

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université de la Manouba

Institut supérieur des Arts Multimédias



rapport de fin de stage d'été

Migration vers Talend pour l'optimisation des Processus ETL.

Réalisé Par :

Khelifi Amen

Encadré Par :

Mahmoudi Oussama

année universitaire : 2023/2024

TABLE DES MATIÈRES

Introduction générale.....	3
Technologies utilisées.....	4
Installation des logiciels.....	4
1. Télécharger PostgreSQL.....	4
2. Télécharger Talend.....	5
3. Télécharger SoapUi.....	6
Flux de Travail.....	8
1. Extraction de Données.....	8
2. Transformation de données.....	16
3. chargement de données.....	19
4. dimensions créés.....	22
5. Résultats.....	23
6. surveillance et gestion des erreurs.....	31
Conclusion.....	34
1. Objectifs Atteints.....	34
2. Défis Rencontrés.....	34

Introduction générale

Dans le cadre de mon stage d'été au sein de l'entreprise **Neolledge**, j'ai eu l'opportunité de travailler sur un projet de migration de processus ETL (Extract, Transform, Load) vers la plateforme **Talend**. Ce projet s'inscrit dans une démarche d'optimisation des flux de données, permettant d'améliorer la qualité, la fiabilité et la rapidité de traitement des informations critiques pour l'entreprise.

Mon domaine d'études, une licence en **Big Data et analyses de données**, m'a permis d'acquérir des compétences solides en gestion de données volumineuses et en développement d'architectures de traitement de données. Ce stage m'a donné l'occasion d'appliquer ces connaissances dans un cadre professionnel, avec pour objectif principal la mise en place d'un mécanisme de synchronisation basé sur des **web services (SOAP UI)**. Ce mécanisme permet d'extraire les données depuis la base principale de l'entreprise pour les migrer vers une base de données simplifiée. Cette dernière est spécialement conçue pour servir de source pour la génération de **rapports BI (Business Intelligence)**, favorisant ainsi une prise de décision plus rapide et plus informée.

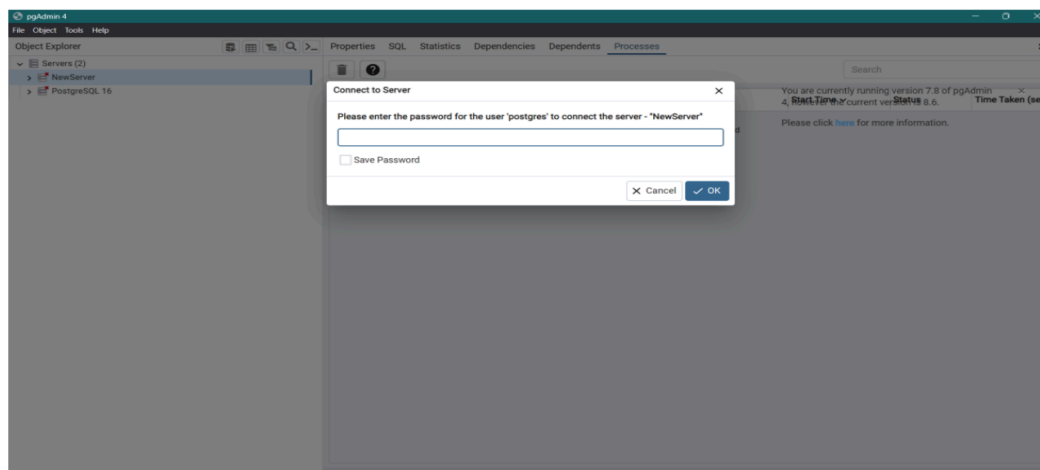
Technologies utilisées

1. **SOAP UI** (web services/ API Elise) : est utilisé pour tester et valider les services web (SOAP et REST).
2. **PostgreSQL/ Microsoft SQL server (SSMS)** : Stocke et gère les données extraites et transformées.
3. **TalendV8.1**: Intègre, transforme et charge les données depuis les services web dans PostgreSQL.

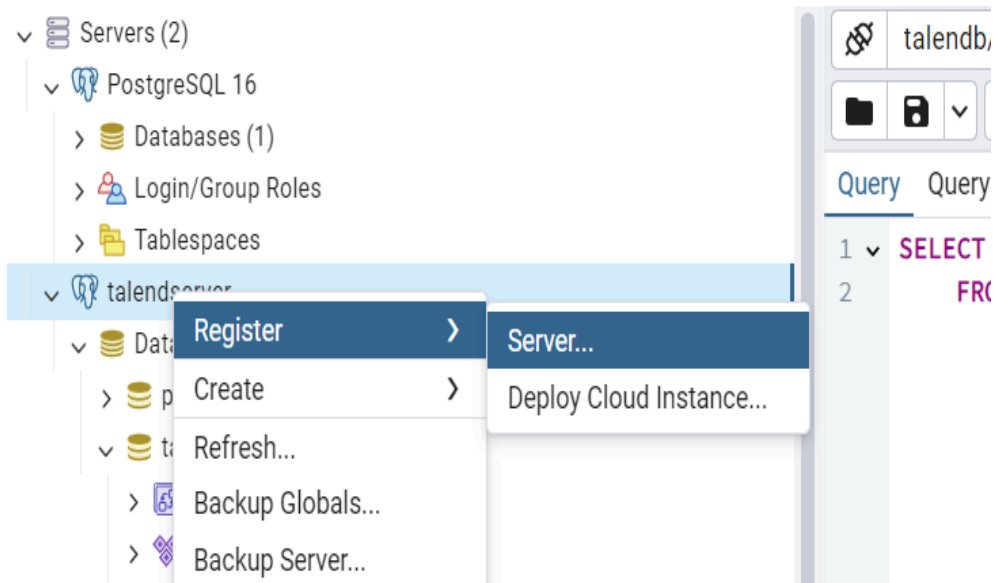
Installation des logiciels

1. Télécharger PostgreSQL

- Allez sur le [site officiel de PostgreSQL](#).
- Sélectionnez votre système d'exploitation (Windows, macOS, Linux).
- Cliquez sur le lien de téléchargement correspondant puis exécutez l'installateur.
- entrez le mot de passe déjà choisi dans les étapes de l'installation .



- créez un nouveau serveur .



Mon serveur, nommé **talendserver**, est configuré pour se connecter via le **port par défaut 5432**.

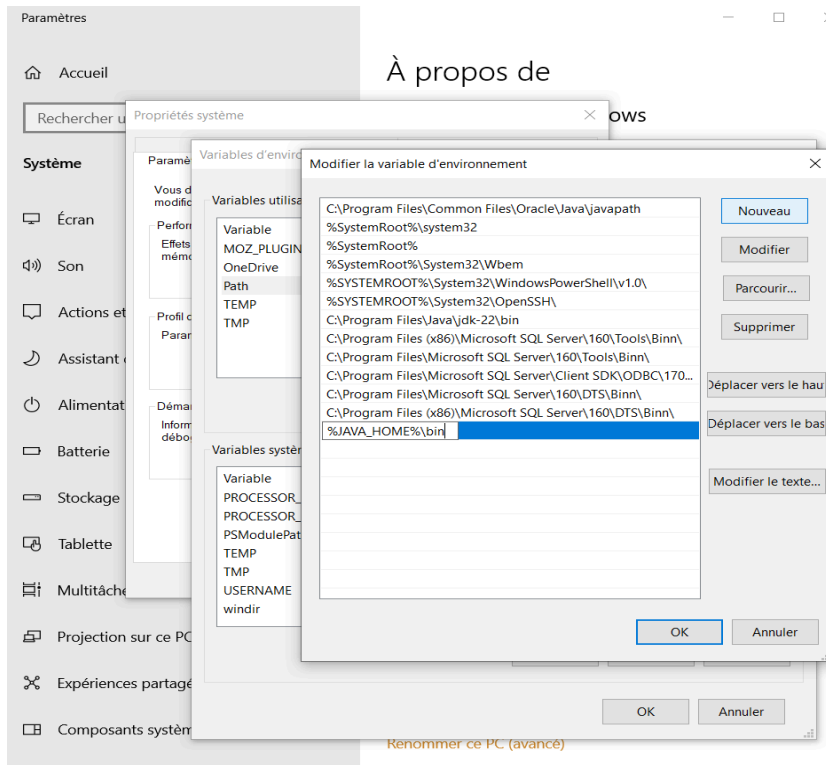
- Créez une base de données qui contiendra les dimensions générées ultérieurement par Talend. entrer un nom de votre BD par exemple “talendb” puis cliquez sur SAVE .

2. Télécharger Talend

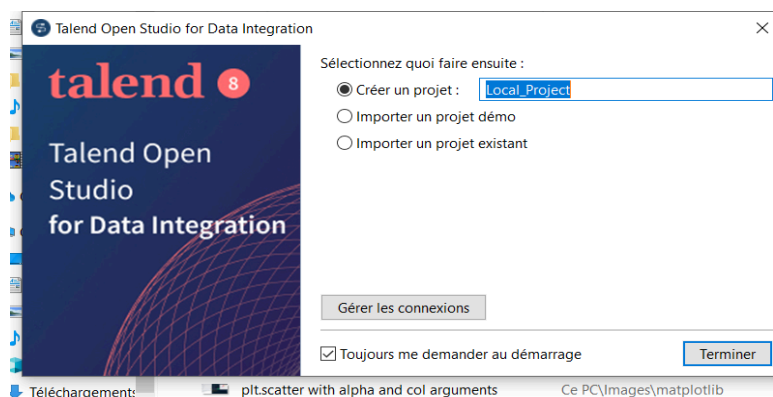
- **il est nécessaire de configurer l'environnement Java avant d'installer Talend.**
- Téléchargez et installez le JDK (Java Development Kit)

Par défaut, le JDK sera installé dans un dossier comme **C:\Program Files\Java\jdk-11** sur Windows.

- Configurer la variable d'environnement JAVA_HOME.

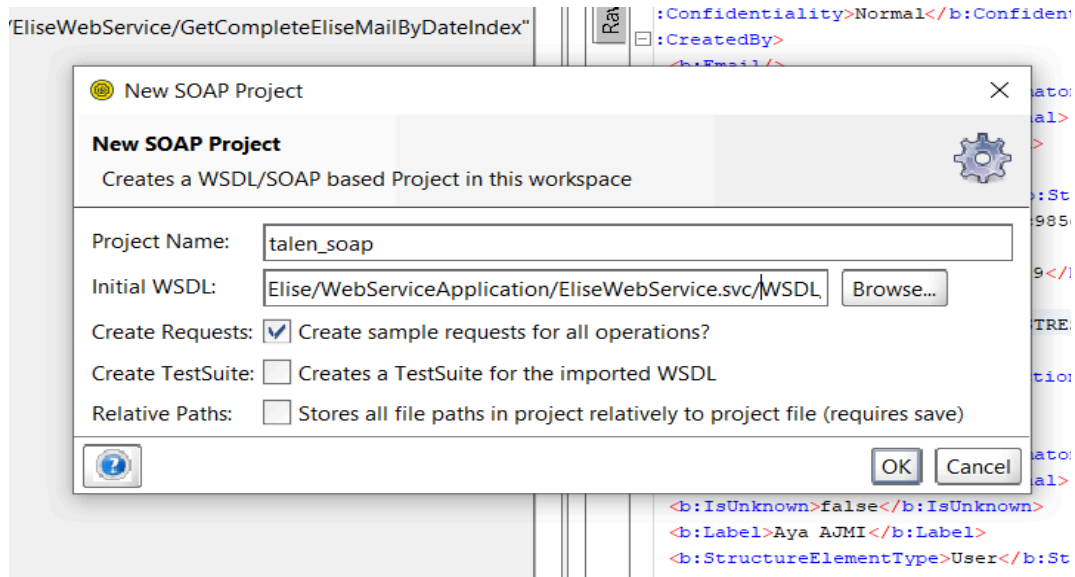


- Cliquez sur ce lien <https://gofile.io/d/ulmHb7> . (pour ce projet on a choisi la version 8.0.1).
- Extrayez le dossier zip, puis lancez l'installation en exécutant le fichier **.exe**.



