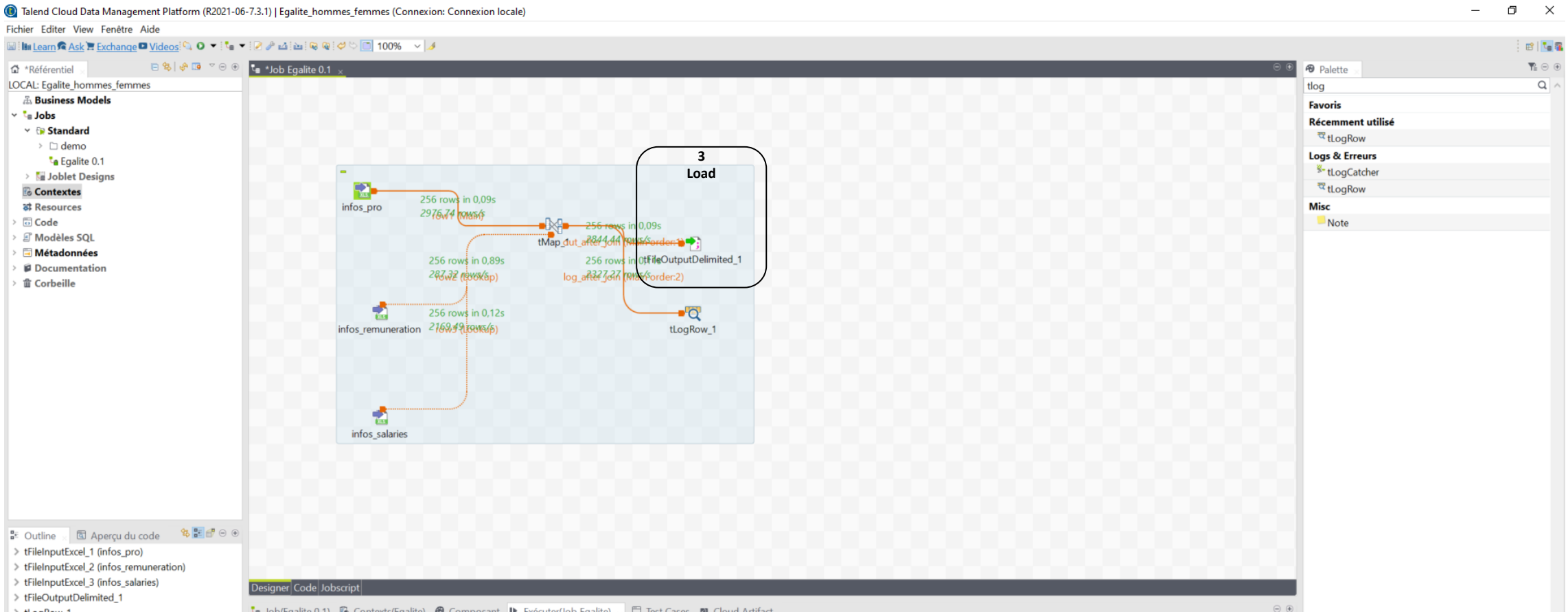


tMap constructeur d'expression : calcul du salaire total



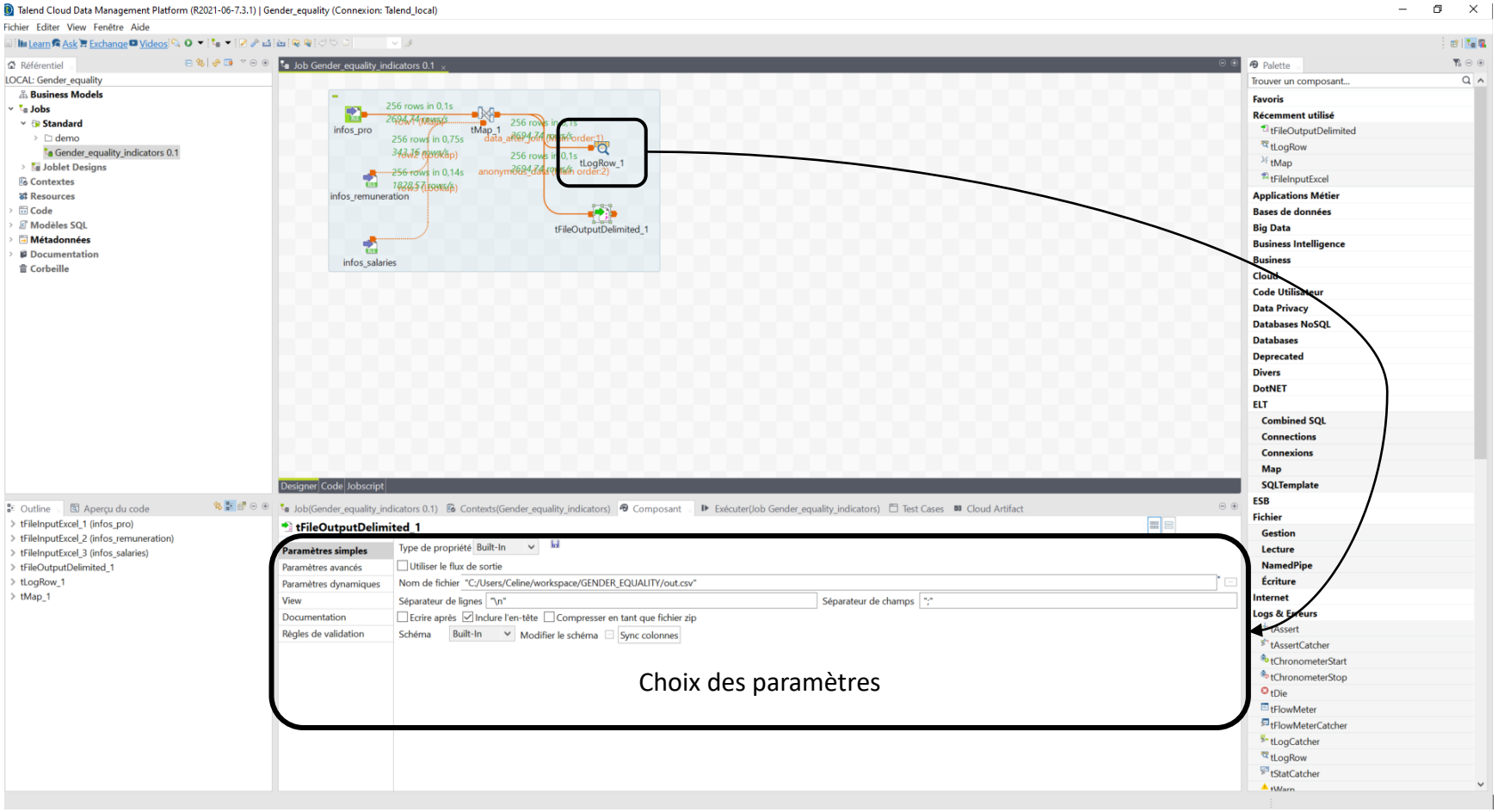
3 Load

3 Load : tFileOutputDelimited

