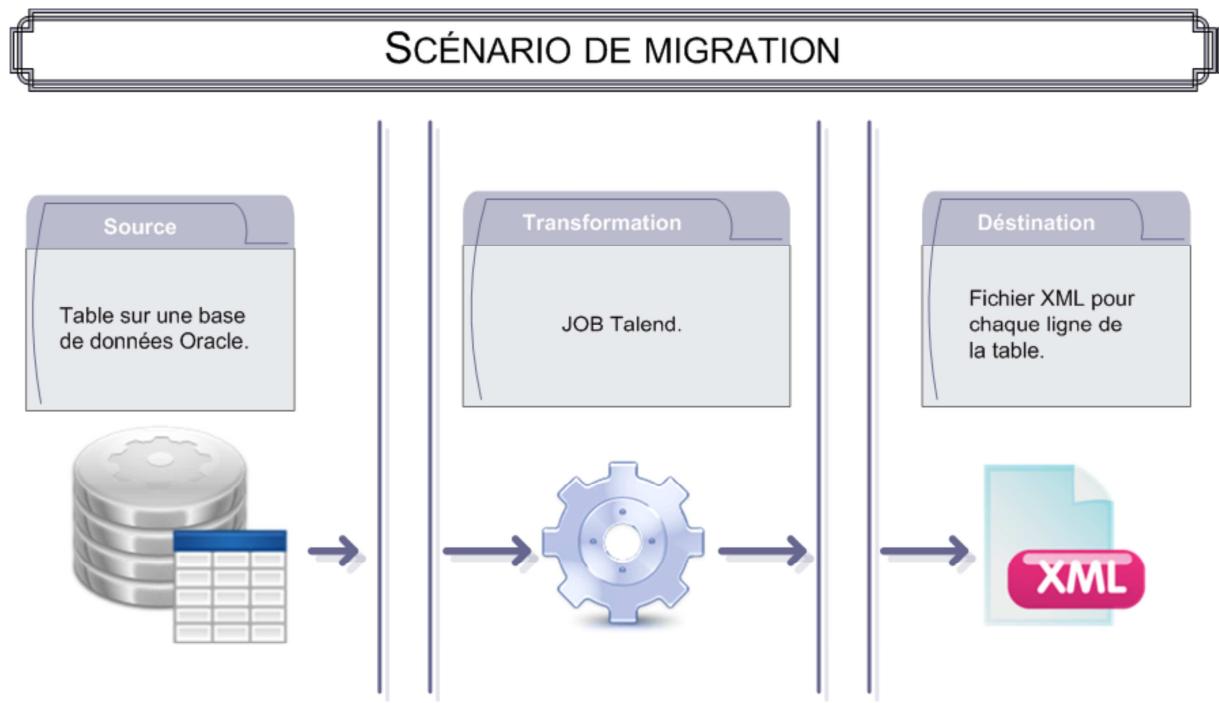


TPS : Création de job avec l'ETL Talend Open Studio

Ce document est issu de <http://www.developpez.com> et reste la propriété exclusive de son auteur
<http://haskouse.developpez.com/tutoriels/etl/talend-open-studio/creation-job/>

Le document suivant nous présente les étapes de création d'un job sous Talend Open Studio, il aura comme but de migrer les lignes d'une table d'une base de données Oracle vers des fichiers XML. Le schéma suivant représente ce scénario :



II - La source de données

Avant de parler de la création du job TOS on va tout d'abord présenter la source de données, qui n'est autre qu'une table dans une base de données Oracle qui a été créée dans le schéma « TEST » avec le script suivant :

```
CREATE TABLE CLIENT(
  CODE VARCHAR2(100 BYTE),
  NOM VARCHAR2(100 BYTE),
  PRENOM VARCHAR2(100 BYTE),
  ADRESSE VARCHAR2(100 BYTE),
  EMAIL VARCHAR2(100 BYTE)
);
```

Puis des lignes ont été insérées dans cette table en utilisant ce script :

```
INSERT INTO CLIENT
VALUES
('cli001',
'ELHASSAK',
'MUSTAPHA',
'N8 BLOC10 RUE20 CASA',
'ELHASSAK.M@HOTMAIL.COM');
INSERT INTO CLIENT
VALUES
('cli002',
```

```

'Esbais',  

'MOUAD',  

'N9 BLOC310 RUE21 RABAT',  

'ESBAISS.M@HOTMAIL.COM');  

INSERT INTO CLIENT  

VALUES  

('cli003',  

'zAKI',  

'ANASS',  

'N18 BLOC1 RUE33 CASA',  

'ZAKI.A@HOTMAIL.COM');  

INSERT INTO CLIENT  

VALUES  

('cli004',  

'MoufRIJE',  

'KHALID',  

'N82 BLOC33 RUE220 FES',  

'MOUFRIJE.K@HOTMAIL.COM');  

INSERT INTO CLIENT  

VALUES  

('cli005',  

'hadri',  

'ZAKARIA',  

'N1 BLOC21 RUE10 TANGER',  

'HADRI.Z@HOTMAIL.COM');

```

Finalement on se retrouve avec le résultat suivant :

III - La destination des données

SQL> select * from client;				
CODE	NOM	PRENOM	ADRESSE	EMAIL
cli001	ELHASSAK	MUSTAPHA	N8 BLOC10 RUE20 CASA	ELHASSAK.M@HOTMAIL.COM
cli002	Esbais	MOUAD	N9 BLOC310 RUE21 RABAT	ESBAISS.M@HOTMAIL.COM
cli003	zAKI	ANASS	N18 BLOC1 RUE33 CASA	ZAKI.A@HOTMAIL.COM
cli004	MoufRIJE	KHALID	N82 BLOC33 RUE220 FES	MOUFRIJE.K@HOTMAIL.COM
cli005	hadri	ZAKARIA	N1 BLOC21 RUE10 TANGER	HADRI.Z@HOTMAIL.COM

Dans cette partie on va présenter le format de sortie de nos données, autrement dit du fichier XML résultant du job.

Ce dernier aura la forme suivante :

```

<?xml version="1.0">
<Client code="_CODE_">
<Nom>_NOM_</Nom>
<Prenom>_PRENOM_</Prenom>
<Coordonnes>
<Adresse>_ADRESSE_</Adresse>
<Email>_EMAIL_</Email>
</Coordonnes>
</Client>

```

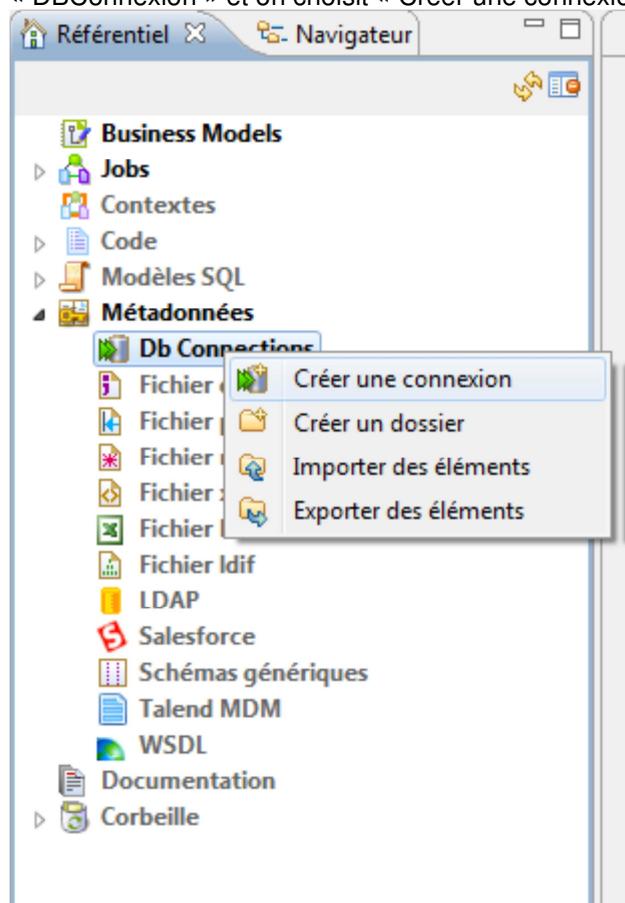
Les valeurs entre deux " _ " représentent les parties qui vont changer suivant chaque ligne de la table. Après cette description des données d'entrée et du format de sortie, on va montrer les étapes de création de notre job TOS.

IV - Job Talend Open Studio

La création du job TOS va se dérouler suivant les étapes :

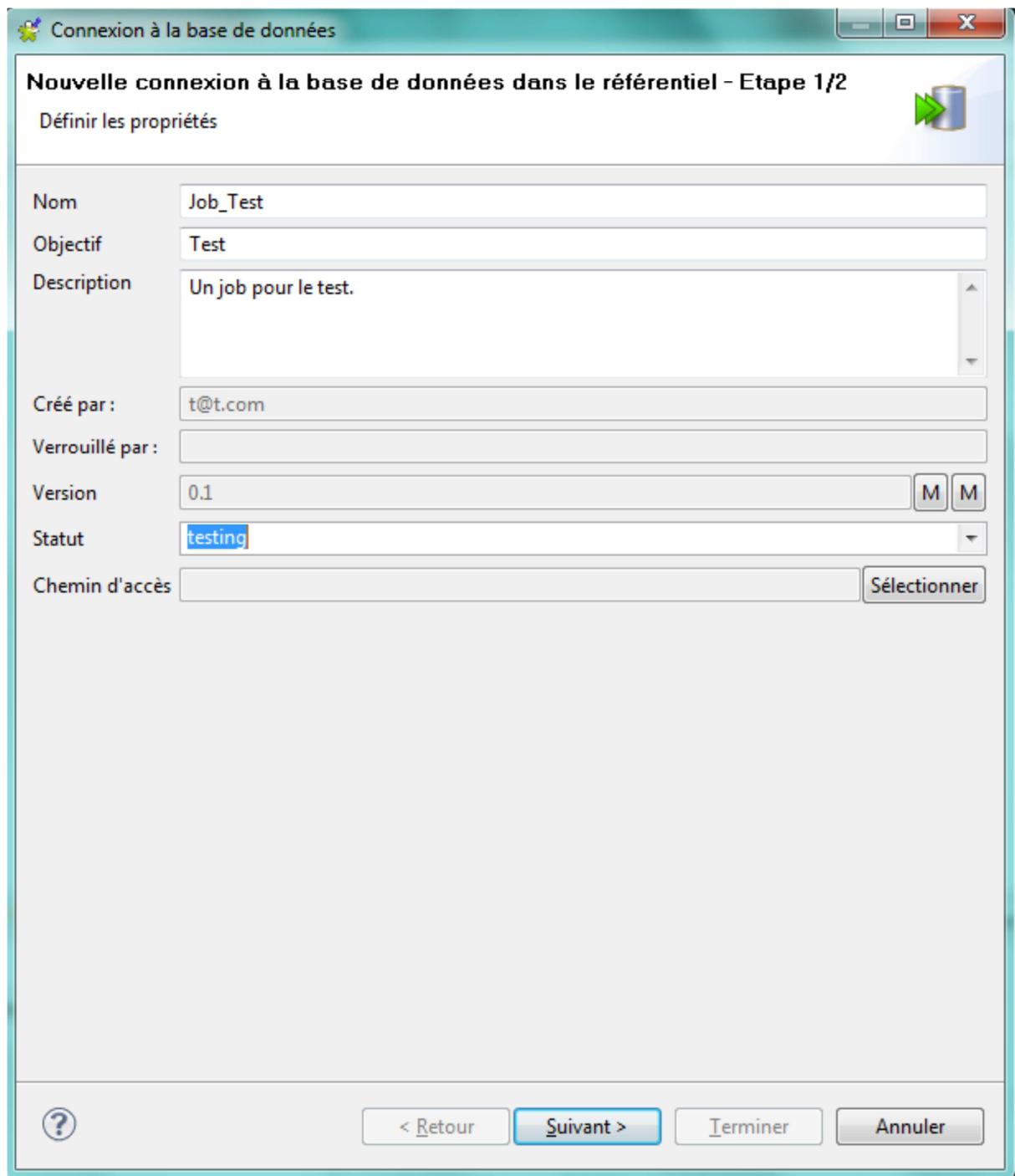
- création d'une connexion avec la base de données ;
- création du schéma de notre table ;
- création de la requête SQL à utiliser ;
- définition du job.