3. Télécharger SoapUi

- Téléchargez sur ce lien <https://www.soapui.org/downloads/soapui/>
- cliquez sur SOAP et entrer ces valeurs :



WEBSERVICE NAME : GetCompleteEliseMailByDateIndex

soap config :

```
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/"
  xmlns:arc="Archimed.Elise.Application.Services"
  xmlns:arc1="http://schemas.datacontract.org/2004/07/Archimed.Elise.Application.Services"
  xmlns:arc2="http://schemas.datacontract.org/2004/07/Archimed.Elise.Models">
```

```
  <soapenv:Header/>
```

```
  <soapenv:Body>
```

```
    <arc:GetCompleteEliseMailByDateIndex>
```

```
      <!--Optional:-->
```

```
      <arc:session>
```

```
        <arc1:ApplicationID>EliseTest</arc1:ApplicationID>
```

```
      <arc1:ApplicationKey>8725d1a2e5aa4b9691aefe3b8235da62</arc1:ApplicationKey>
```

```
      <arc1:EliseVersionRequired>5.3</arc1:EliseVersionRequired>
```

```
      <arc1:Instance>GED</arc1:Instance>
```

```
      <arc1:Language>FR</arc1:Language>
```

```
      <arc1:Password>AdminGED</arc1:Password>
```

```

<arc1:UserLogin>AdminGED</arc1:UserLogin>

<arc2:MaxCount>1</arc2:MaxCount>

</arc:session>

<!--Optional:-->

<arc:completeEliseMailByDateIndexParameters>

<arc2:MaxCount>-1</arc2:MaxCount>

</arc:completeEliseMailByDateIndexParameters>

</arc:GetCompleteEliseMailByDateIndex>

</soapenv:Body>

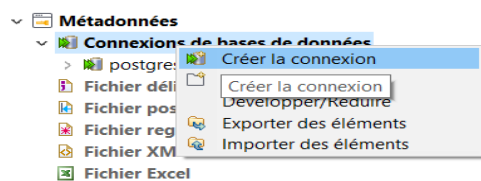
</soapenv:Envelope>

```

Flux de Travail

1. Extraction de Données

- Établissez la connexion avec notre base de données nommée **talendb** dans PostgreSQL.



Connexion base de données

Nouvelle connexion à la base de données dans le référentiel - Étape 1/2

⚠ Il est déconseillé de laisser vide le champ Objectif.

Nom	postgres
Objectif	
Description	
Créé par :	user@talend.com
Verrouillé par	
Version	0.1 M m
Statut	
Chemin d'accès	Sélectionner

< Back Next > Finish Cancel

Connexion base de données

Nouvelle connexion à la base de données dans le référentiel - Étape 2/2

① Vous devez cliquer sur le bouton Vérifier pour vérifier les paramètres de la base de données

Type de BdD PostgreSQL

Version de la base de données	v9 and later
Chaîne de caractères de connexion	jdbc:postgresql://localhost:5432/talendb?
Connexion	postgres
Mot de passe	••••••••
Serveur	localhost
Port	5432
Base de données	talendb
Schéma	
Paramètres supplémentaires	

Tester la connexion V

Exporter en tant que contexte Revenir au contexte précédent

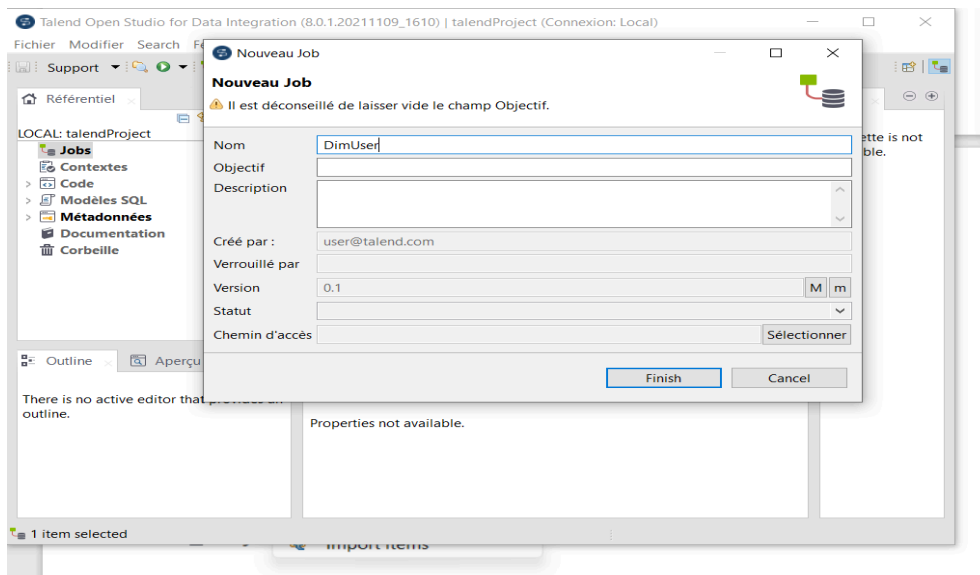
[Installer un pilote](#)

< Back Next > Finish Cancel

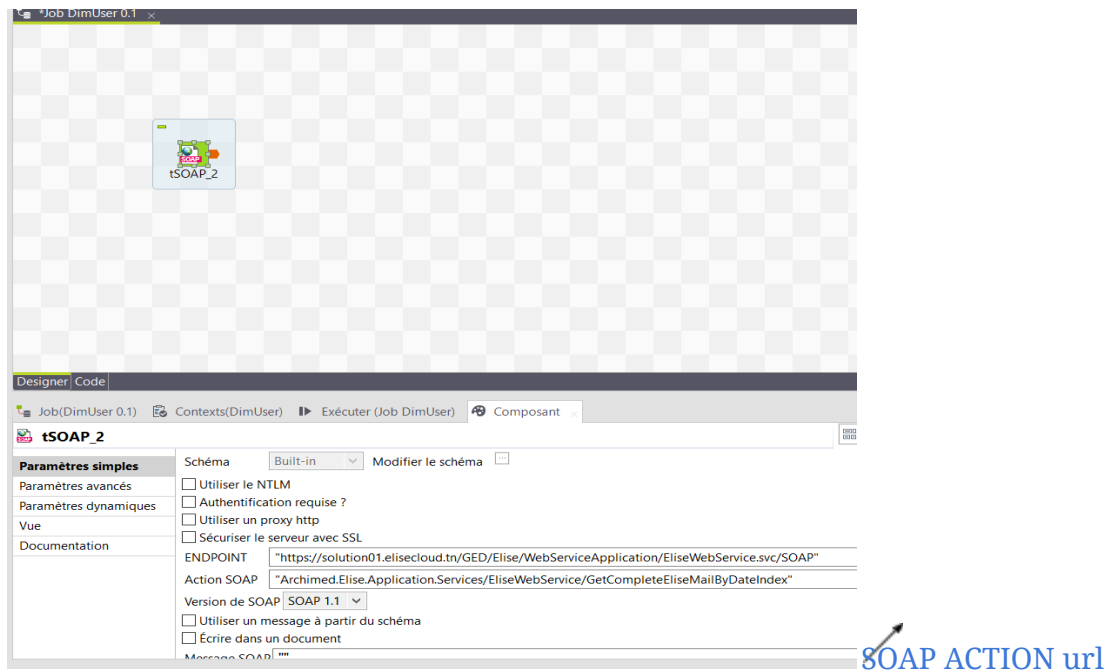
- **Configurer la connexion à PostgreSQL:**
 - **Sélectionner le type de base de données :** Choisissez **PostgreSQL** dans la liste des types de bases de données.
 - **Nom de la connexion :** Donnez un nom à la connexion (par exemple, postgres).
 - **Hôte :** Entrez l'adresse de votre serveur PostgreSQL (par exemple,

`localhost` ou l'adresse IP de votre serveur).

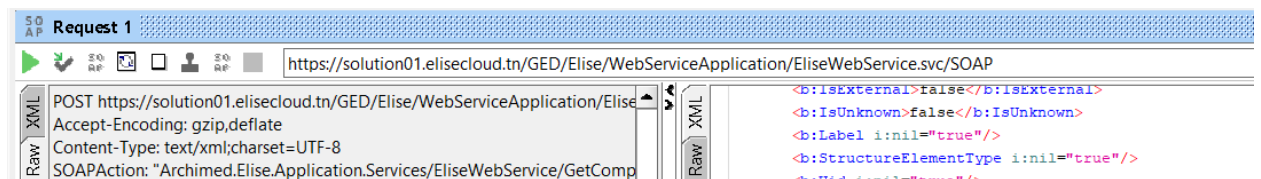
- **Port** : Entrez le port de connexion de PostgreSQL, généralement **5432**.
 - **Nom de la base de données** : Entrez le nom de votre base de données, ici **talendb**.
 - **Mot de passe** : Entrez le mot de passe correspondant à l'utilisateur.
 - Cliquez sur le bouton **Test** pour vérifier que Talend peut se connecter à la base de données. Assurez-vous que le test est réussi.
- **Créez un Job Talend** :
 - Donnez un nom à votre Job dans la fenêtre contextuelle (par exemple, "**Dim_user**"), ajoutez une description si souhaité, et cliquez sur "**Finish**".



- Utilisez le composant **tSOAP** pour extraire les données des services web.
 - Pour se connecter avec SOAP UI, il est nécessaire de saisir le **bon ENDPOINT**, la **SOAP ACTION** appropriée et le **Message SOAP**, que vous trouverez dans SOAP UI.



