

## Exercice : ETL

a) Quelle composante Talend serait appropriée pour extraire des données d'un fichier plat comme csv ? **tFileInputDelimited**

b) Vous avez besoin de concaténer le noms et prénom du client. Quelle composante Talend utiliseriez-vous ? Décrivez les étapes à suivre. **tMap**

**Etapes à suivre :**

1. Ajouter le composant tMap entre la source et la sortie
2. Dans tMap, créer une nouvelle colonne de sortie nommée par exemple 'nom\_complet'
3. Dans l'expression de cette colonne, utiliser : `row1.prenom + " " + row1.nom`

c) Vous devez nettoyer les données extraites en supprimant les doublons. Quelle composante Talend utiliseriez-vous ? **tUniqRow**

d) Quelle composante Talend utiliseriez-vous pour attribuer un ID en séquence à chaque enregistrement dans le fichier des produits pharmaceutiques ? Donnez le code Java à utiliser.

**tMap**

**Code Java :**

**`Numeric.sequence("S1", 1, 1)`**

e) Quelle composante Talend utiliseriez-vous pour exécuter cette requête :

`"Select *`

`From médicaments`

`Order by catégorie ;"`

**tSortRow**

f) Dans quel cas utiliseriez-vous la composante Talend TFilterRow dans le cadre de l'entreprise pharmaceutique ? **Un cas d'usage : Filtrer les ventes dépassant un certain montant.**

g) Comment utiliseriez-vous la composante Talend TAggregateRow pour calculer des statistiques sur les ventes de médicaments ?

**Les étapes à suivre pour calculer le montant total des ventes par produit :**

1. Connecter tAggrateRow à la source
2. Dans la zone Group by, ajouter la colonne produit
3. Dans la zone Operation, ajouter la colonne « montant », et choisir la fonction « Somme »

h) Quelle composante Talend utiliseriez-vous pour exécuter cette requête :

`Select *`

`From client`

`Limit 5,7..9,20 ; "`

**tSampleRow**

i) Quelle composante Talend utiliseriez-vous pour combiner les données de plusieurs sources de ventes de médicaments en un seul fichier de sortie pour une analyse globale ? **tUnite**