Dans l'onglet « Référentiel » on développe l'arborescence jusqu'à « Métadonnées » puis on fait un clic droit sur « DBConnexion » et on choisit « Créer une connexion ».

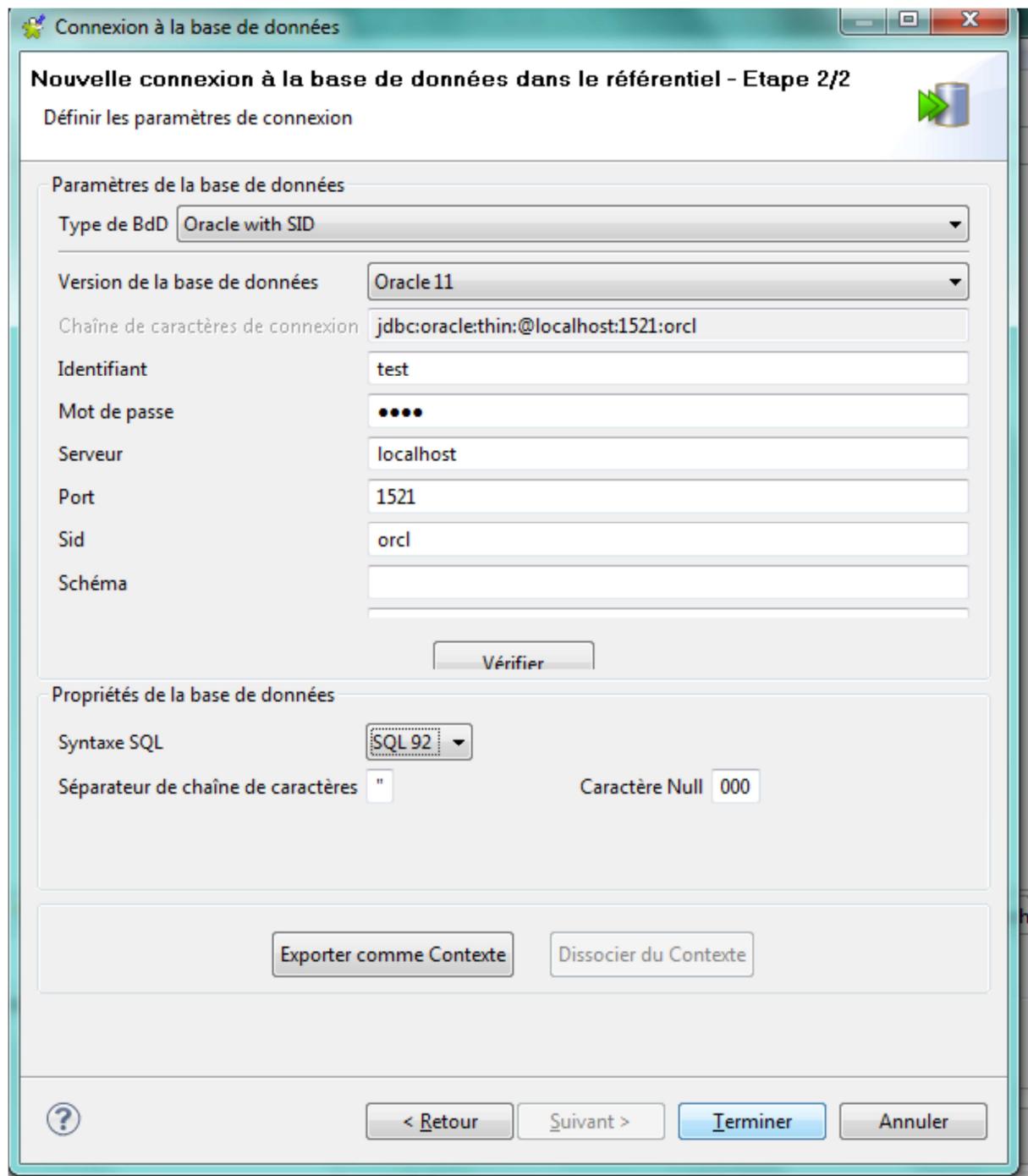


L'assistant de création de connexion s'affiche.

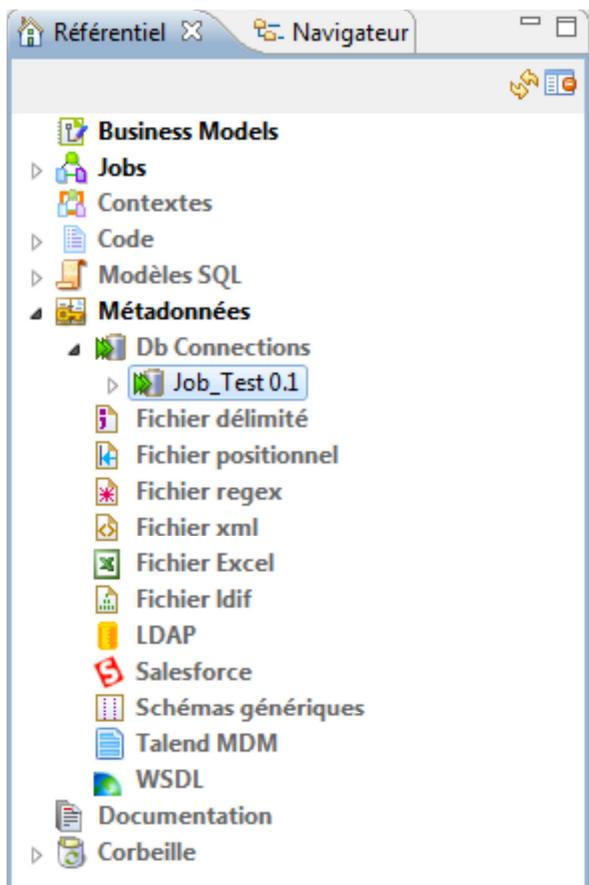
- On remplit les différents champs en saisissant le nom, l'objectif et une description.
- On clique sur « Suivant ».



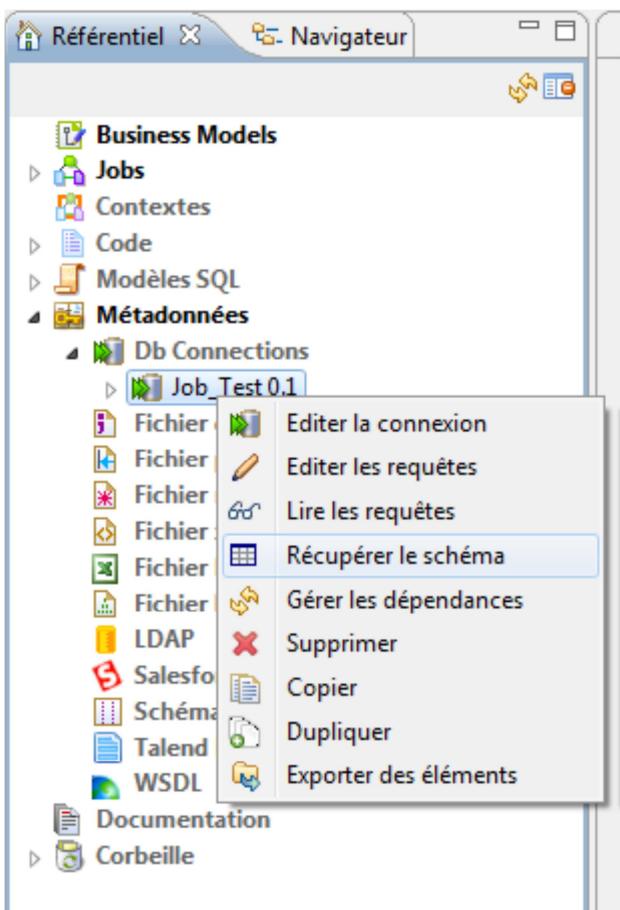
- On choisit le type de la BD dans le combobox. Dans notre exemple c'est Oracle.
- On remplit les différentes informations pour se connecter à la base.
- On clique sur « Terminer ».



Voilà notre connexion est créée et on peut la voir dans l'arborescence de l'onglet « référentiel ».