et finalement le message soap qu'on le trouve en cliquant sur xml dans SOAPUI.



```

<?xml version='1.0' encoding='UTF-8'>
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/" xmlns:a="http://schemas.xmlsoap.org/soap/envelope/">
  <soapenv:Header/>
  <soapenv:Body>
    <arc:GetCompleteEliseMailByDateIndex>
      <!--Optional:-->
      <arc:session>
        <arc1:ApplicationID>EliseTest</arc1:ApplicationID>

        <arc1:ApplicationKey>8725d1a2e5aa4b9691aefe3b8235da62</arc1:ApplicationKey>

        <arc1:EliseVersionRequired>5.3</arc1:EliseVersionRequired>

        <arc1:Instance>GED</arc1:Instance>

        <arc1:Language>FR</arc1:Language>

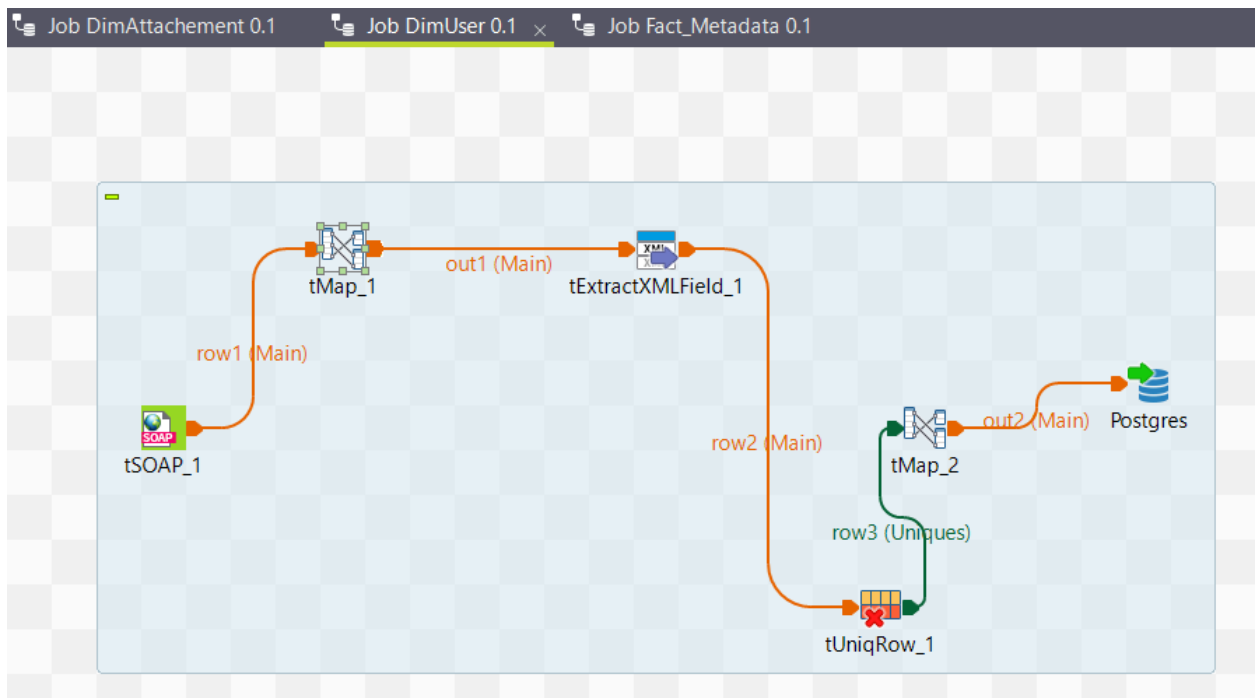
        <arc1:Password>AdminGED</arc1:Password>

        <arc1:UserLogin>AdminGED</arc1:UserLogin>

      </arc:session>
      <!--Optional:-->
      <arc:completeEliseMailByDateIndexParameters>
        <arc2:MaxCount>1</arc2:MaxCount>
      </arc:completeEliseMailByDateIndexParameters>
    </arc:GetCompleteEliseMailByDateIndex>
  </soapenv:Body>
</soapenv:Envelope>

```

- Ajoutez les composants nécessaires pour votre Job et configurez les.



Voici le schéma complet de notre Job intitulé **DimUser**. Chaque composant sera clarifié et détaillé dans les sous-titres suivants.

Maintenant que nous avons configuré le composant **tSOAP**, la réponse du service SOAP

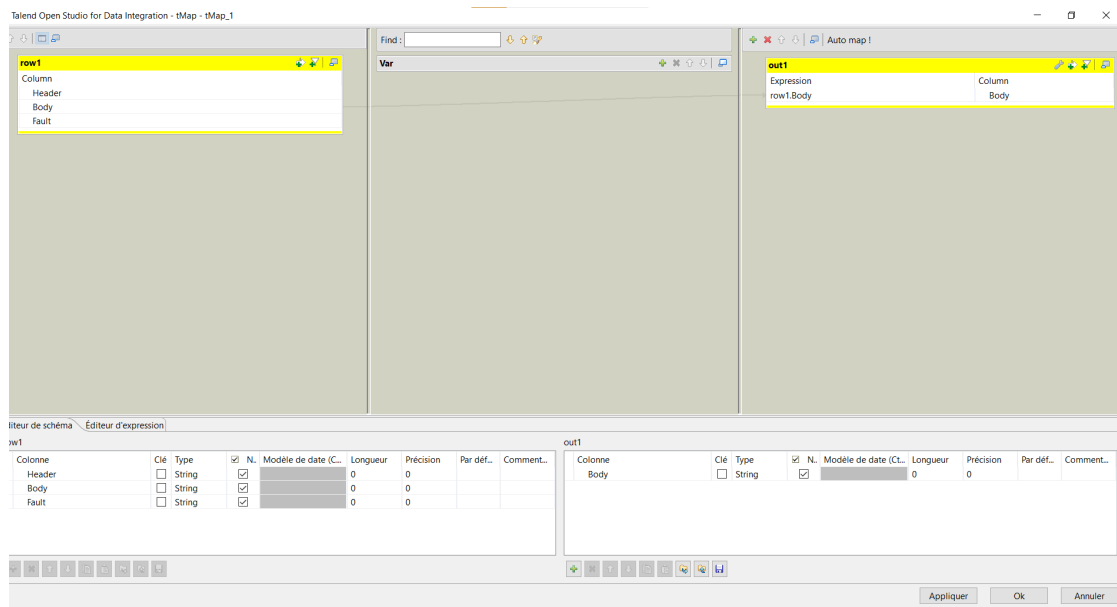
est divisée en trois parties :

- **Head, Body et Fault.**

Pour extraire les données, seule la partie **Body** nous intéresse. Nous allons donc utiliser **tMap** pour extraire le contenu de la **Body**.

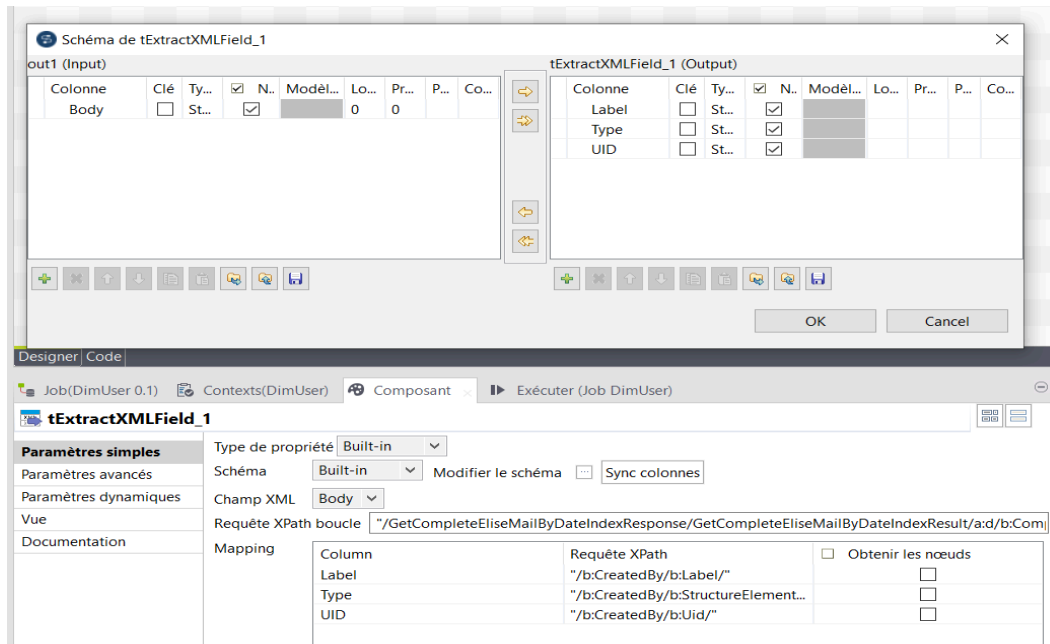
Voici les étapes :

- Cliquez sur le panneau, tapez **tMap**, puis sélectionnez-le.
- Reliez ensuite le composant **tSOAP** au composant **tMap** afin de traiter la réponse XML.
- cliquez sur l'élément tmap , cette fenêtre s'affiche:



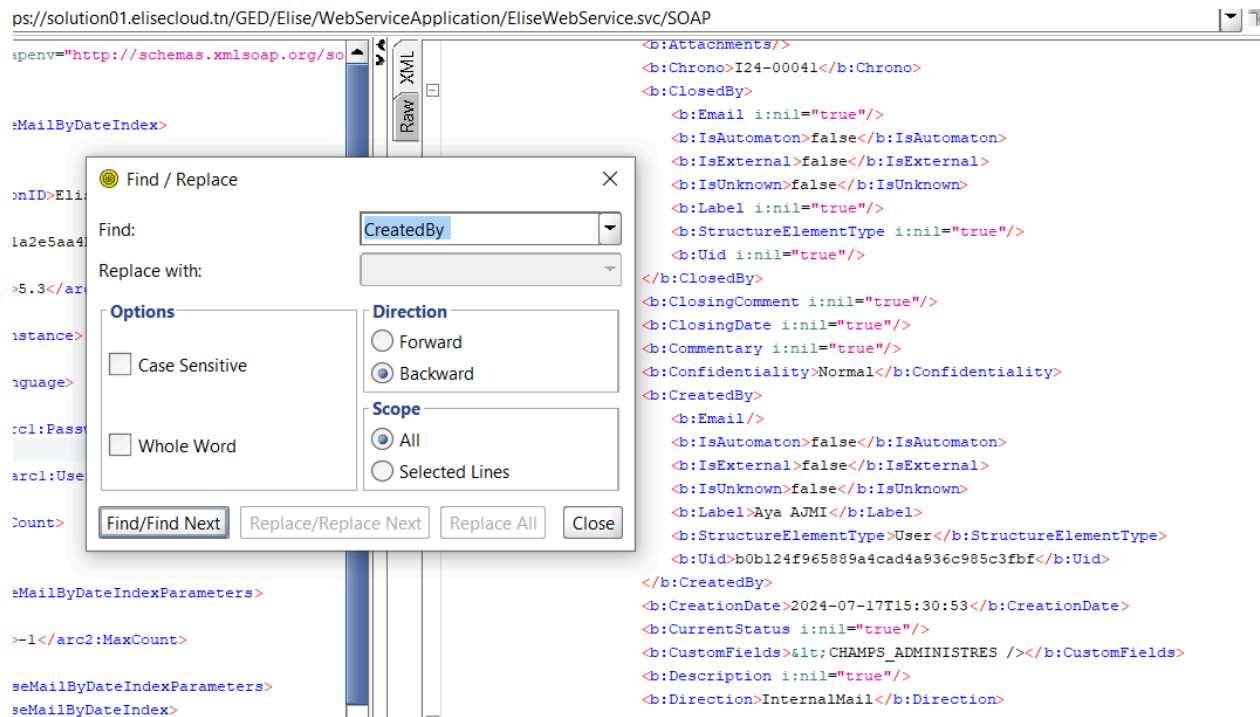
Dans cette étape, vous allez **glisser-déposer** le contenu de la **Body** de la section principale vers la sortie, que vous allez créer.

- **tExtractXMLField** pour traiter le contenu XML de la réponse SOAP :



- il est nécessaire de configurer l'élément tExtractXMLField :
 - au premier lieu , on définira la requête XPath boucle qui est `"/GetCompleteEliseMailByDateIndexResponse/GetCompleteEliseMailByDateIndexResult/a:d/b:CompleteEliseMailCollection/b:CompleteEliseMail"`
 → Cela indique à Talend que chaque **CompleteEliseMail** est un élément répétitif et doit être traité séparément. Les données contenues dans chaque **CompleteEliseMail** seront ensuite extraites en fonction des colonnes définies dans le schéma.
- ensuite , Il faut entrer manuellement les noms des colonnes de votre sortie ainsi que les données associées à ces noms, qui sont les **requêtes XPath**. Pour commencer à entrer le nom des colonnes -Cliquez sur **Edit schema** (Modifier le schéma).
- Pour sélectionner les colonnes, nous utilisons une requête SOAP dans SoapUI. En appuyant sur **Ctrl+F**, nous recherchons les informations dont nous avons besoin. Dans notre exemple concernant la dimension **User**, nous allons rechercher les informations relatives à l'utilisateur dans les balises

correspondantes de SoapUI. Les données requises se trouvent sous la balise **CreatedBy**.



on va extraire les informations suivantes : label , structureElementType,Uid sous CreatedBy .

