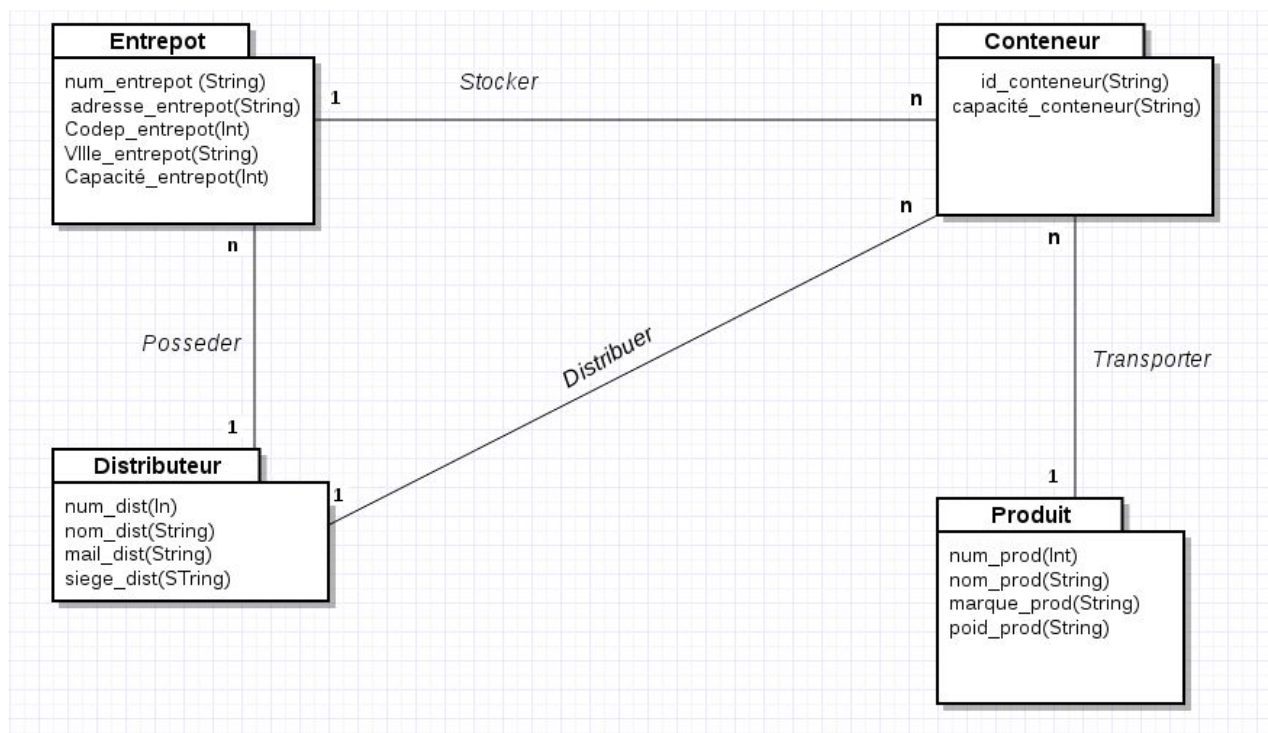


Présentation du Scénario

Etude de cas pour une société dans le domaine: Import/Export

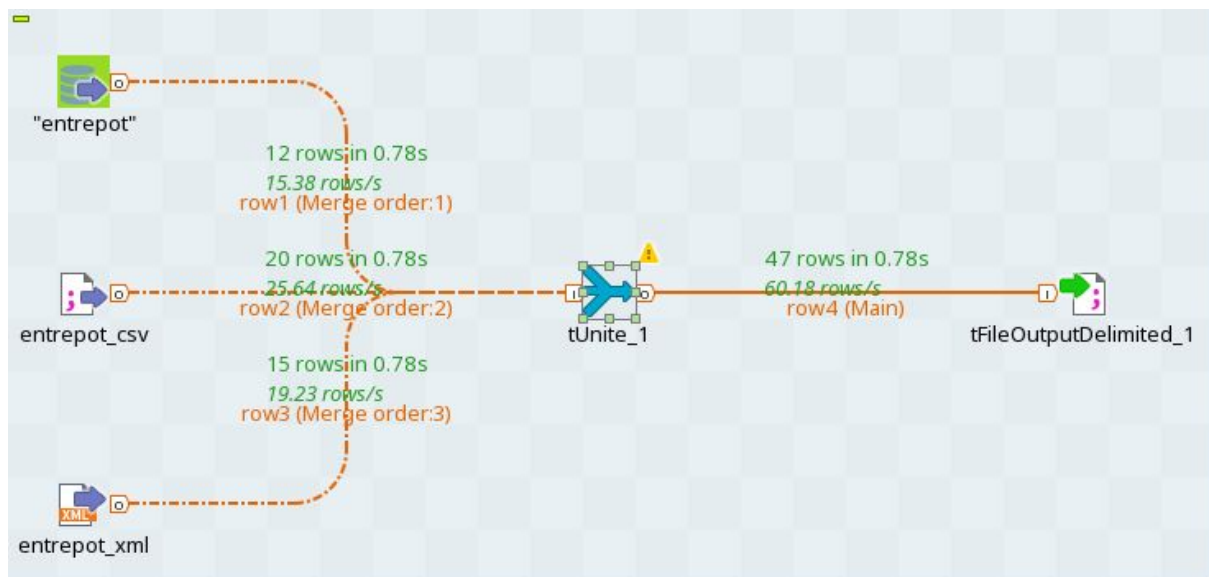
Diagramme Uml de la structure des bases de données:



Hypothèses :

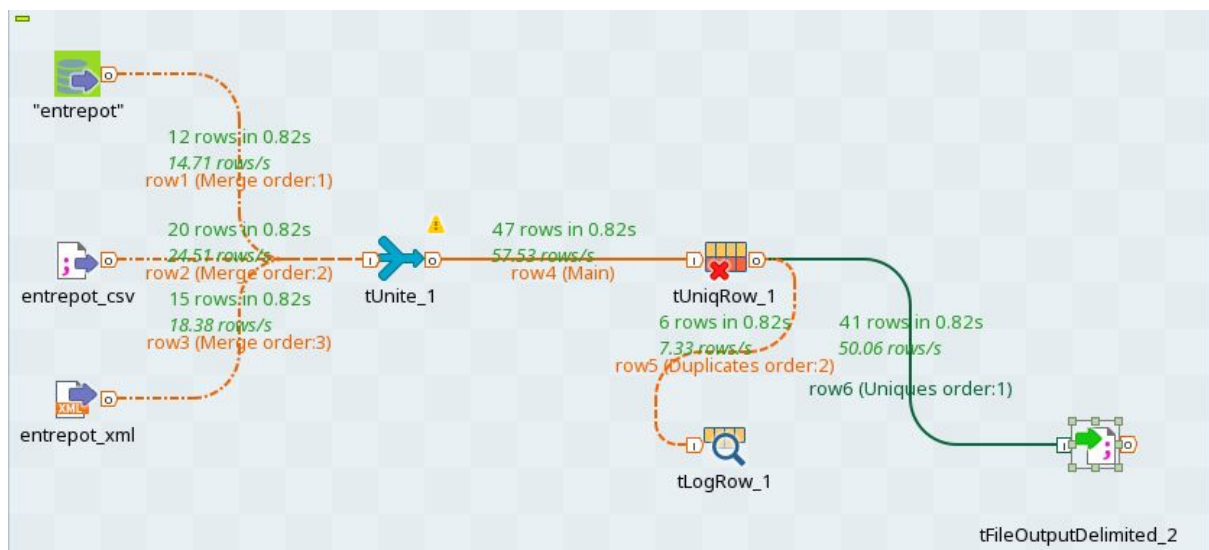
- ⇒ Un entrepôt n'appartient qu'à un seul distributeur.
- ⇒ Un conteneur est distribué par un seul distributeur.
- ⇒ Un conteneur Transporte qu'un seul type de produit.
- ⇒ Un type de Produit peut être transporté dans plusieurs conteneurs.
- ⇒ La capacité du conteneur est mesuré en unité de poid comme le poid du produit.
- ⇒ 3 types de stockage de base de données seront présentés dans notre étude de cas: XML, CSV, SQL .
- ⇒ 4 types de problemes seront explicités dans cette étude de cas:
 - ⇒ Problème de doublons.
 - ⇒ Problème de granularité.
 - ⇒ Problème de complétude.
 - ⇒ Problème de Conformité à un format.