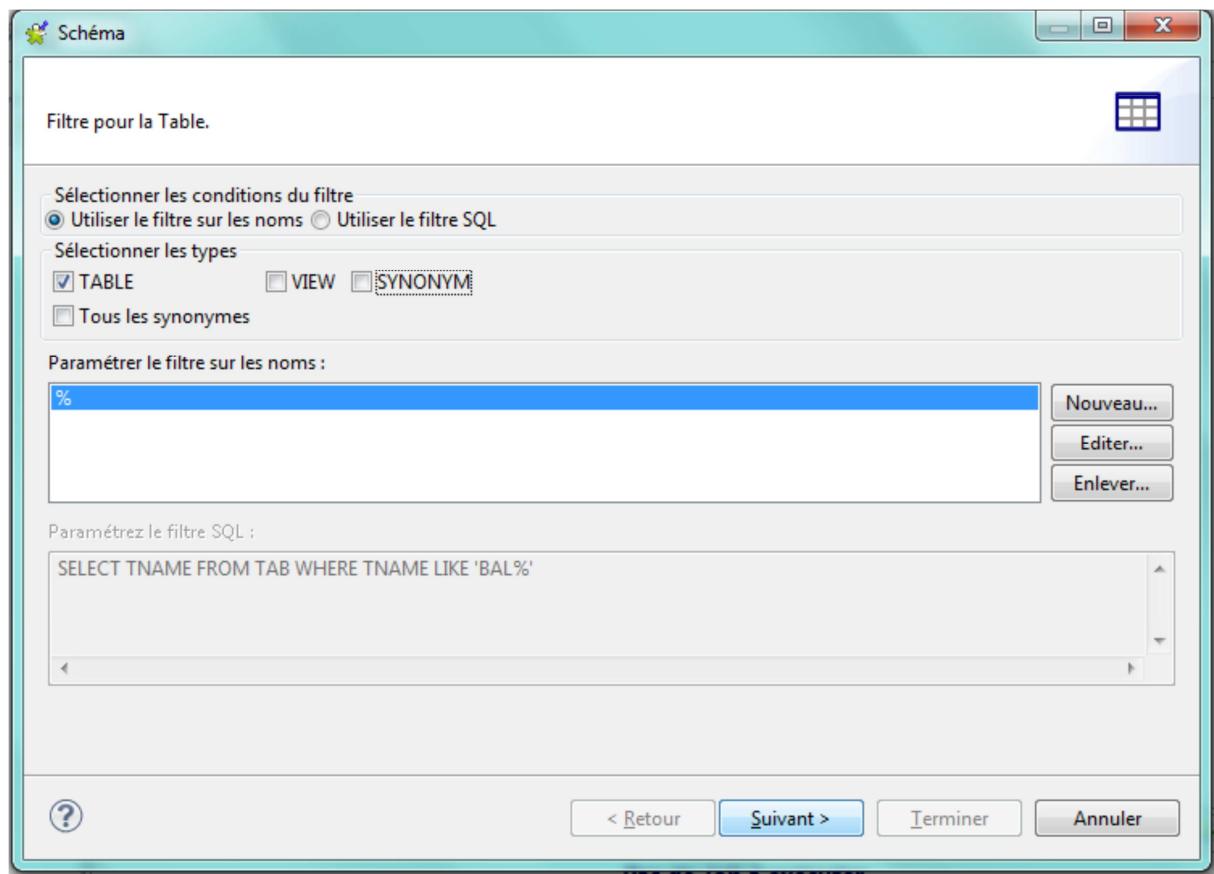


Dans l'onglet « Référentiel » on développe l'arborescence jusqu'à arriver à la connexion précédemment créée. On fait un clic droit puis on clique sur « Récupérer le schéma ».

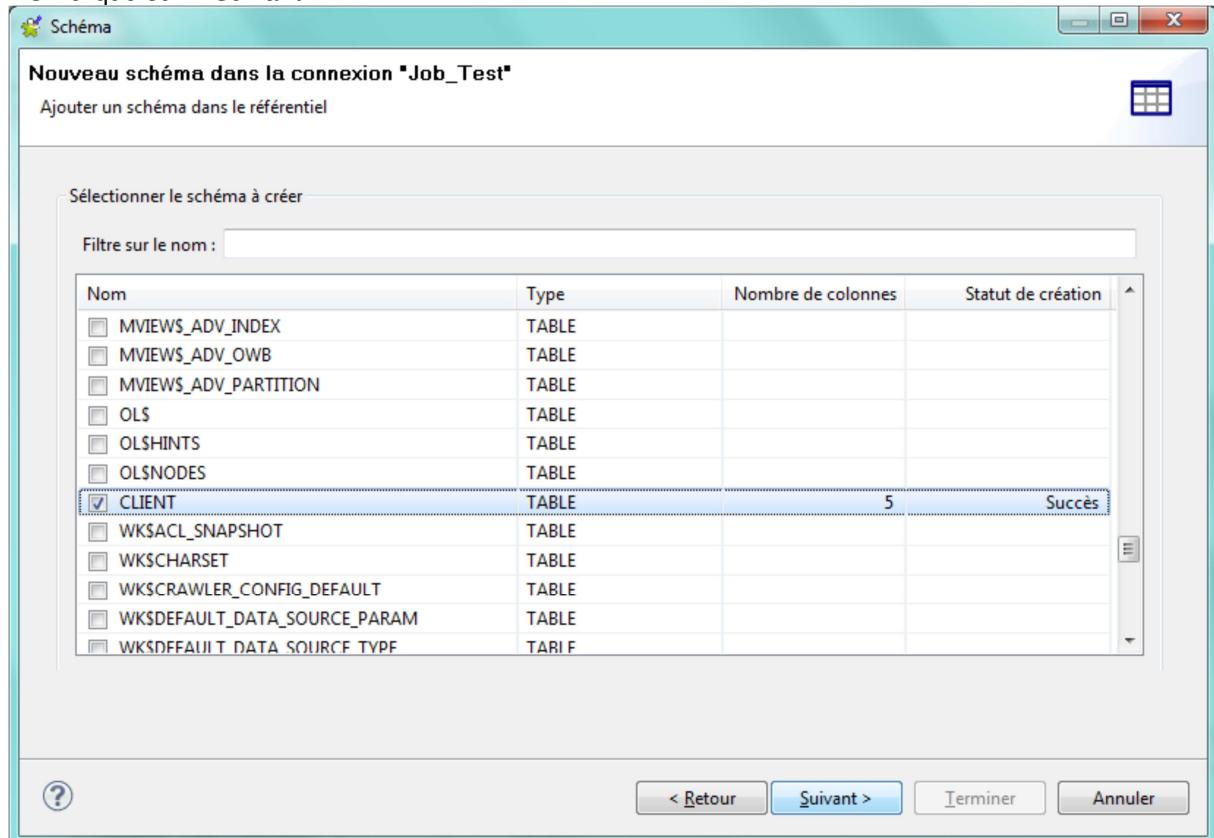


L'assistant de récupération de schéma s'affiche.

- On sélectionne le type d'objet à récupérer , ici on choisit « TABLE ».
- On clique sur « Suivant ».

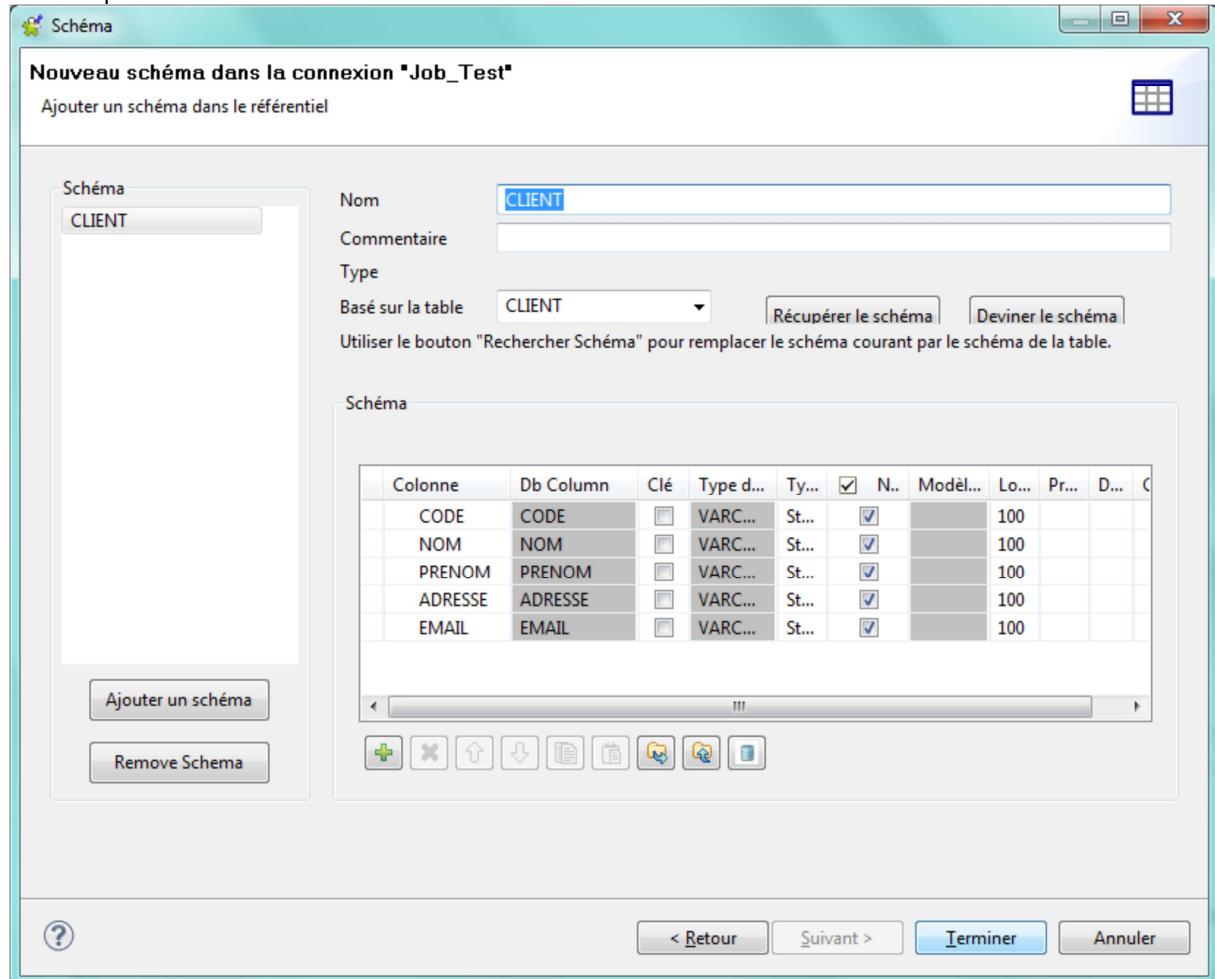


- Les noms de toutes les tables s'affichent, on coche la ou les tables qui nous intéressent.
- On clique sur « Suivant ».