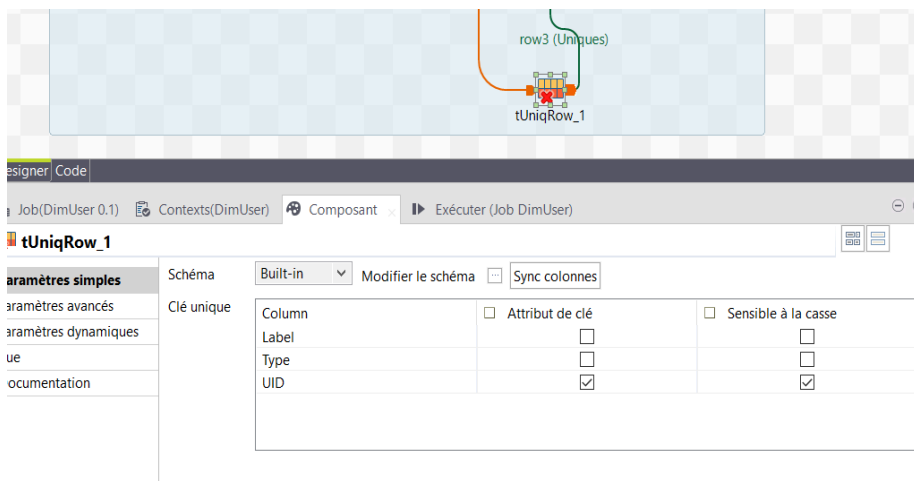
Pour extraire des données comme le **Label** et le **Uid**, vous rempliriez les champs de requête XPath comme suit :

- Pour le Label : `"/b:CreatedBy/b:Label/"`
- Pour le Uid : `"/b:CreatedBy/b:Uid/"`

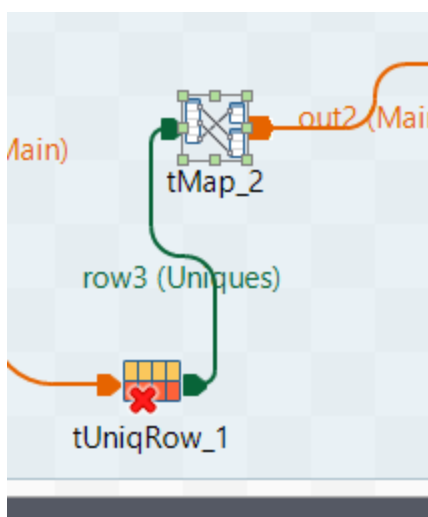
Ces requêtes vous permettent d'extraire les valeurs "Aya AJMI" pour le **Label** et "b0b124f965889a4cad4a936c985c3fbf" pour le **Uid**.

2. Transformation de données

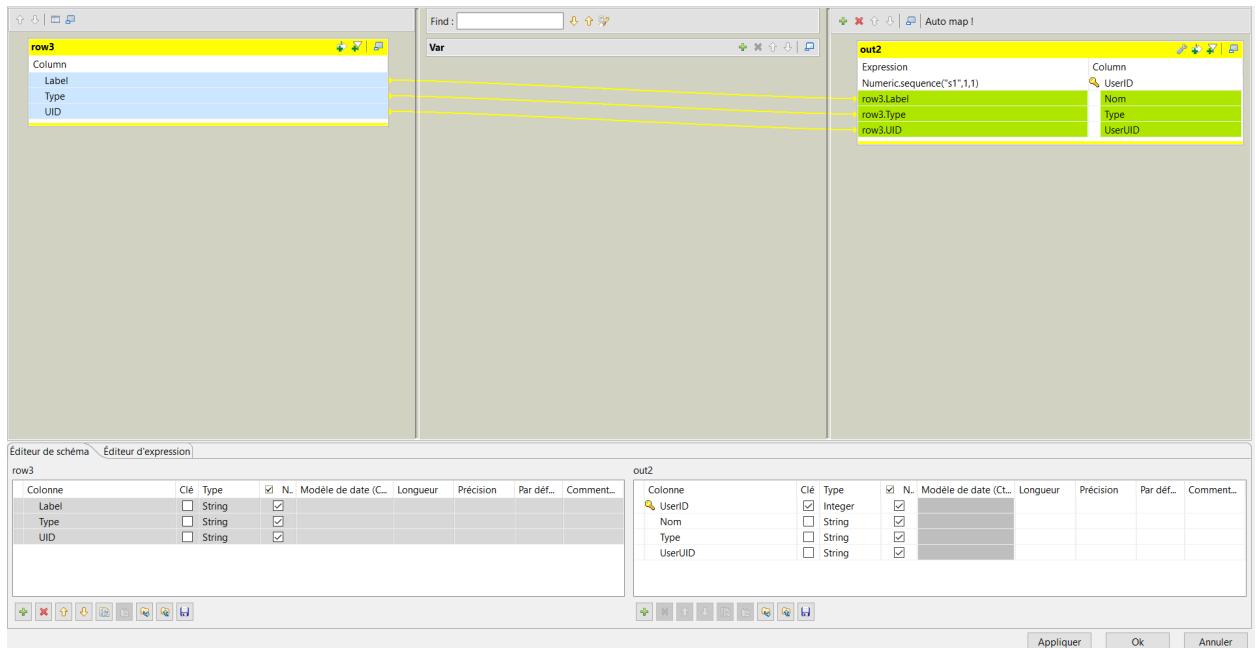
- **Ajoutez le composant tUniqRow** : dont le rôle est de filtrer les lignes dupliquées en conservant uniquement les lignes uniques selon les colonnes que vous avez définies. Dans notre cas, la clé utilisée pour **tUniqRow** est le **Uid**.



- **Ajoutez le composant tMap** : permet d'assurer une intégration de données optimisée et personnalisée entre nos différentes sources et cibles dans le cadre de la migration ETL vers Talend.



connecter tmap avec tUniqRow et passons à la configuration de l'élément tmap_2.

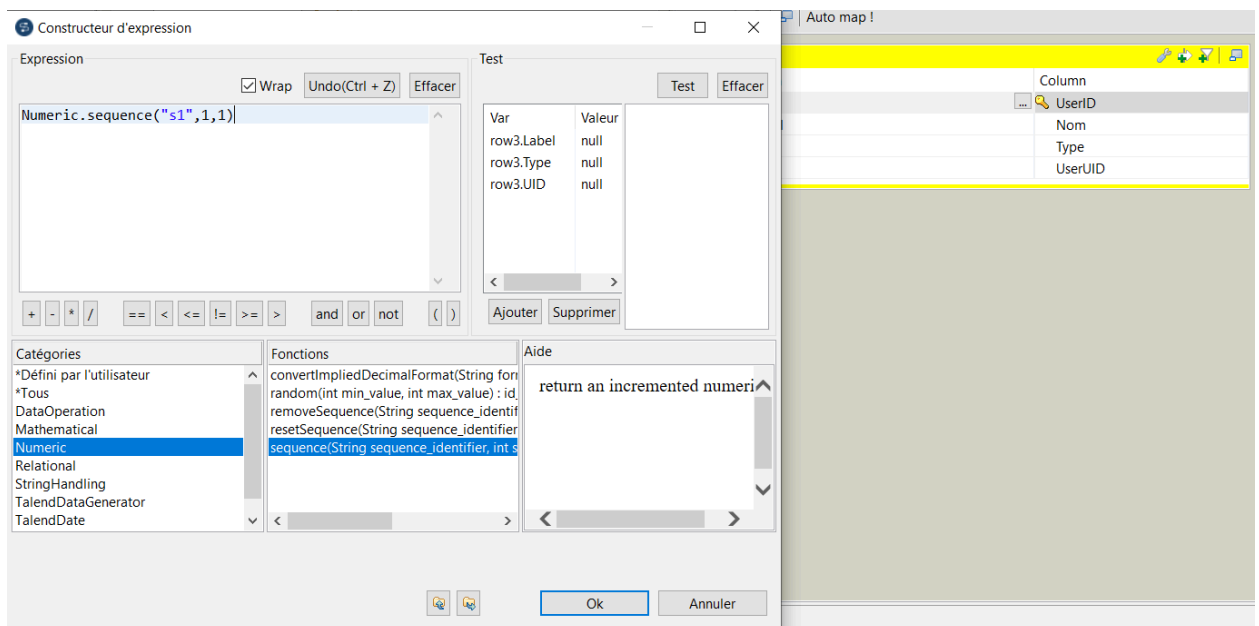


- Pour mapper les colonnes, glissez-déposez les colonnes d'entrée (à gauche) vers les colonnes de sortie (à droite).
- il faut ajouter un clé substitut dont son rôle est de fournir un identifiant unique et stable pour chaque enregistrement (par exemple, **UserID**). Assurez-vous de spécifier son type comme **Integer**.
- Cochez la case **Primary Key** (Clé primaire) pour cette colonne.
- Dans le cas où la colonne **longueur** et **précision** admet une valeur, il est préférable de les supprimer pour éviter les erreurs lors de la compilation.

Colonne	Clé	Type	<input checked="" type="checkbox"/> N..	Modèle de date (Ct...	Longueur	Précision	Par déf...	Comi
UserID	<input checked="" type="checkbox"/>	Integer	<input checked="" type="checkbox"/>					
Nom	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>					
Type	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>					
UserID	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>					

out2	
Expression	Column
	UserID
row3.Label	Nom
row3.Type	Type
row3.UID	UserID

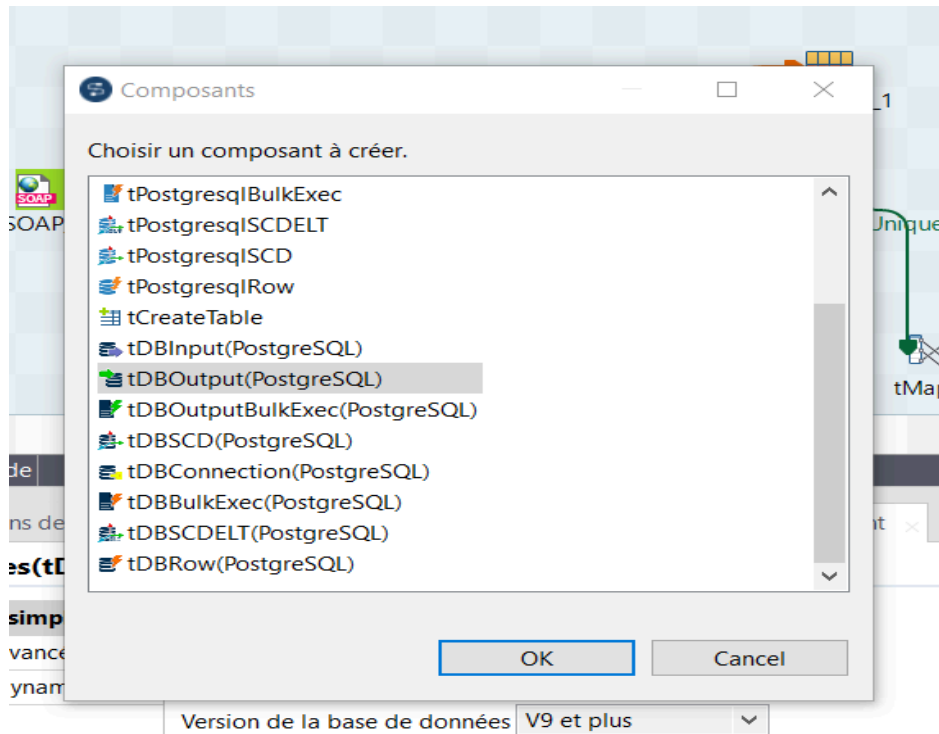
1. Dans le champ **Expression** de la colonne **UserID**, cliquez sur les trois points ... à droite.
2. Dans la fenêtre **Expression Builder** (Constructeur d'expressions), sélectionnez **Numeric** (Numérique).
3. Glissez-déposez **sequence** dans la boîte vide en haut à gauche pour définir la séquence numérique.
4. Cliquez sur **OK** pour valider et fermer l'éditeur d'expressions.