Premier test : Probleme de doublons:



Exemple de test sur les tables "entrepôt" des différents base de données

La première étape vise à regrouper les 3 sources différentes afin qu'on puisse détecter les problèmes en utilisant le composant **tUnit** qui centralise des données provenant de sources diverses et hétérogènes.

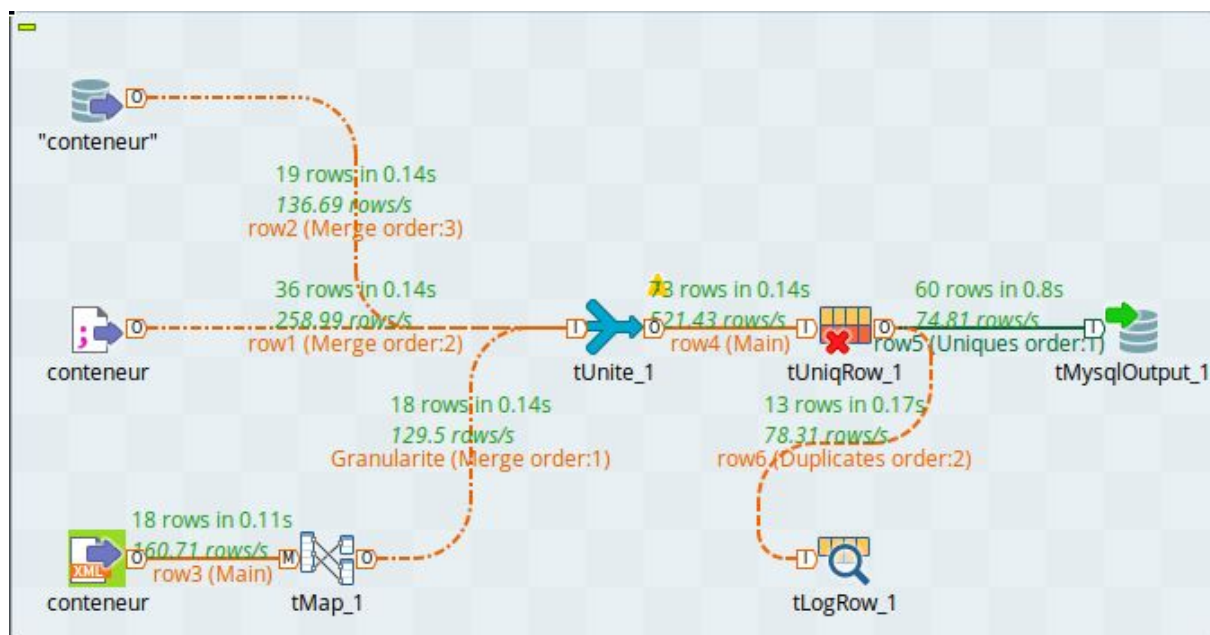


Exemple de résolution du problème des doublons sur les table "entrepôt".

Pour détecter les doublons, nous avons choisi 4 critères (*adr_ent*, *codep_ent*, *ville_ent*, *capacite_ent*), sous Talend on utilise le composant **tUniqRow** qui compare et supprimer les doublons du flux d'entrées.

N.B: tLogRow: Affiche les données ou les résultats dans la console RUN.

#1. tLogRow_1	#3. tLogRow_1	#5. tLogRow_1
key	key	key
value	value	value
num_ent	num_ent	num_ent
adr_ent	adr_ent	adr_ent
codep_ent	codep_ent	codep_ent
ville_ent	ville_ent	ville_ent
capacite_ent	capacite_ent	capacite_ent
numdist	numdist	numdist
#2. tLogRow_1	#4. tLogRow_1	#6. tLogRow_1
key	key	key
value	value	value
num_ent	num_ent	num_ent
adr_ent	adr_ent	adr_ent
codep_ent	codep_ent	codep_ent
ville_ent	ville_ent	ville_ent
capacite_ent	capacite_ent	capacite_ent
numdist	numdist	numdist



Le problème de granularité peut être trouvé sur l'attribut '**capacite**' de la table Conteneur (en g ou en Kg).

Pour rectifier le problème on calcule la division des valeurs sur 1000 pour avoir tous les valeurs en "Kg" en utilisant le composant tMap où on peut définir des fonctions mathématiques, et où en fait aussi la jointure avec le flux de sortie.