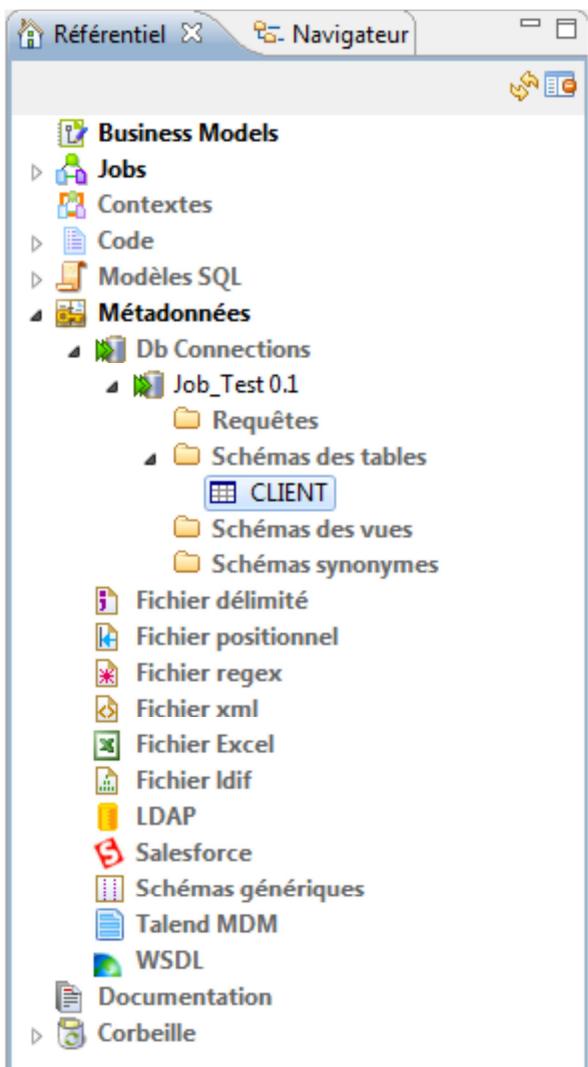


- On donne un nom au schéma.
- On vérifie les colonnes récupérées par l'assistant dans la vue « Schéma ».

- On clique sur « Terminer ».

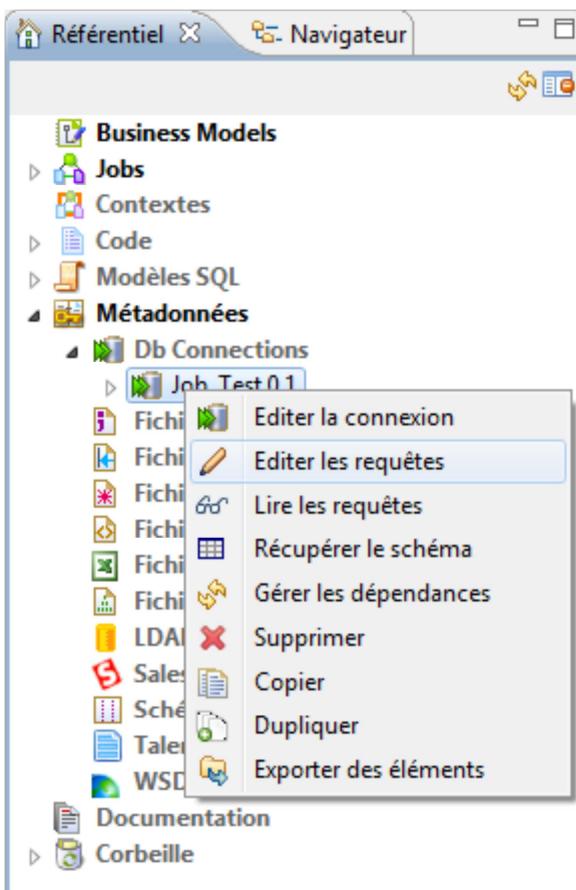


Le schéma est maintenant créé et on le voit dans l'arborescence.



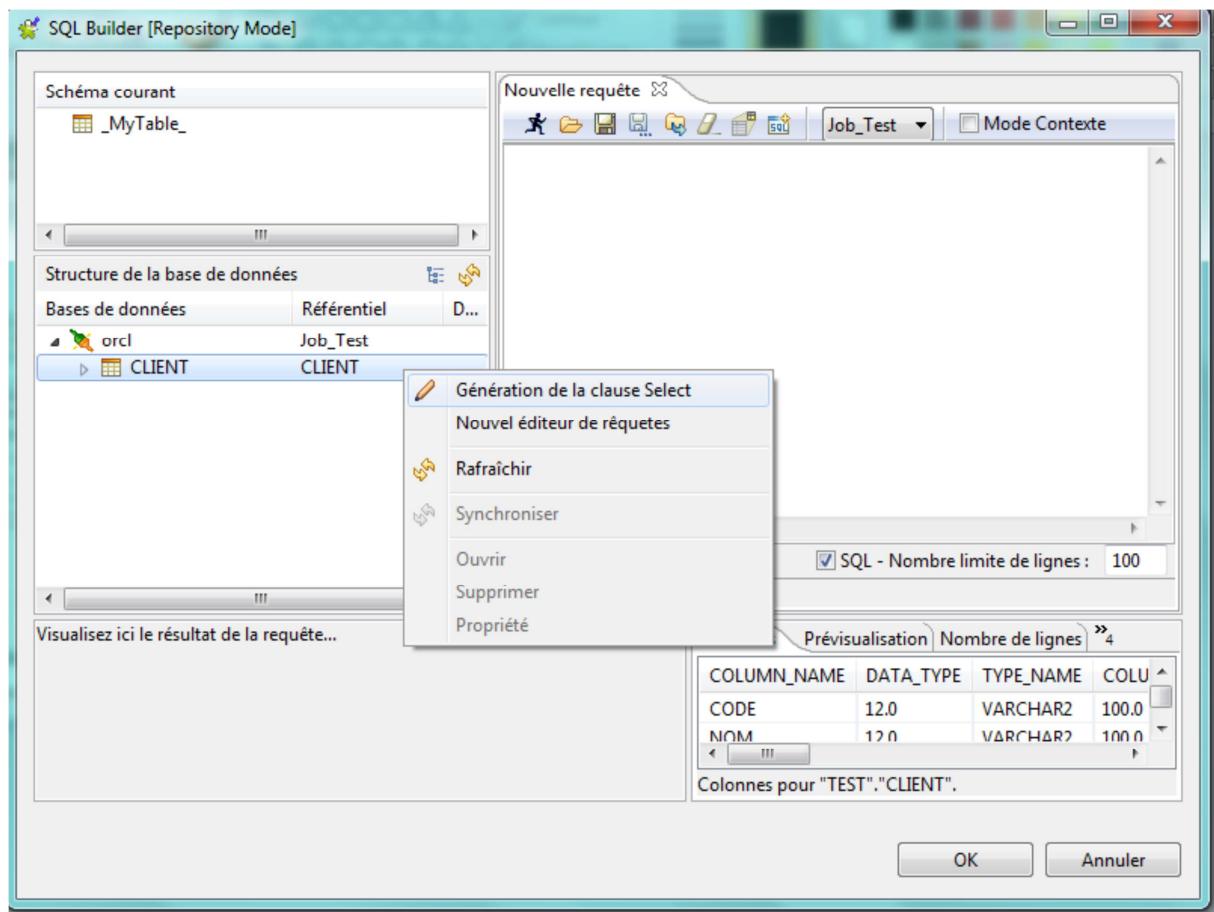
IV-C - Création de la requête SQL à utiliser

Dans l'onglet « Référentiel » on développe l'arborescence jusqu'à arriver à la connexion précédemment créée. On fait un clic droit puis on clique sur « Editer les requêtes ».



L'assistant de création de requêtes s'affiche.

- Dans la partie « Structure de la base de données » on sélectionne la table qui nous intéresse , on fait un clic droit et on choisit : « Génération de la clause select ».



- La requête est créée à droite, il ne nous reste plus qu'à la sauvegarder en lui donnant un nom.

SQL Builder [Repository Mode]

Schéma courant: _MyTable_

Nouvelle requête *Nouvelle requête

Bases de données Référentiel

orcl Job_Test

CLIENT CLIENT

Structure de la base de données

CLIENT

CODE
NOM
PRENOM
ADRESSE

select CLIENT.CODE,CLIENT.NOM,CLIENT.PRENOM,CLIENT.ADRESSE,CLIENT.EMAIL
from CLIENT

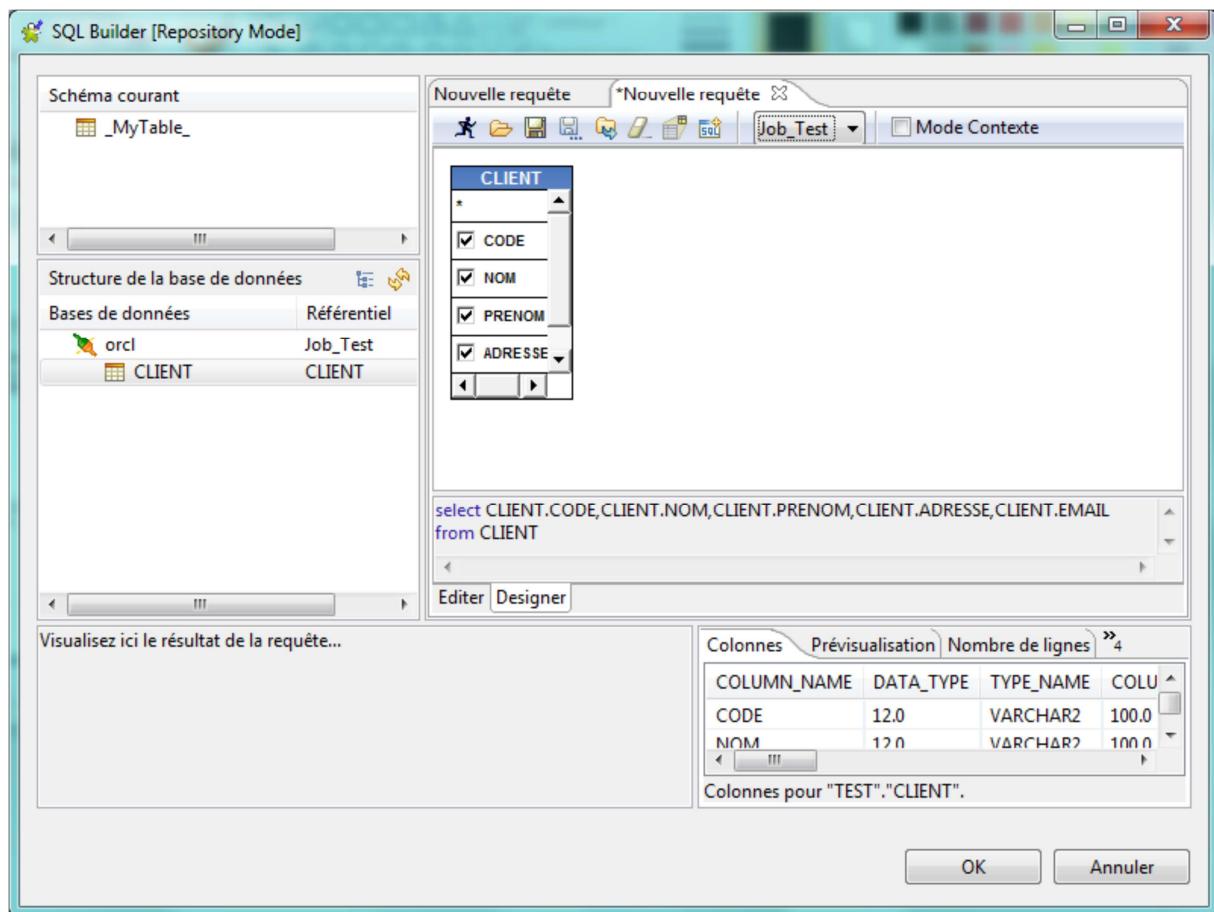
Visualisez ici le résultat de la requête...