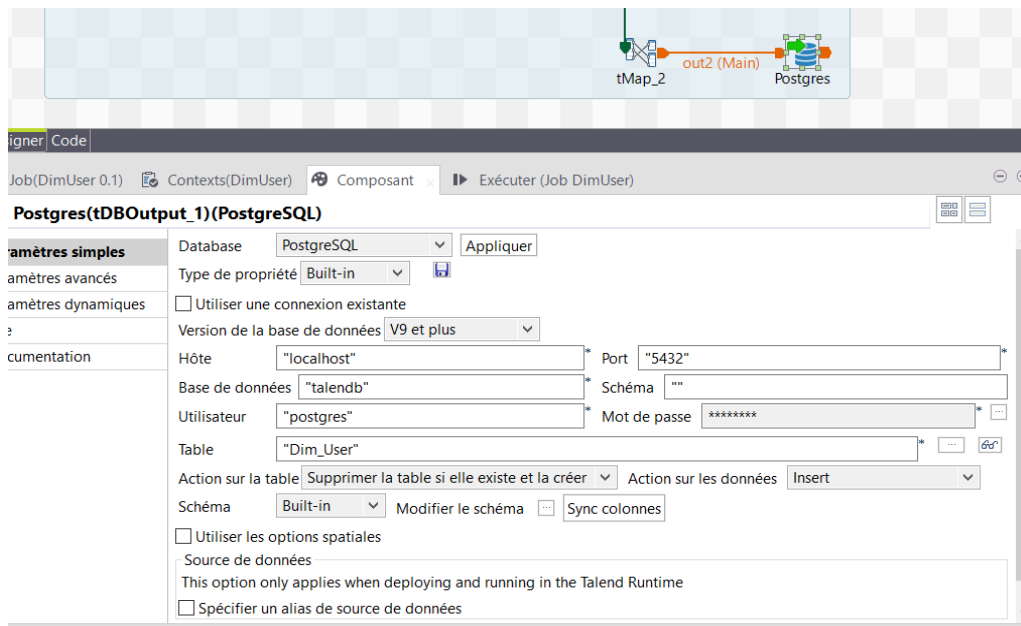


- Enfin, cliquez sur **OK** pour fermer l'éditeur **tMap**. Cette configuration définit la colonne **ID** comme clé primaire avec une séquence numérique dans votre transformation.

3. chargement de données

- Utilisez le composant **tDBOutput(PostgreSQL)** pour charger les données transformées dans votre base de données PostgreSQL.





- Entrez le nom de la table que vous souhaitez créer dans votre base de données PostgreSQL. Dans notre exemple, il s'agit de Dim_User.
- **Pour vérifier la création de l'output de notre job dans PostgreSQL :**
 1. Ouvrez votre base de données PostgreSQL.
 2. Sous le schéma public, trouvez les dimensions créées.
 3. Faites un clic droit sur la table correspondante.
 4. Sélectionnez Script puis Select Script.

Cela vous permettra de visualiser les données et vérifier que la table a été correctement créée.

pgAdmin 4

File Object Tools Edit View Window Help

Object Explorer

- Languages
- Publications
- Schemas (1)
 - public
 - Aggregates
 - Collations
 - Domains
 - FTS Configurations
 - FTS Dictionaries
 - FTS Parsers
 - FTS Templates
 - Foreign Tables
 - Functions
 - Materialized Views
 - Operators
 - Procedures
 - Sequences
 - Tables (7)
 - Dim_Documents
 - Dim_Support
 - Dim_Type
 - Dim_User**
 - Columns
 - Constraints
 - Indexes

talendb/postgres@talendserver

Count Rows
Create
Delete
Delete (Cascade)
Refresh...
Restore...
Backup...
Import/Export Data...
Reset Statistics
ERD For Table
Maintenance...
Scripts

- CREATE Script
- SELECT Script**
- INSERT Script
- UPDATE Script
- DELETE Script

Truncate
View/Edit Data
Search Objects...
PSQL Tool
Query Tool
Properties...

SQL

	UserID	UserUID	Nom	UserType
1	BUILTIN-SERVICE	CrystalFramework Batch S		
2	e7b7a90e8560874eb71951d6eff66584	AdminGED		
3	b0b124f965889a4cad4a936c985c3fbf	Aya AJMI		
4	LEXICON_00000039	Délice Holding		

Query Query History

```
1 SELECT "UserID", "UserUID", "Nom", "UserType"
2 FROM public."Dim_User";
```

Data Output Messages Notifications

SQL

	UserID [PK] integer	UserUID character varying	Nom character varying	UserType character varying
1	1	BUILTIN-SERVICE	CrystalFramework Batch Service	User
2	2	e7b7a90e8560874eb71951d6eff66584	AdminGED	User
3	3	b0b124f965889a4cad4a936c985c3fbf	Aya AJMI	User
4	4	LEXICON_00000039	Délice Holding	Service
5	5	27da2ead1443024589aa9cb1355f73...	OpNumTun1	User
6	6	ecf32550bdf3e64b8ff510be277f147b	مكتب الضبط	User
7	7	LEXICON_00000000	GED	Service
8	8	320bd15d7c6b0743b8c3508e839822...	Delice	User
9	9	6e0b57e1c2605842aa4abc86215edc6f	AgentBOTun1	User
10	10	ddb72fb1b46ca49a72124b03eb0a7...	Abir amri	User
11	11	07a9a7de8a6de749ae4df46a4ca3ef6f	Nadia	User
12	12	79b116f9e9a7d94484bfd5beed9ec14	Admin1	User
13	13	00420fdc944cb7479d18e7c2ca09ba6c	Quentin	User
14	14	56006bc81f503d408d23581079f88543	Agent BO	User
15	15	7cbbc4c8faa6844780dfce0ea4cd61a	Oussama	User
16	16	757ed7340b83ef4690378c931a03cf8a	test1	User
17	17	b9e0060de8902b41aba88ba1e1568e...	Agent BO	User
18	18	e7ef91a6e4db134bb43df7e6e2287805	test007	User
19	19	e96d43b740c21b40ad6690f815fd0f91	Test2	User

Total rows: 21 of 21 Query complete 00:00:00 142 In 1 Col 1

De la même manière pour les autres dimensions créées, on répète le même processus pour **DimAttachment**, **DimSupport**, **DimSens**, et **DimType**

4. dimensions créés

Nom Dimension	tExtractXMLField	tmap_2																											
DimAttachment	<div>Mapping</div> <table><thead><tr><th>Column</th><th>Requête XPath</th><th><input type="checkbox"/> Obtenir les nœuds</th></tr></thead><tbody><tr><td>AttachmentId</td><td>"/b:Attachments/b:Attachment/...</td><td><input type="checkbox"/></td></tr><tr><td>NomDoc</td><td>"/b:Attachments/b:Attachment/...</td><td><input type="checkbox"/></td></tr><tr><td>TypeDoc</td><td>"/b:Attachments/b:Attachment/...</td><td><input type="checkbox"/></td></tr><tr><td>TailleDoc</td><td>"/b:Attachments/b:Attachment/...</td><td><input type="checkbox"/></td></tr></tbody></table>	Column	Requête XPath	<input type="checkbox"/> Obtenir les nœuds	AttachmentId	"/b:Attachments/b:Attachment/...	<input type="checkbox"/>	NomDoc	"/b:Attachments/b:Attachment/...	<input type="checkbox"/>	TypeDoc	"/b:Attachments/b:Attachment/...	<input type="checkbox"/>	TailleDoc	"/b:Attachments/b:Attachment/...	<input type="checkbox"/>	<div>out2</div> <table><thead><tr><th>Expression</th><th>Column</th></tr></thead><tbody><tr><td>Numeric.sequence("s1",1,1)</td><td> AttachID</td></tr><tr><td>row3.AttachmentId</td><td>IdDoc</td></tr><tr><td>row3.NomDoc</td><td>Nomdoc</td></tr><tr><td>row3.TypeDoc</td><td>TypeDoc</td></tr><tr><td>row3.TailleDoc</td><td>TailleDoc</td></tr></tbody></table>	Expression	Column	Numeric.sequence("s1",1,1)	AttachID	row3.AttachmentId	IdDoc	row3.NomDoc	Nomdoc	row3.TypeDoc	TypeDoc	row3.TailleDoc	TailleDoc
Column	Requête XPath	<input type="checkbox"/> Obtenir les nœuds																											
AttachmentId	"/b:Attachments/b:Attachment/...	<input type="checkbox"/>																											
NomDoc	"/b:Attachments/b:Attachment/...	<input type="checkbox"/>																											
TypeDoc	"/b:Attachments/b:Attachment/...	<input type="checkbox"/>																											
TailleDoc	"/b:Attachments/b:Attachment/...	<input type="checkbox"/>																											
Expression	Column																												
Numeric.sequence("s1",1,1)	AttachID																												
row3.AttachmentId	IdDoc																												
row3.NomDoc	Nomdoc																												
row3.TypeDoc	TypeDoc																												
row3.TailleDoc	TailleDoc																												
DimSupport	<table><thead><tr><th>Column</th><th>Requête XPath</th></tr></thead><tbody><tr><td>SupportLabel</td><td>"/b:Support/b:Title/"</td></tr><tr><td>Support</td><td>"/b:Support/b:ExternalId/"</td></tr></tbody></table>	Column	Requête XPath	SupportLabel	"/b:Support/b:Title/"	Support	"/b:Support/b:ExternalId/"	<div>out2</div> <table><thead><tr><th>Expression</th><th>Column</th></tr></thead><tbody><tr><td>Numeric.sequence("s1",1,1)</td><td> SupportID</td></tr><tr><td>row3.SupportLabel</td><td>SupportLabel</td></tr><tr><td>row3.Support</td><td>Support</td></tr></tbody></table>	Expression	Column	Numeric.sequence("s1",1,1)	SupportID	row3.SupportLabel	SupportLabel	row3.Support	Support													
Column	Requête XPath																												
SupportLabel	"/b:Support/b:Title/"																												
Support	"/b:Support/b:ExternalId/"																												
Expression	Column																												
Numeric.sequence("s1",1,1)	SupportID																												
row3.SupportLabel	SupportLabel																												
row3.Support	Support																												
DimSens	<table><thead><tr><th>Column</th><th>Requête XPath</th></tr></thead><tbody><tr><td>Direction</td><td>"/b:Direction/"</td></tr></tbody></table>	Column	Requête XPath	Direction	"/b:Direction/"	<div>out2</div> <table><thead><tr><th>Expression</th><th>Column</th></tr></thead><tbody><tr><td>Numeric.sequence("s1",1,1)</td><td> SensID</td></tr><tr><td>row3.Direction</td><td>Sens</td></tr></tbody></table>	Expression	Column	Numeric.sequence("s1",1,1)	SensID	row3.Direction	Sens																	
Column	Requête XPath																												
Direction	"/b:Direction/"																												
Expression	Column																												
Numeric.sequence("s1",1,1)	SensID																												
row3.Direction	Sens																												
DimType	<table><thead><tr><th>Column</th><th>Requête XPath</th></tr></thead><tbody><tr><td>TypeLabel</td><td>"/b:MailType/b:Title/"</td></tr><tr><td>Type</td><td>"/b:MailType/b:ExternalId/"</td></tr></tbody></table>	Column	Requête XPath	TypeLabel	"/b:MailType/b:Title/"	Type	"/b:MailType/b:ExternalId/"	<div>out2</div> <table><thead><tr><th>Expression</th><th>Column</th></tr></thead><tbody><tr><td>Numeric.sequence("s1",1,1)</td><td> TypelD</td></tr><tr><td>row3.TypeLabel</td><td>TypeLabel</td></tr><tr><td>row3.Type</td><td>Type</td></tr></tbody></table>	Expression	Column	Numeric.sequence("s1",1,1)	TypelD	row3.TypeLabel	TypeLabel	row3.Type	Type													
Column	Requête XPath																												
TypeLabel	"/b:MailType/b:Title/"																												
Type	"/b:MailType/b:ExternalId/"																												
Expression	Column																												
Numeric.sequence("s1",1,1)	TypelD																												
row3.TypeLabel	TypeLabel																												
row3.Type	Type																												