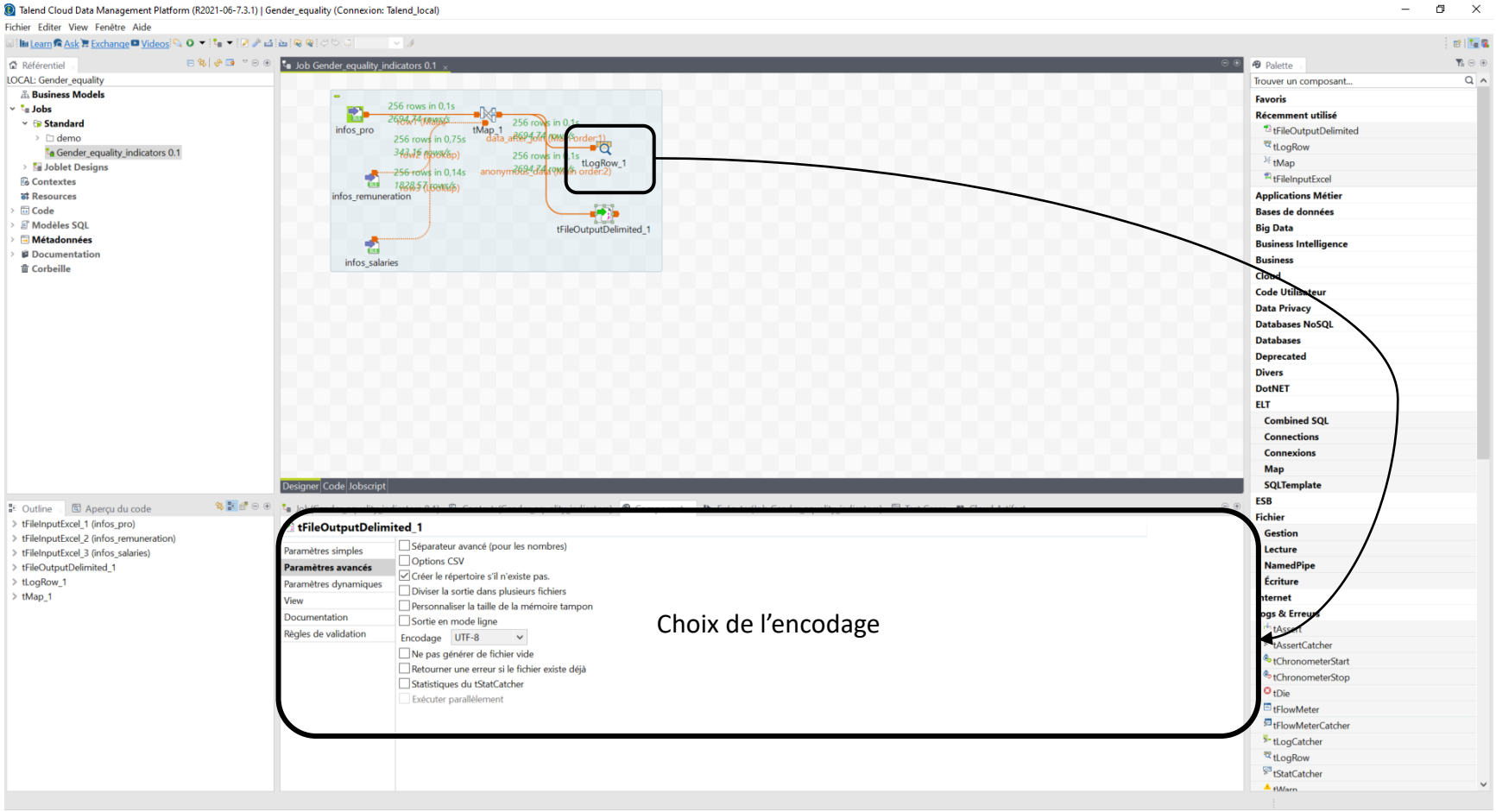


talend

tFileOutputDelimited : création d'un fichier .CSV



tFileOutputDelimited : encodage UTF-8



3 Load :tLog, visualisation de la table de sortie dans l'interface Talend ®