Colonnes Prévisualisation Nombre de lignes »4

COLUMN_NAME	DATA_TYPE	TYPE_NAME	COLU
CODE	12.0	VARCHAR2	100.0
NOM	12 0	VARCHAR2	100 0

Colonnes pour "TEST"."CLIENT".

OK Annuler



"Nouvelle requête" Enregistrer la requête sous

Enregistrer sous

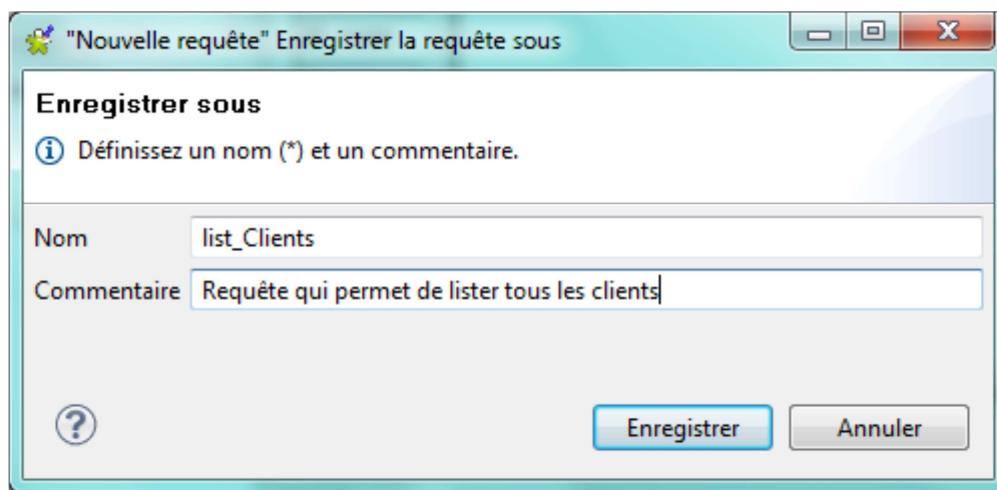
❶ Définissez un nom (*) et un commentaire.

Nom: list_Clients

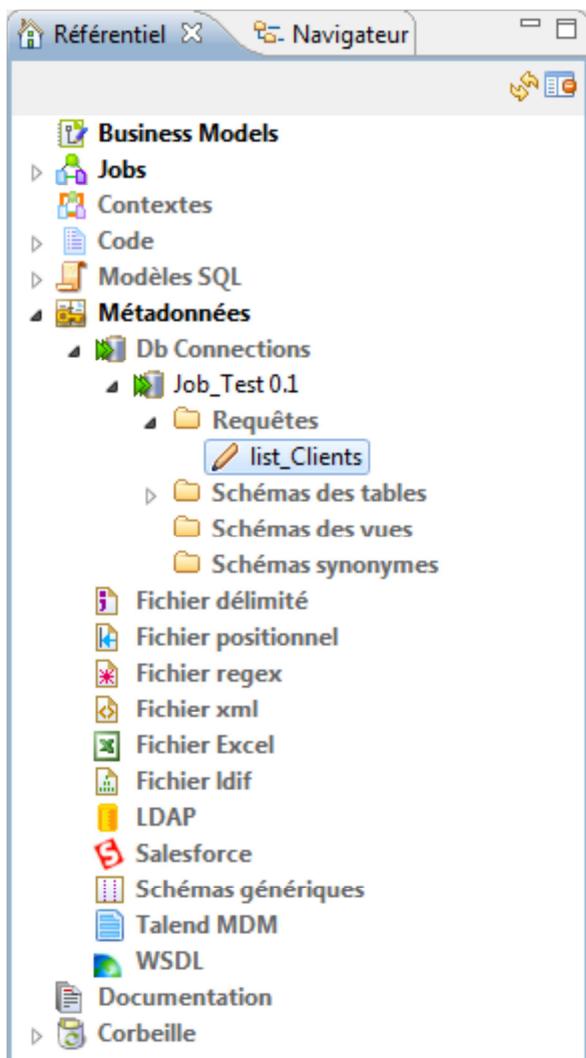
Commentaire: Requête qui permet de lister tous les clients

?

Enregistrer Annuler

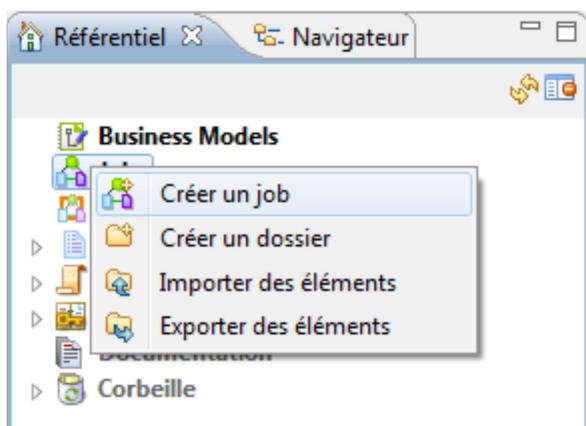


- Finalement on clique sur « OK » pour terminer.
Notre requête est créée, elle apparaît dans l'arborescence.



IV-D - Définition du job

Toujours dans l'onglet « Référentiel » on fait un clic droit sur « Jobs », ensuite on clique sur « Créer un job ». L'assistant de création de job s'affiche.



Nouveau Job

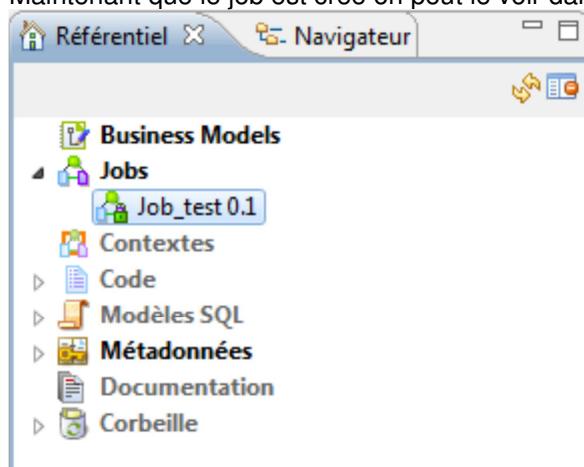
Ajoute un Job dans le Référentiel

Nom	Job_test
Objectif	Création de fichier XML
Description	Transformation de chaque ligne de la table "client" à un fichier XML.
Créé par :	t@t.com
Verrouillé par :	
Version	0.1
Statut	testing
Chemin d'accès	<input type="button" value="Sélectionner"/>

?

- On remplit les différents champs décrivant le job : Nom, Objectif, Description.
- On clique sur « Terminer ».

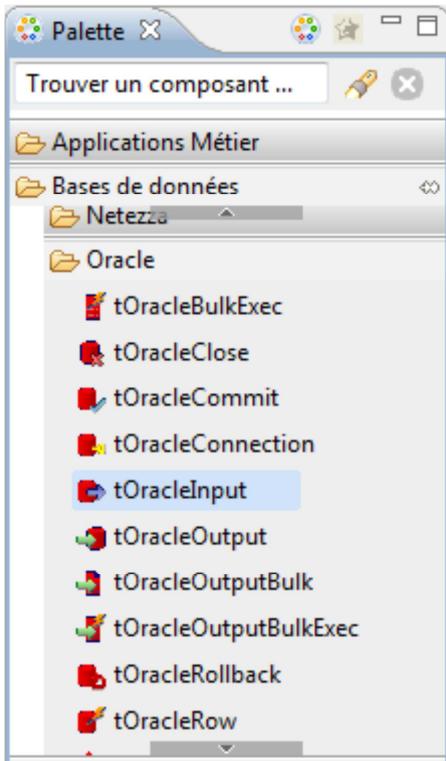
Maintenant que le job est créé on peut le voir dans l'arborescence.



On va déposer ses différents composants.

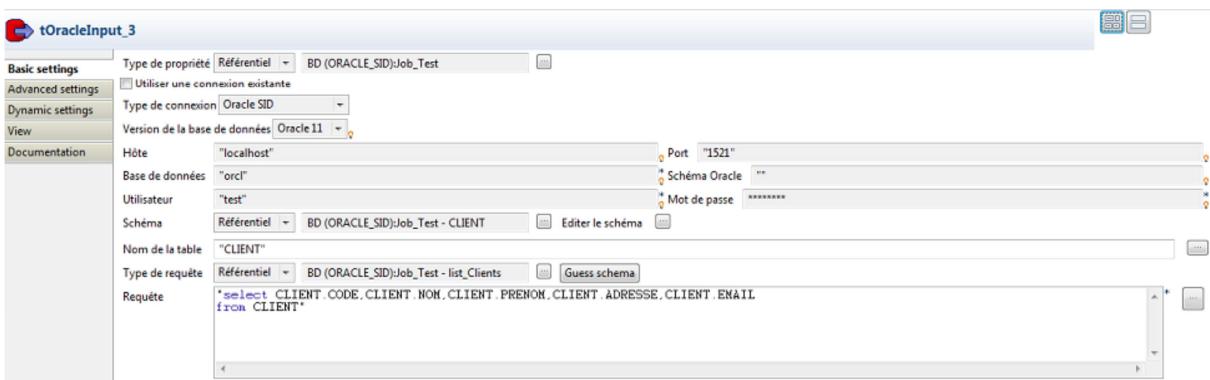
IV-D-1 - Création de la connexion Oracle

- Dans l'onglet « Palette » on choisit le composant « tOracleInput », on le glisse et le dépose dans le « Designer » du job.