5. Résultats

Le but de la création des dimensions DimAttachment, DimSupport, DimSens, DimType, et DimUser est de structurer et d'organiser les données extraites de différentes sources pour faciliter l'analyse et les rapports.

La dimension document contient les informations extraites de ces autres dimensions, consolidant ainsi les données pour une vue complète et cohérente des informations pertinentes.

Column	Requête XPath	<input type="checkbox"/> Obtenir les nœuds
Titre	"/b:Subject/"	<input type="checkbox"/>
ExternalId	"/b:MailId/"	<input type="checkbox"/>
Sens	"/b:Direction/"	<input type="checkbox"/>
Support	"/b:Support/b:ExternalId/"	<input type="checkbox"/>
Type	"/b:MailType/b:ExternalId/"	<input type="checkbox"/>
CreatedBy	"/b:CreatedBy/b:Uid/"	<input type="checkbox"/>
CreatedAt	"/b:CreationDate/"	<input type="checkbox"/>

on va ajouter le composant tDBInput dans le but de se connecter à une base de données (dans ce cas, PostgreSQL) et d'extraire des données d'une table spécifique

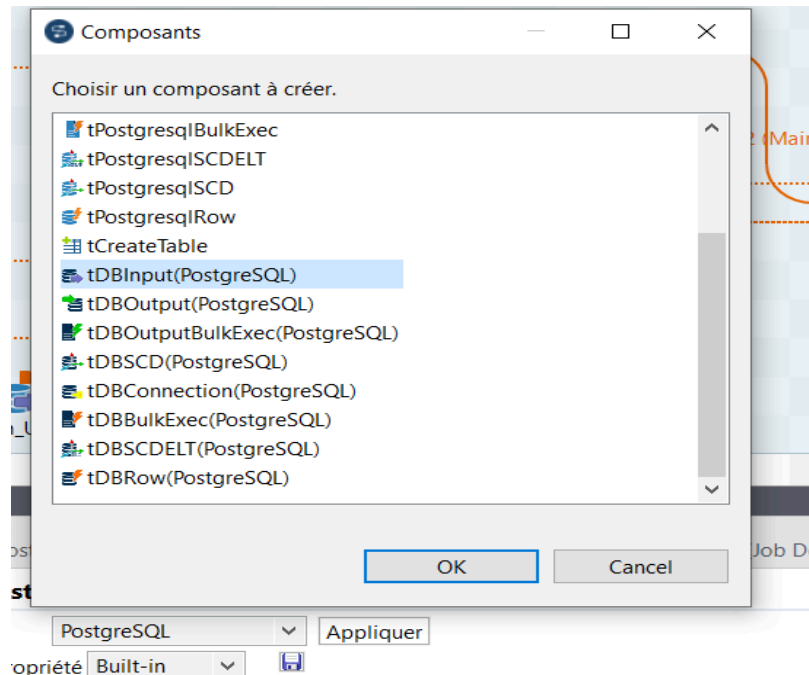
Dans notre contexte, tDBInput est utilisé pour récupérer les données de la table DimSupport afin que la dimension DimDocument puisse les exploiter.

Passons maintenant à l'ajout du composant tDBInput PostgreSQL qui chargera les données pour le job DimSupport, permettant ainsi à DimDocument d'extraire les données depuis la table DimSupport.

1. Ajout du composant tDBInput :

- Glissez-déposez l'icône PostgreSQL située sous l'onglet Métadonnées > Connexion de base de données.

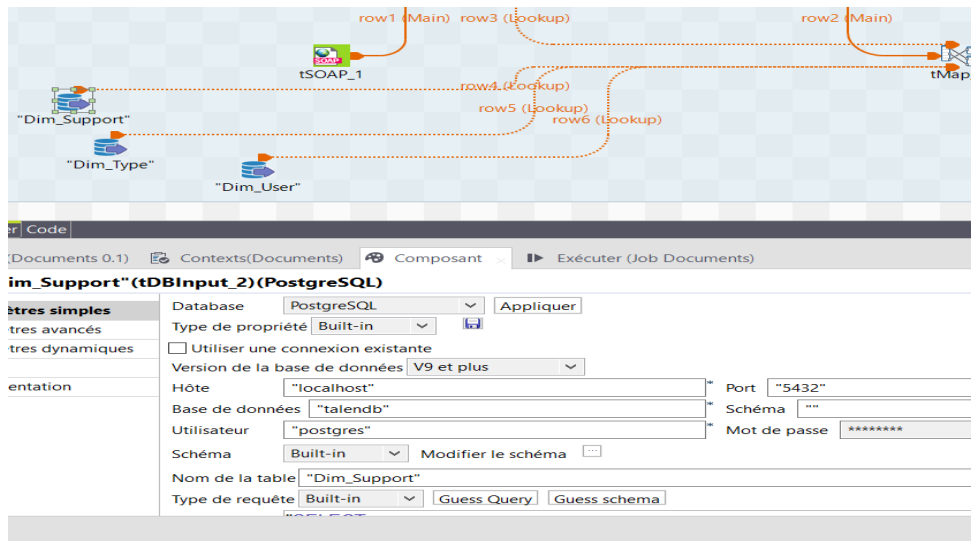
- Une fenêtre apparaîtra. Choisissez le composant tDBInput pour commencer sa configuration



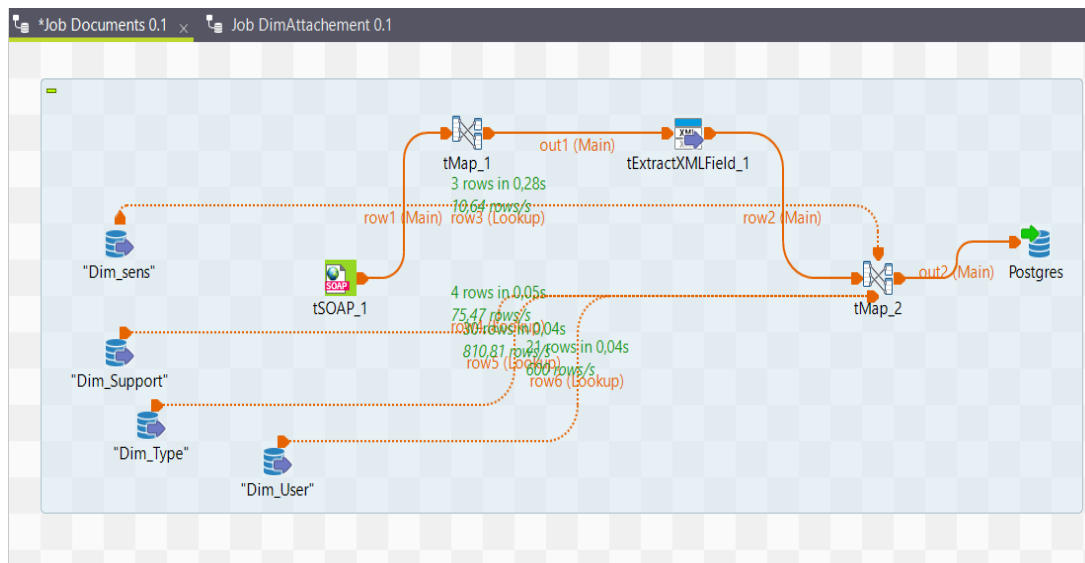
- une connexion existante

2. Configuration de la connexion :

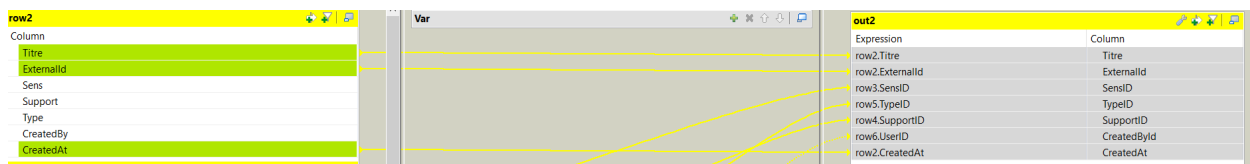
- Dans la fenêtre de configuration du composant tDBInput, renseignez les détails de connexion à la base de données PostgreSQL.
- Nom de la table : Tapez le nom de la table créée à partir du job précédemment effectué, qui est DimSupport.

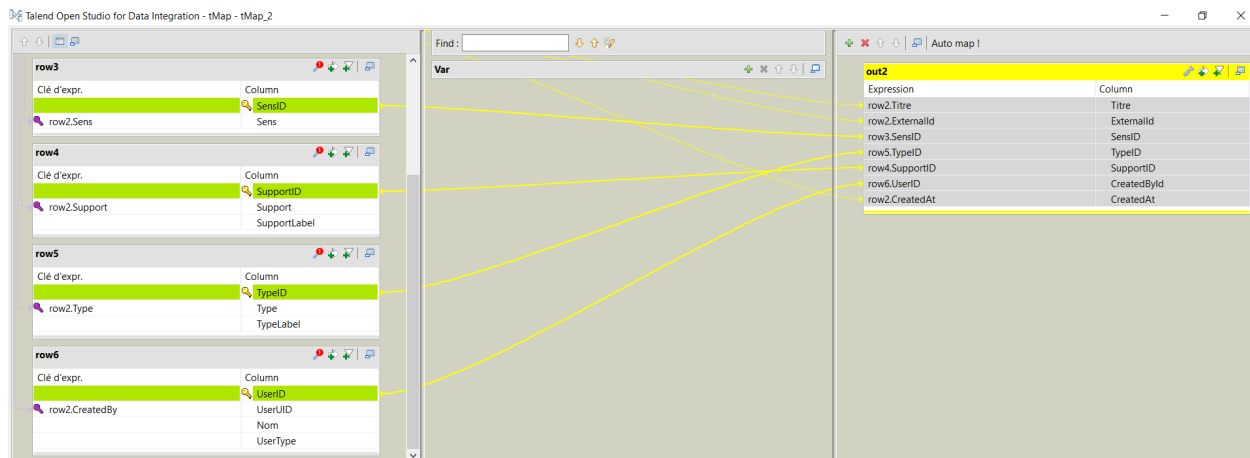


On procède par les mêmes étapes pour ajouter les autres composants tDBInput afin de charger les données des autres tables.



ci dessus le schéma complet du job dim documents. on va se focaliser sur la configuration de tmap_2.





le panneau de gauche affiche les flux de données entrants, correspondant aux composants qui alimentent le tMap. Par exemple, si vous reliez un tDBInput PostgreSQL à tMap, les colonnes de la table source s'affichent ici.