- On sélectionne le composant « tOracleInput », puis on va dans l'onglet « Component ».
- On remplit les différents champs pour configurer la connexion avec la BD, ou bien on choisit une connexion déjà créée ce qui est notre cas.
- Dans le combobox « Type de propriété » on choisit la valeur : « Référentiel ».
- On clique sur le bouton « . » qui apparaît à droite, puis on parcourt les différentes connexions pour choisir celle qui nous intéresse. Puis on clique sur « OK ».
- Après ça dans le combobox « Schéma » on choisit la valeur « Référentiel » puis récupérer le schéma déjà créé (ce qui est notre cas) , ou bien on clique sur « éditer le schéma » pour en créer un nouveau.

- On clique sur le premier bouton « . » pour choisir notre schéma. Une fenêtre s'affiche avec les différents schémas créés. On choisit celui qui nous intéresse. Puis on clique sur « OK ».

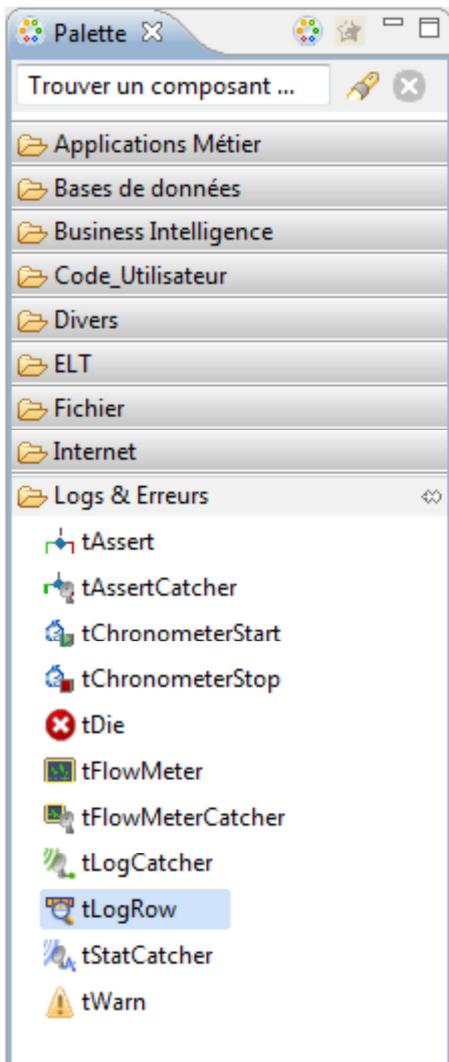


- Après ça il faut spécifier la requête à utiliser. Dans le combobox « Type de requête » on choisit la valeur « Référentiel ».
- On clique sur le bouton « . » qui apparaît à droite, puis on parcourt les différentes requêtes pour choisir celle qui nous intéresse. Puis on clique sur « OK ».

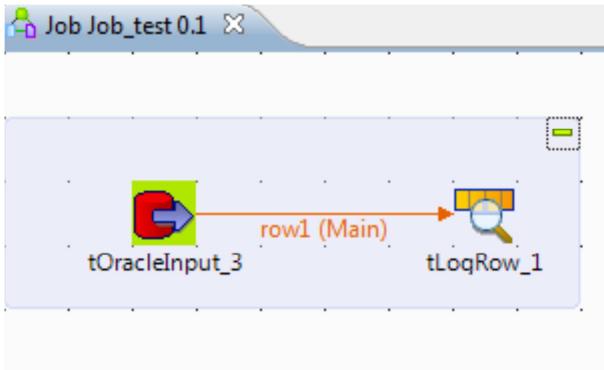
Voilà notre connexion avec la base est configurée. On passe maintenant à créer le composant de journalisation.

IV-D-2 - Création du composant de journalisation

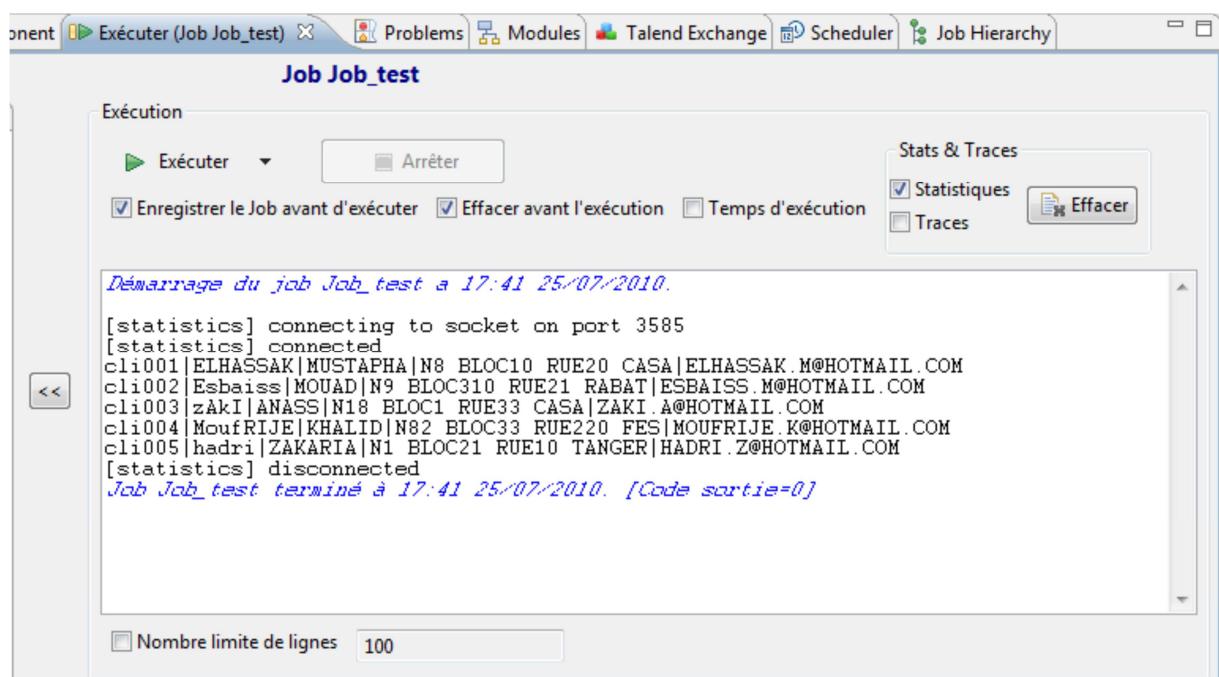
- Dans l'onglet « Palette » on choisit le composant « tLogRow », on le glisse et le dépose dans le « Designer » du job.



- On fait un clic droit sur le composant « tOracleInput » puis on le glisse vers le composant « tLogRow ».
- On sélectionne le composant « tLogRow », puis on va dans l'onglet « Component ».
- On clique sur le bouton « Synchroniser les colonnes » pour récupérer le schéma depuis le composant d'entrée (ici c'est le « tOracleInput »).
- On choisit les autres différentes options pour notre composant.

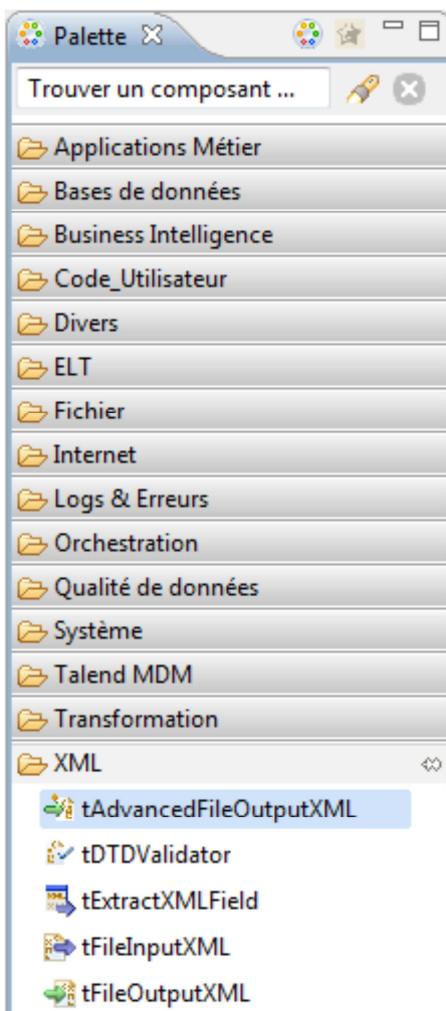


Le composant est maintenant fin prêt, et on peut déjà tester le job pour voir s'il marche en appuyant sur la touche « F6 » ou bien en allant dans l'onglet « Exécuter » et en cliquant sur le bouton « Exécuter ».

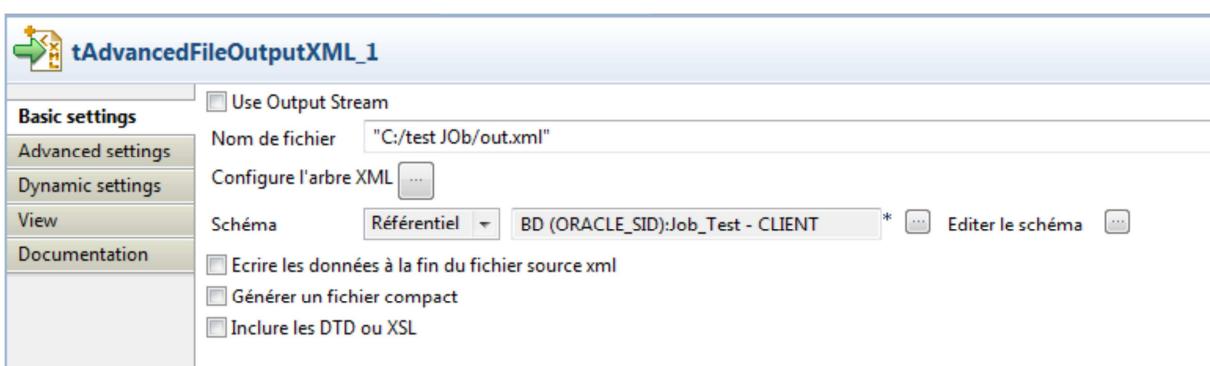


IV-D-3 - Création du fichier de sortie XML

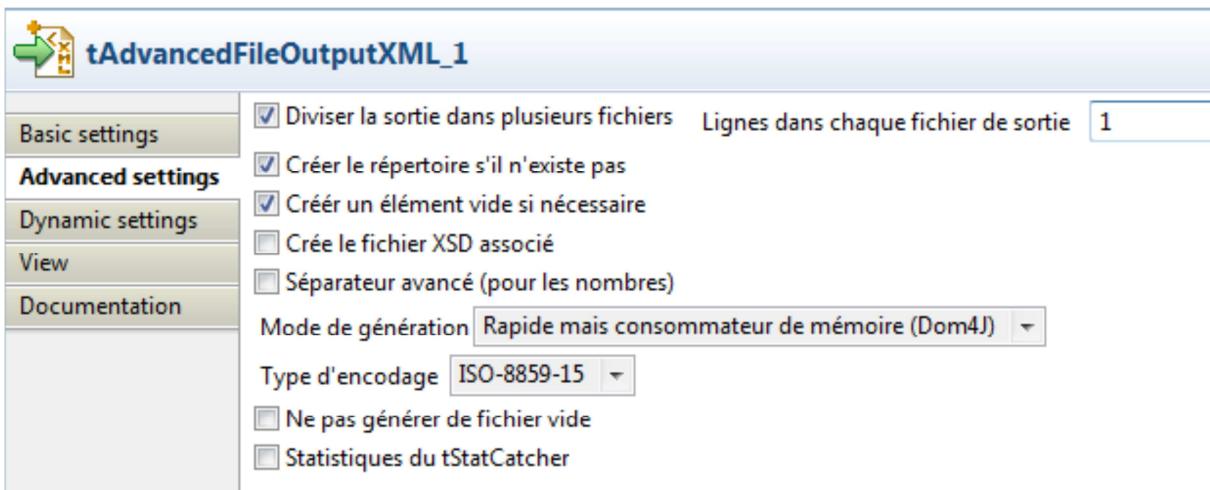
- Dans l'onglet « Palette » on choisit le composant « tAdvancedFileOutputXML », on le glisse et le dépose dans le « Designer » du job.