- ☐ Chaque table d'entrée peut contenir des colonnes d'entrée (ou clés d'entrée) qui sont les champs récupérés depuis la base de données correspondante. on va glisser-déposer ces colonnes d'entrée vers la droite pour les mapper aux colonnes de sortie, ou appliquer des transformations.
- Pour relier les colonnes de la table d'entrée (Documents) avec les colonnes de chaque table provenant des différents composants tDBInput, voici les étapes et l'approche à suivre, en vous assurant de faire référence aux clés appropriées :
- **Clé de Support** : Reliez la colonne Support (issue de la table DimSupport dans le composant tDBInput) à la colonne correspondante dans Documents.
 - **Par exemple** : **DimSupport.SupportID** → **Documents.Support**.
- **Clé de Sens** : Reliez la colonne Sens (issue de la table DimSens) à la colonne Sens dans Documents.
 - **Exemple** : **DimSens.SensID** → **Documents.Sens**.
- **Clé de Type** : Reliez la colonne Type (issue de DimType) à la colonne Type de Documents.
 - **Exemple** : **DimType.TypeID** → **Documents.Type**.
- **Clé de User** : Reliez la colonne User (issue de DimUser) à la colonne CreatedBy dans Documents.

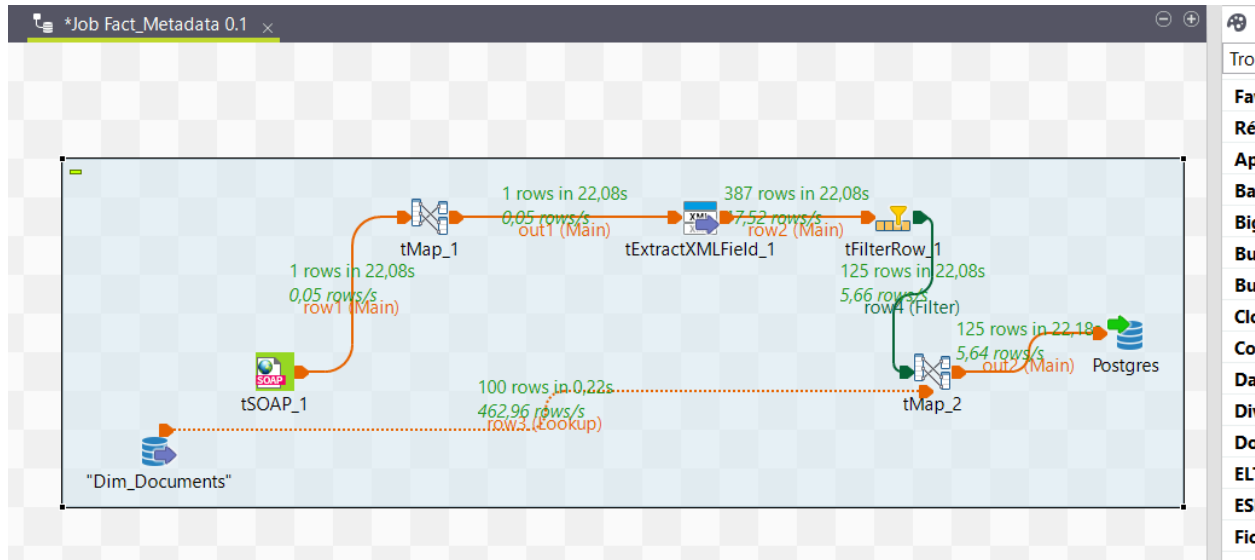
- Exemple : DimUser.UserID → Documents.CreatedBy.

row4	
Clé d'expr.	Column
row2.Support	SupportID
	Support
	SupportLabel
row5	
Clé d'expr.	Column
row2.Type	TypeID
	Type
	TypeLabel
row6	
Clé d'expr.	Column
row2.CreatedBy	UserID
	UserID
	Nom
	UserType

ExternalId	
Sens	
Support	
Type	
CreatedBy	
CreatedAt	
row3	
Clé d'expr.	
row2.Sens	
row4	
Clé d'expr.	
row2.Support	
row5	
Clé d'expr.	
row2.Type	
row6	

★ Création du JOB Fact_Metadata:

On procède par les mêmes étapes pour ajouter les composants tSOAP_1, tDBInput , tMAP , tExtractXMLField , Postgres.



Designer | Code

Job(Fact_Metadata 0.1) Contexts(Fact_Metadata) Composant Exécuter (Job Fact_Metadata)

tExtractXMLField_1

Paramètres simples Paramètres avancés Paramètres dynamiques Vue Documentation

Type de propriété Built-in

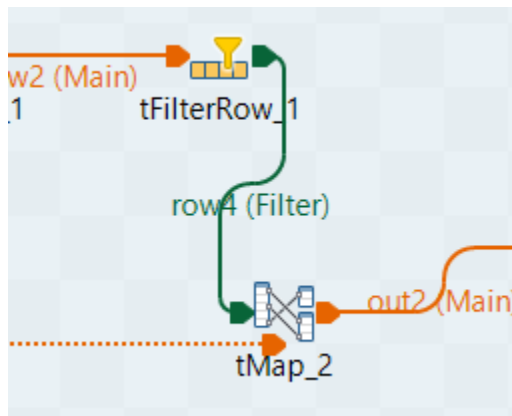
Schéma Built-in La variable associée à ce paramètre est : _PROPERTY_TYPE_

Champ XML Body

Requête XPath boucle "/GetCompleteEliseMailByDateIndexResponse/GetCompleteEliseMailByDateIndexResult/a:d/b:CompleteEliseMailColle

Mapping

Column	Requête XPath	Obtenir les nœuds
Metadata	"/b:CustomFields/"	<input type="checkbox"/>
ExternalId	"/b:MailId/"	<input type="checkbox"/>



L'utilisation de tFilterRow permet de s'assurer que seules les actions ayant une valeur non vide dans champs_administres sont prises en compte pour le traitement ultérieur, excluant ainsi les métadonnées vides et focalisant sur les actions pertinentes.

DesignerCode

Job(Fact_Metadata 0.1)Contexts(Fact_Metadata)ComposantExécuter (Job Fact_Metadata)

tFilterRow_1

Paramètres simples

Paramètres avancés

Paramètres dynamiques

Vue

Documentation

SchémaBuilt-inModifier le schémaSync colonnes

Opérateur logique utilisé pour combiner les conditionsEt*

Conditions

Colonne d'entrée	Fonction	Opérateur	Valeur
Metadata	Vider	Différent de	"<CHAMPS_ADMINISTRES />"
Metadata	Vider	Différent de	""

+✕↑↓📄📋

☐ Utiliser le mode avancé

★ résultat tmap_2 :

Talend Open Studio for Data Integration - tMap - tMap_2

Find:

Var

out2

Expression
row2.Metadata
row2.ExternalId

Column
Titre
ExternalId

row4

Column
Metadata
ExternalId

row3

Cle d'expr.

row2.ExternalId

Column
Titre
ExternalId
SensID
TypeID
CreatedById
CreatedAt

Éditeur de schéma Éditeur d'expression

row3

Colonne	Cle	Type	N.	Modèle de date (C...	Longueur	Précision	Par déf.	Comment...
Titre	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		21474836...	0		
ExternalId	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		21474836...	0		
SensID	<input type="checkbox"/>	int	<input type="checkbox"/>		10	0		
TypeID	<input type="checkbox"/>	int	<input type="checkbox"/>		10	0		
CreatedById	<input type="checkbox"/>	int	<input type="checkbox"/>		10	0		
CreatedAt	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		21474836...	0		

out2

Colonne	Cle	Type	N.	Modèle de date (Ct...	Longueur	Précision	Par déf.	Comment...
Titre	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>					
ExternalId	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>					

Appliquer Ok Annuler

process

Fichier Accueil Partage Affichage

Ce PC > Bureau > TALENDPROJECT > process

Rechercher dans : process

Nom	Modifié le	Type	Taille
DimAttachement_0.1	04/09/2024 14:37	Fichier source Pro...	2 Ko
DimSens_0.1	04/09/2024 14:37	Fichier source Pro...	2 Ko
DimSupport_0.1	04/09/2024 14:37	Fichier source Pro...	2 Ko
DimType_0.1	04/09/2024 14:37	Fichier source Pro...	2 Ko
DimUser_0.1	04/09/2024 14:37	Fichier source Pro...	2 Ko
Documents_0.1	04/09/2024 14:37	Fichier source Pro...	2 Ko
Fact_Metadata_0.1.item	04/09/2024 14:37	Fichier ITEM	56 Ko
Fact_Metadata_0.1	04/09/2024 14:37	Fichier source Pro...	2 Ko
Fact_Metadata_0.1.screenshot	04/09/2024 14:37	Fichier SCREENSH...	120 Ko

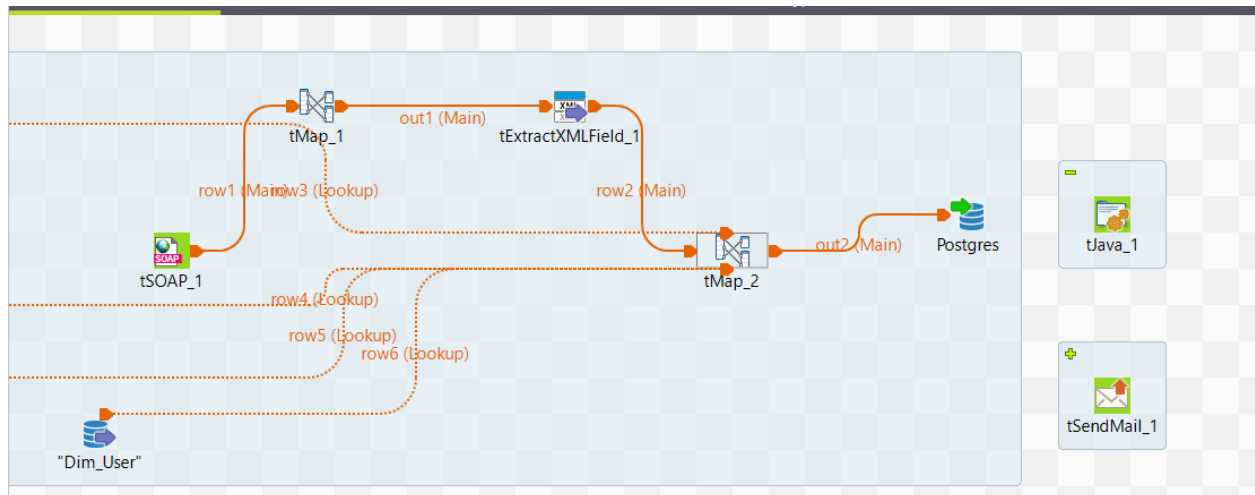
Type: Fichier source Properties
Taille: 1,47 Ko
Modifié le: 04/09/2024 14:37