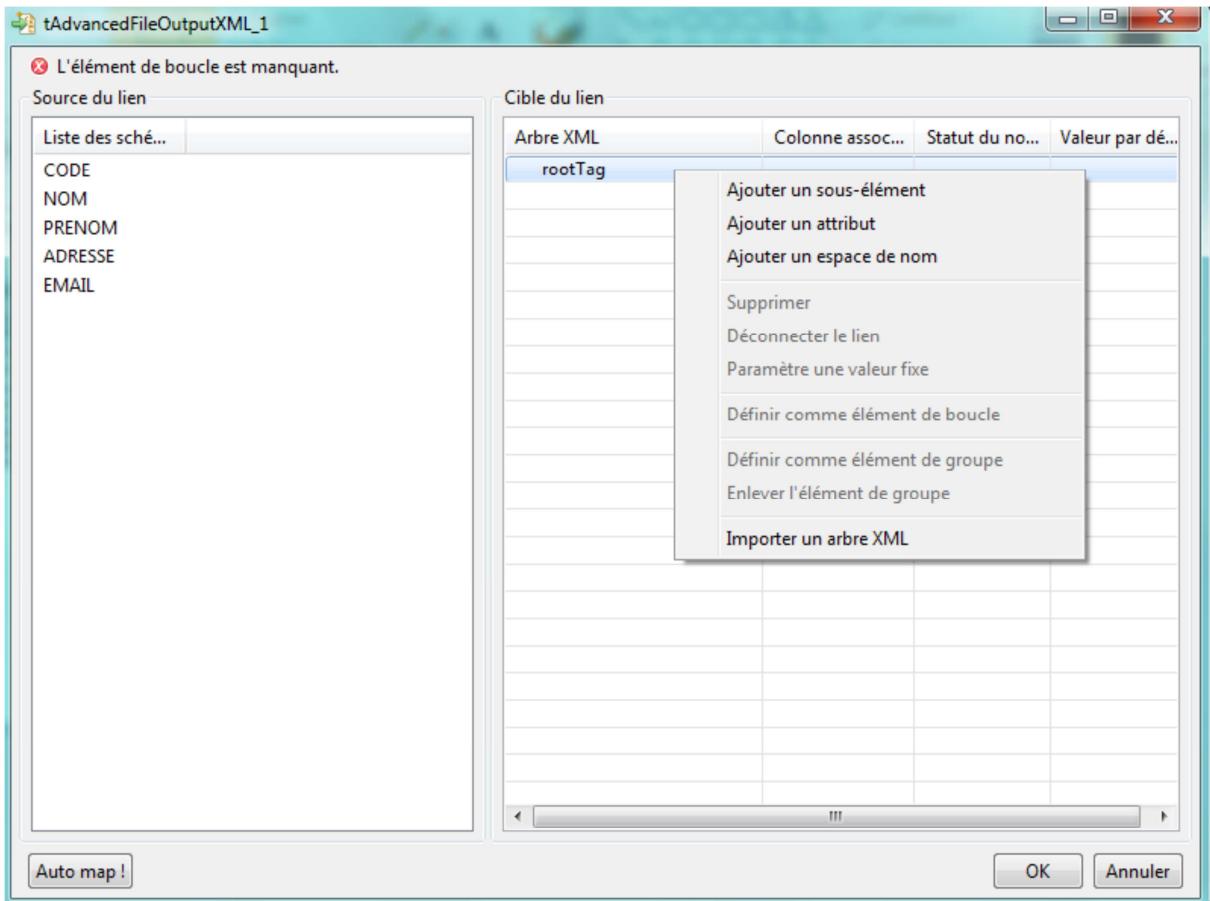
- On sélectionne le composant « tAdvancedFileOutputXML » puis on va dans l'onglet « Component ».
- On spécifie les différentes propriétés du composant, notamment le nom et chemin du fichier de sortie.



- On clique sur le bouton « . » devant le libellé « Configurer l'arbre XML » pour créer la structure de notre XML.



L'assistant de création de l'arbre XML s'affiche. On choisit soit de créer la structure manuellement, soit de faire un clic droit sur le « root Tag » puis cliquer sur « Importer un arbre XML », ce qui est notre cas ici. On sélectionne le fichier XML qui représente la structure désirée puis on clique sur « Open » et enfin sur « OK ».



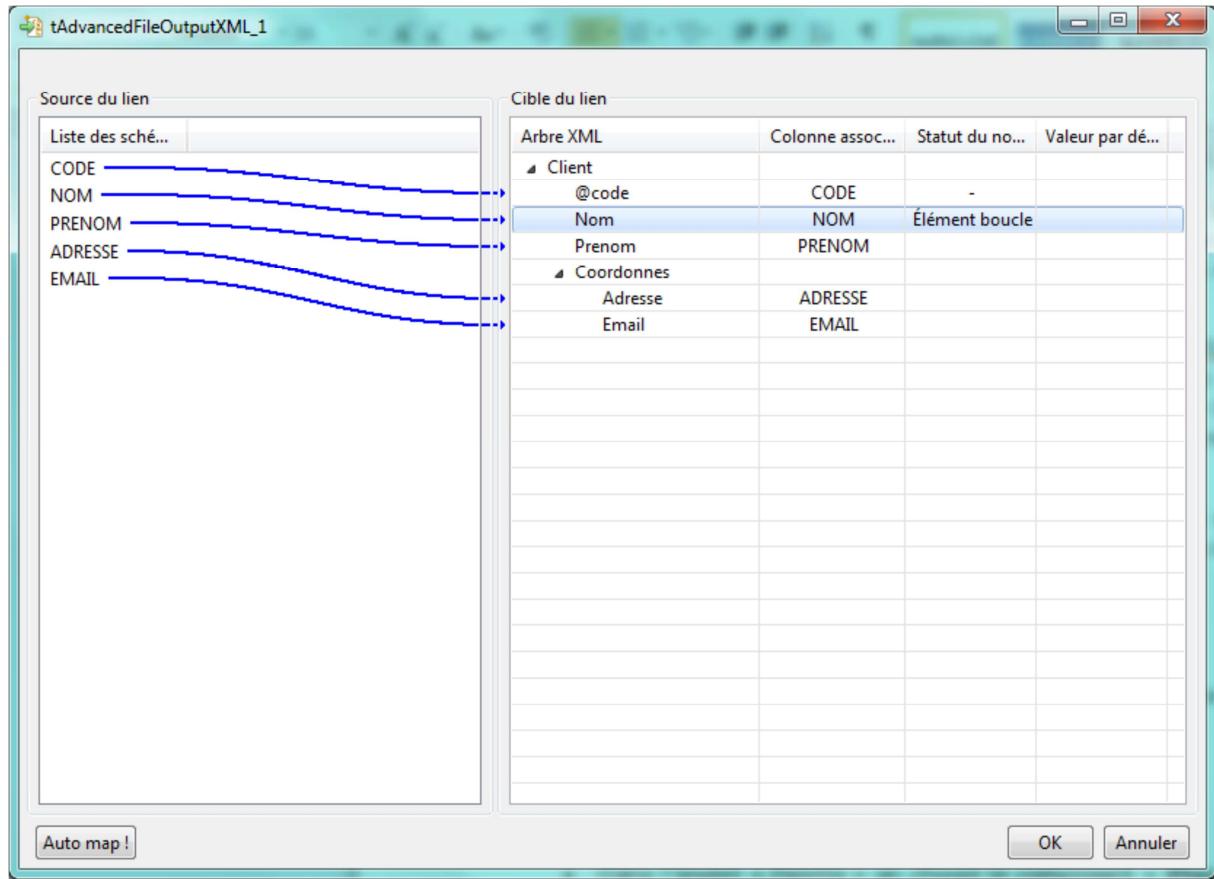
- Après on définit le schéma qui doit être mappé avec la structure du XML, pour cela soit on choisit un schéma du référentiel comme avant, soit on le crée manuellement. Dans notre exemple on choisira le même schéma que celui de la connexion Oracle.

- On fait un double-clic sur le composant « tAdvancedFileOutputXML » pour faire la relation entre les champs en entrée affichés dans la partie « Source du lien » et les balises de sortie dans la partie « Cible du lien ».

Pour cela il suffit de faire des glisser/déplacer depuis la partie gauche vers celle de droite.

- Il faut spécifier un élément de boucle , c'est la balise sur laquelle on va boucler et c'est obligatoire, on choisit ici la balise <Nom>.

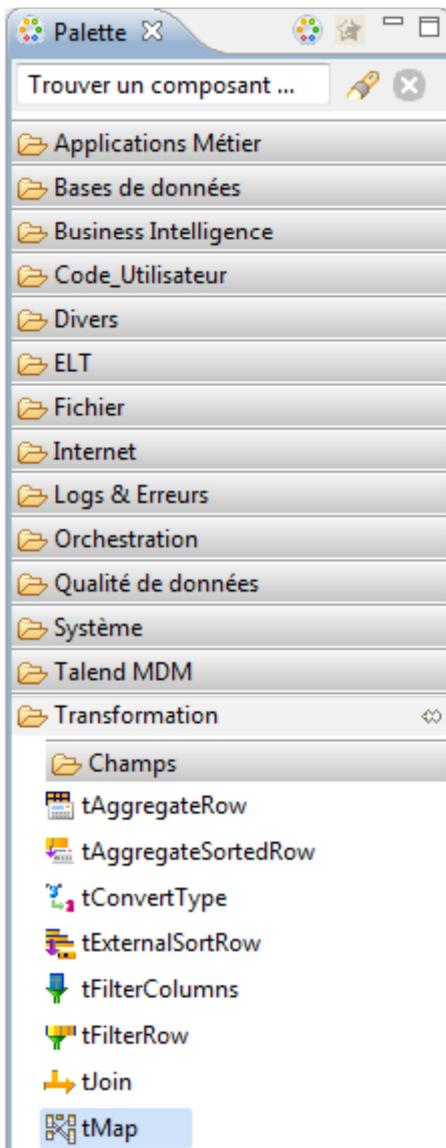
- On clique sur « OK ».