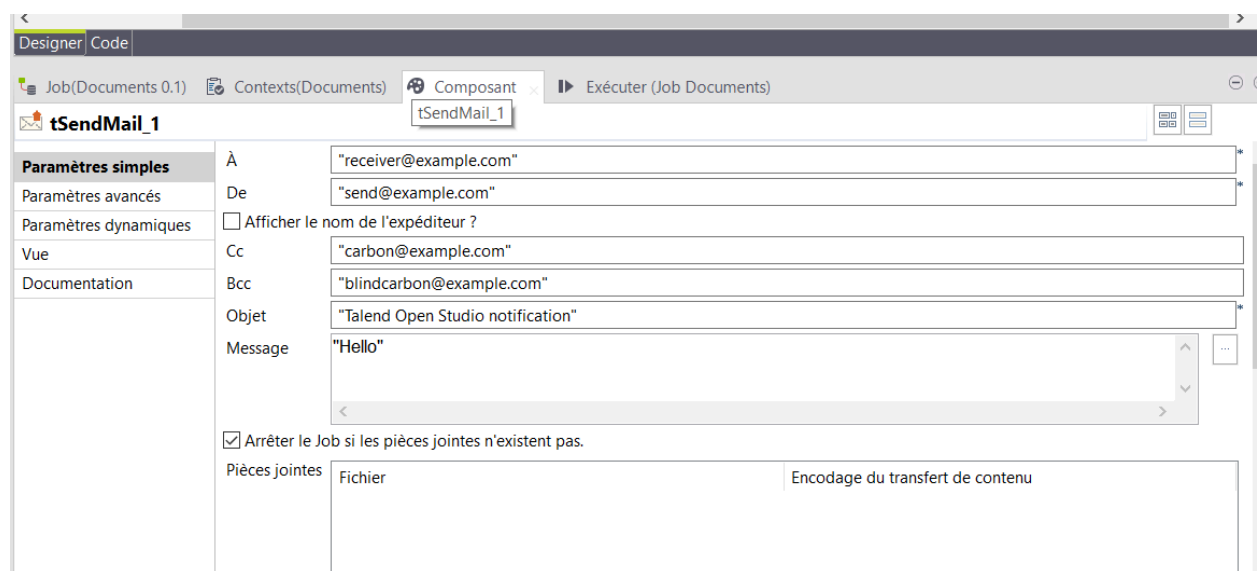
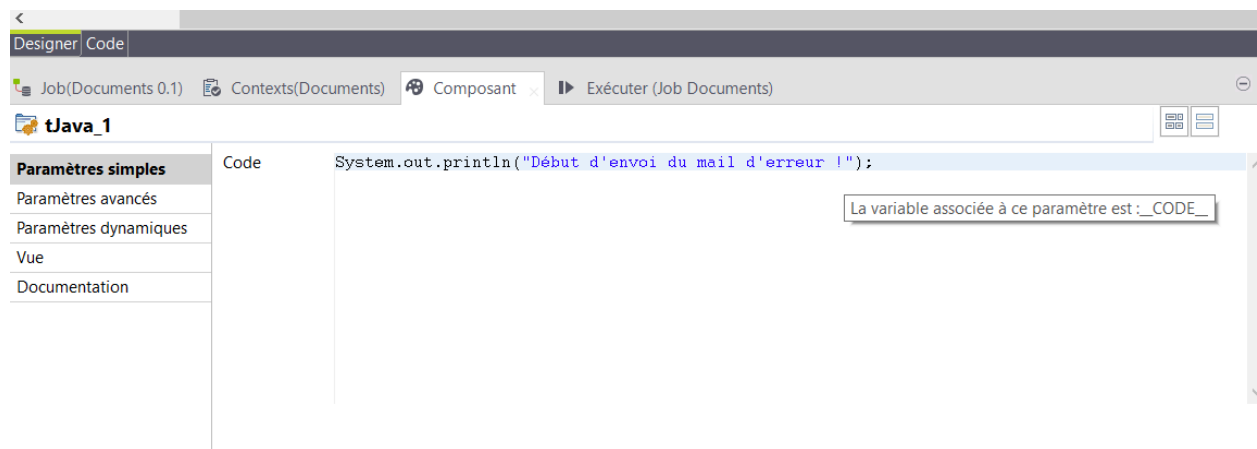
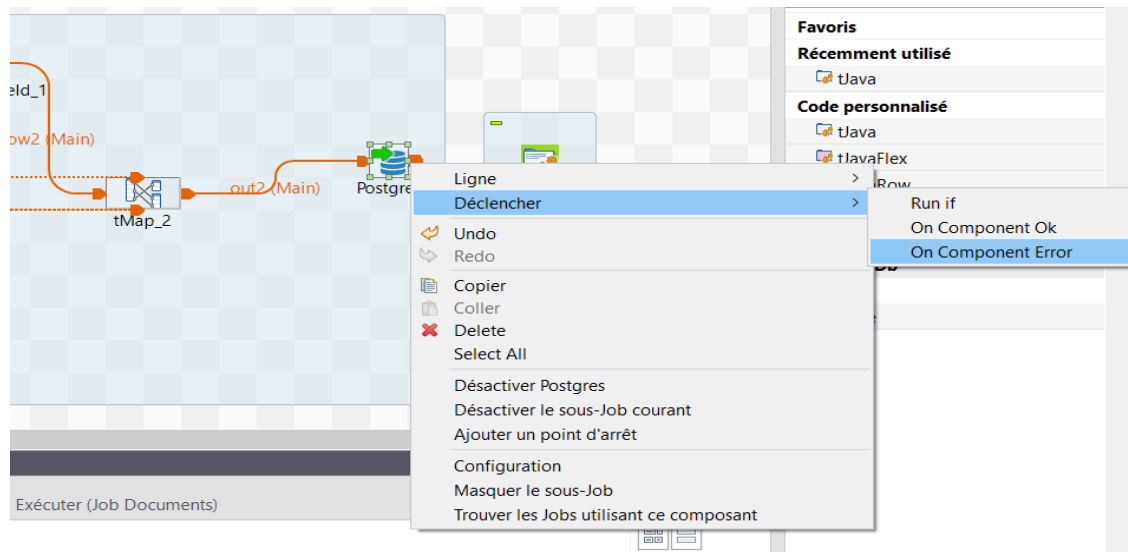
9 élément(s)

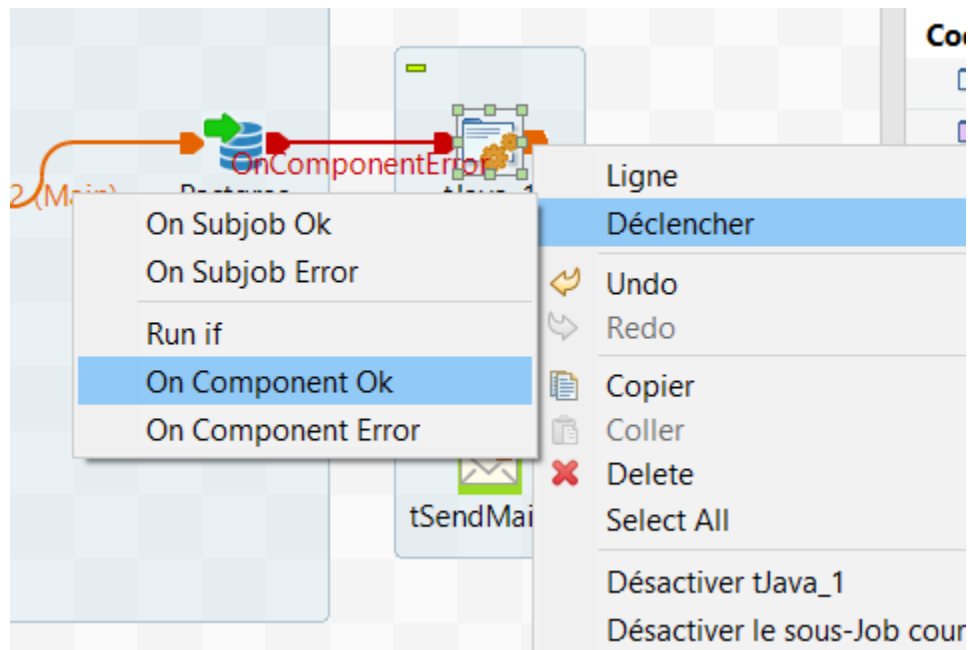
ces fichiers représentent les différentes dimensions créées dans notre projet Talend, ainsi que des métadonnées. Nous pouvons les exécuter ou vérifier leur configuration via l'interface Talend pour nous assurer que les processus sont bien connectés aux bonnes sources de données et qu'ils fonctionnent correctement.

6. surveillance et gestion des erreurs

Notification par E-mail : Utilisez le composant `tSendMail` pour envoyer des notifications par e-mail en cas d'erreur critique ou de succès de l'exécution. Configurez-le pour alerter les administrateurs en cas de problème.







À "amen.khlifi.21@gmail.com" *

De "amen.khlifi.03@gmail.com" *

☒ Afficher le nom de l'expéditeur ? La variable associée à ce paramètre est : _FROM_ *

Nom de l'expéditeur "Neolledge" *

Cc "" *

Bcc "" *

Objet "***ERREUR DU JOB TALEND **" *

Message "Bonjour
Votre Job est en erreur . veuillez vérifier vos paramètres de connexion .
Merci .

☐ Arrêter le Job si les pièces jointes n'existent pas.

Hôte SMTP "smtp.office365.com" * Port SMTP 587 *

☐ Support de SSL ☒ Support de STARTTLS Importance Normal ▾

☒ Authentification requise ?

Utilisateur "amenkhlifi.03@gmail.com" * Mot de passe ***** *

☒ Arrêter en cas d'erreur

Conclusion

Le projet de migration vers Talend et d'optimisation des processus ETL a été mené à bien avec succès, atteignant les objectifs définis et répondant efficacement aux besoins spécifiques en matière de gestion et d'analyse des données. Voici un résumé des principales réalisations et des impacts observés :

1. Objectifs Atteints

- **Migration des Données** : La migration des données vers Talend a été effectuée avec succès en utilisant des mécanismes de synchronisation basés sur les web services (SOAP UI). Cette approche a permis de centraliser et d'améliorer la gestion des informations provenant de la base de données principale.
- **Optimisation des Processus ETL** : Les processus ETL ont été optimisés pour améliorer les performances et réduire le temps de traitement des données. L'intégration des composants Talend a facilité l'automatisation et simplifié les flux de données.
- **Suivi et Alertes** : Un système de notifications par e-mail a été mis en place pour alerter les administrateurs en cas d'erreurs critiques ou de succès significatifs des jobs. Cela assure une réaction rapide et une gestion efficace des incidents.

2. Défis Rencontrés

- **Problèmes de Connexion** : La configuration des connexions aux bases de données a présenté certains défis, mais ceux-ci ont été résolus par des ajustements précis et une analyse approfondie.
- **Problèmes lors de l'Ajout des Composants** : Nous avons rencontré des difficultés, notamment avec le composant **tSOAP** en raison de problèmes de configuration des messages SOAP. De plus, des complications sont survenues avec le composant **tExtractXMLField** lors de la spécification des requêtes XPath et de la gestion des boucles.
- **Gestion des Erreurs** : La gestion des erreurs et des exceptions au sein des processus ETL a exigé une attention particulière pour garantir une fiabilité maximale des flux de données.

En conclusion, le projet a non seulement atteint ses objectifs initiaux mais a également posé des bases solides pour des améliorations futures. Les connaissances acquises et les

compétences développées au cours de cette initiative seront essentielles pour la gestion continue et l'optimisation des processus de données.