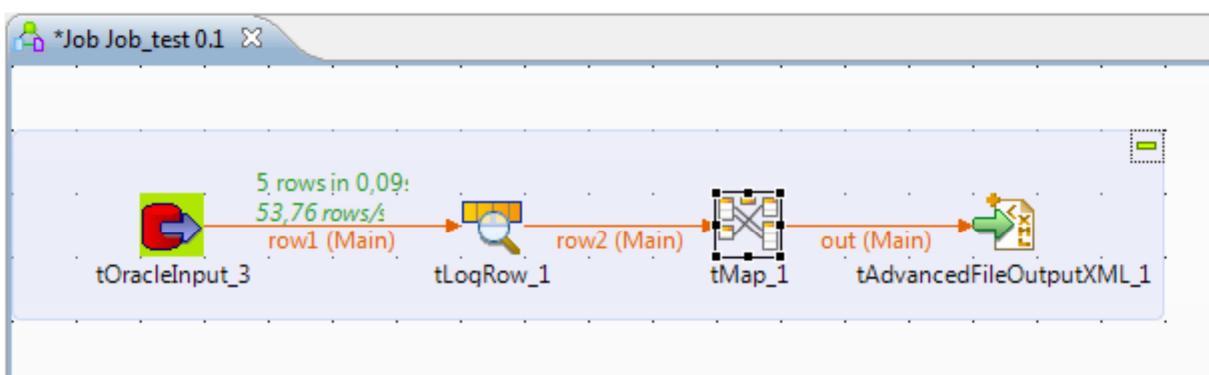
Maintenant on va placer un composant « tMap » qui nous permettra de faire quelques opérations sur les champs avant de les mapper.

IV-D-4 - Création du composant tMap

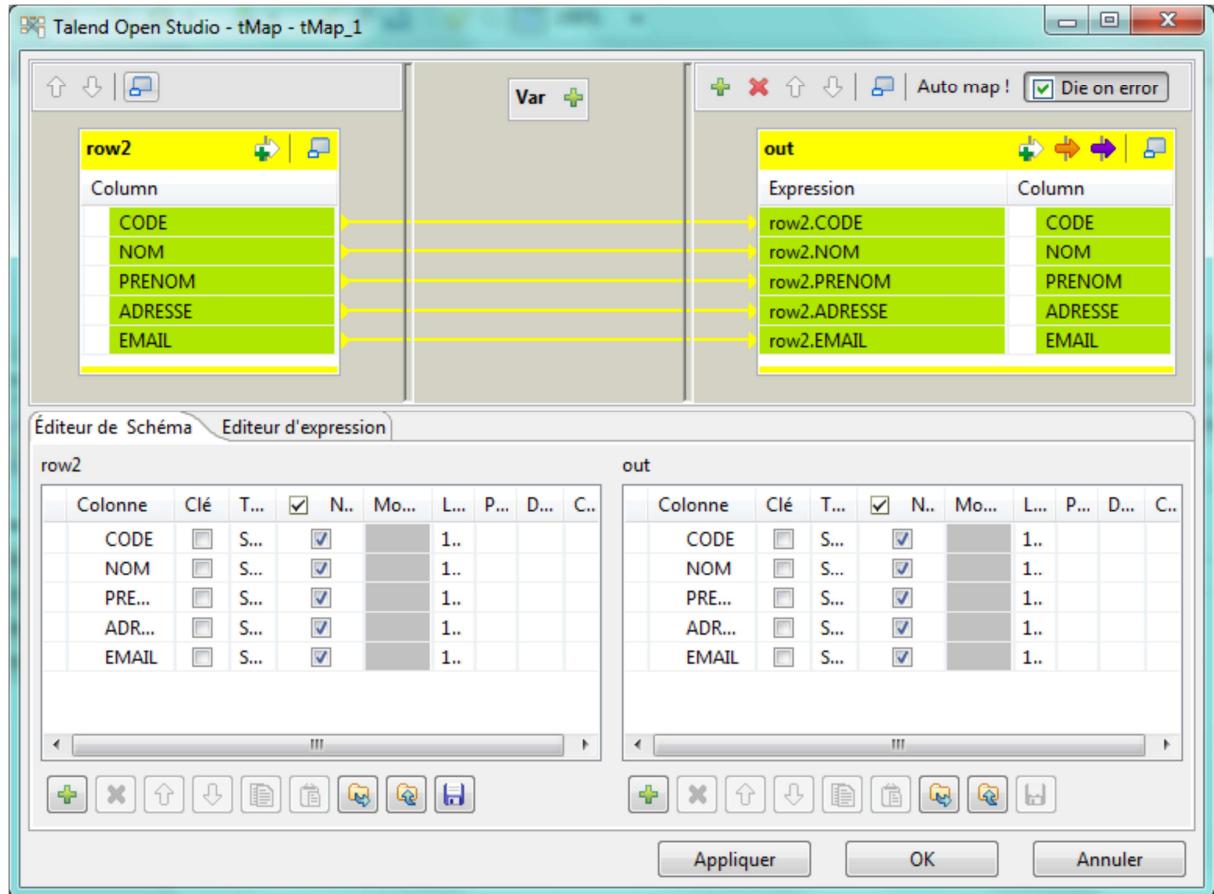
- Dans l'onglet « Palette » on choisit le composant « tMap », on le glisse et le dépose dans le « Designer » du job.



- On fait un clic droit sur le composant « tLogRow » puis on glisse vers le composant « tMap ».
- On fait un clic droit sur le composant « tMap » puis on glisse vers le composant « tAdvancedFileOutputXML », on donne un nom au lien qui sera créé. Puis on clique sur « Oui » pour récupérer les schémas de la cible.



- On fait un double-clic sur le composant « tMap » pour faire le mapping.
- Soit on fait des glisser/déplacer entre le schéma d'entrée et celui de sortie, soit si les noms des colonnes sont identiques on clique sur le bouton « autoMap ».



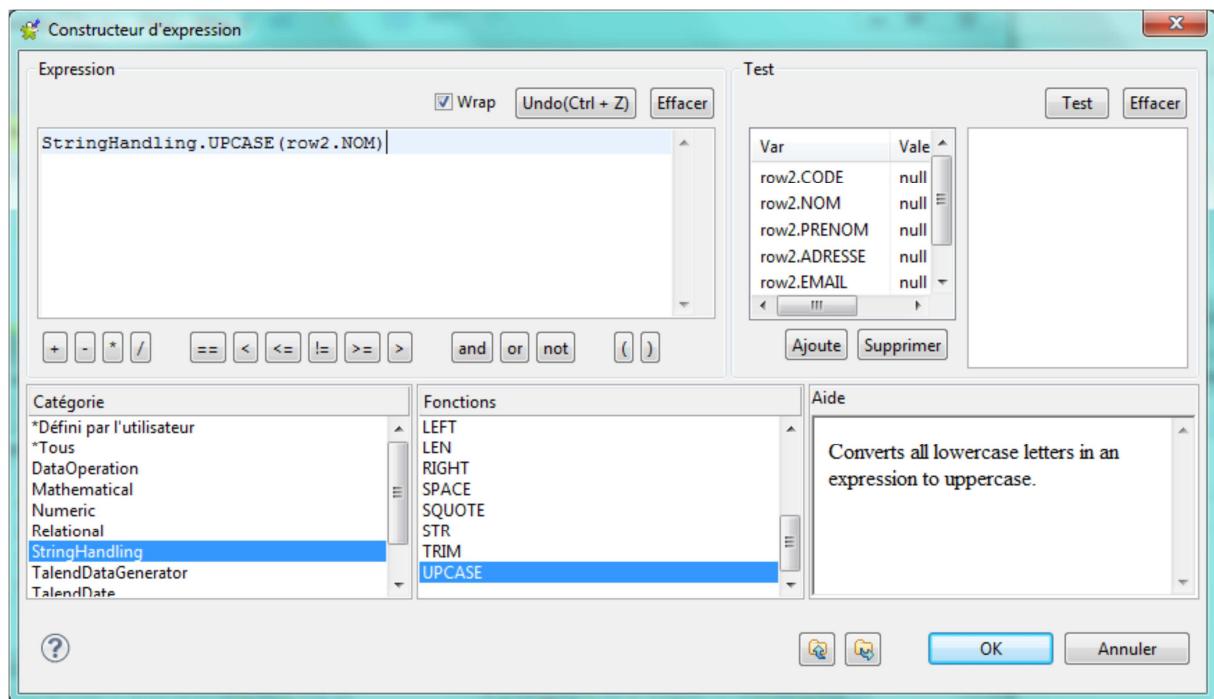
Le job est prêt à être exécuté, on appuie sur le bouton « F6 », et si on va sur le dossier de sortie on retrouvera nos fichiers XML.

```
<?xml version="1.0" encoding="ISO-8859-15" ?>
- <Client code="cli003">
  <Nom>zAkI</Nom>
  <Prenom>ANASS</Prenom>
- <Coordonnees>
  <Adresse>N18 BLOC1 RUE33 CASA</Adresse>
  <Email>ZAKI.A@HOTMAIL.COM</Email>
</Coordonnees>
</Client>
```

IV-D-5 - Transformer un champ grâce au « tMap »

On suppose qu'on veut transformer le champ « Nom » des clients en majuscules.

- On fait un double-clic sur le « tMap ».
- On sélectionne la colonne à transformer dans le schéma de sortie.
- Il y a un bouton « . » qui apparaît, on clique dessus.
- L'assistant de construction d'expression s'affiche.
- On choisit en bas à gauche la catégorie de fonction qui nous intéresse (dans notre cas on choisira « StringHandling »), puis dans la colonne de droite on choisit la fonction désirée, dans notre cas ce sera « UPCASE ».
- On fait un double-clic dessus.
- On modifie l'expression du haut comme dans l'aperçu en bas.
- Finalement on clique sur « OK ».



On peut maintenant relancer le job on remarque que les noms des clients sont maintenant tous en majuscules.