



ENET'COM

DATA LAKE

EXTRACTION, TRANSFORMATION ET SAUVEGARDE DE
DONNÉE
RAPPORT

ETL avec Talend

Élèves :

SALHI ZIED

MOALLA SAIF

Enseignant :

Ben MASSAOUD INES

28 février 2023

Table des matières

1	Les outils utilisés	2
1.1	Le logiciel Talend	2
1.2	XAMP(MySQL)	2
2	Extraction des données	4
2.1	L'outil utilisé pour extraire les données	5
2.2	Modification manuelle des bases de donnée	6
3	Creation d'une base de donnée avec XAMP/phpmyadmin)	6
4	Transformer les données :Travailler avec Talend	6
4.1	Liaison avec la base de donnée	6
4.2	Importer les fichier CSV	7
4.3	Création des Jobs/Création des inputs/Création des outputs	8
4.4	Configuration au niveau du tMap	9
5	Sauvegarde des données(Load)	9
5.1	Exécuter les Jobs	9
6	Modèle Multidimensionnel	10

1 Les outils utilisés

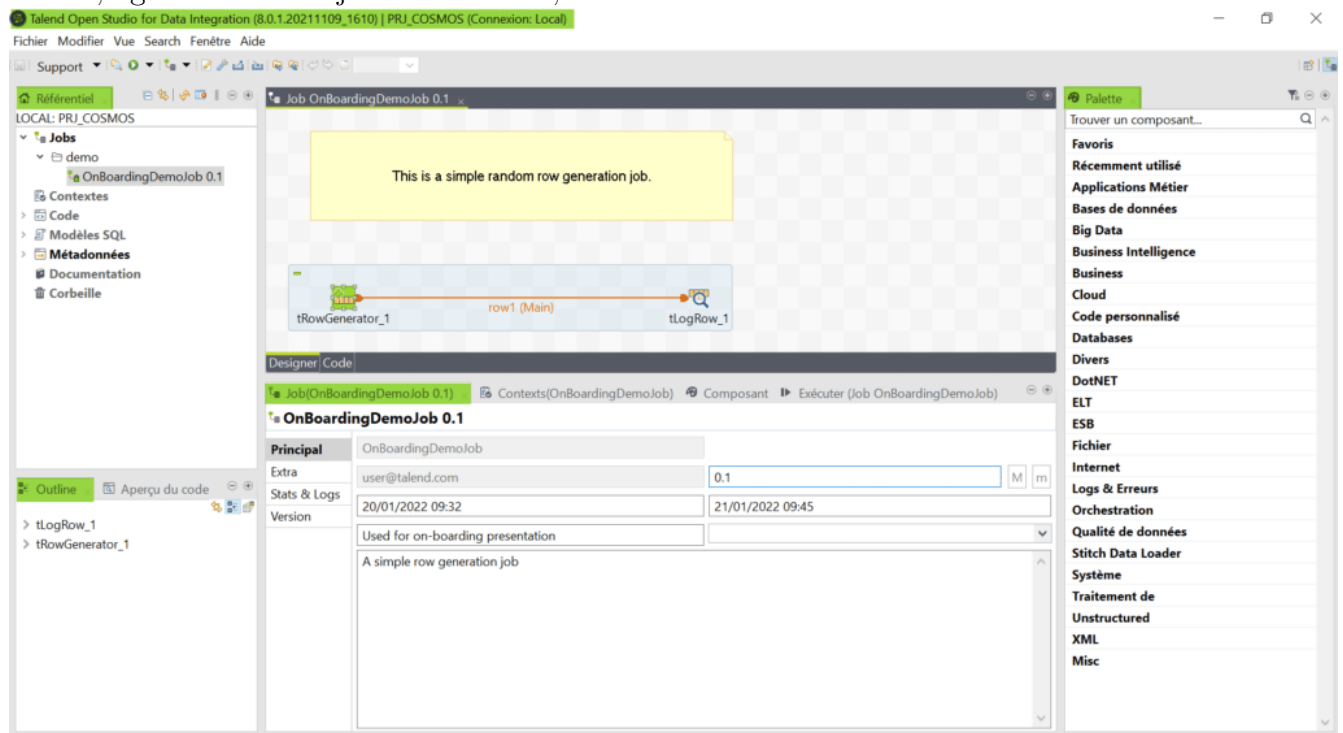
1.1 Le logiciel Talend

Il y a plusieurs logiciels appartenants à la fammille Talend,on a **Talend pour data Integration,pour Big Data...**

Dans mon projet j'ai travailler avec **Talend for Data Integration.**

Talend for Data Integration offre plusieurs fonctionnalités telques l'intégration des données ,création des tables de fait et de dimension.

Ci joint l'intrface du logiciel Talend,il y l'interface de palette ou on trouve des options de travail,à gauche ou on ajoute les tables ,les fichiers...

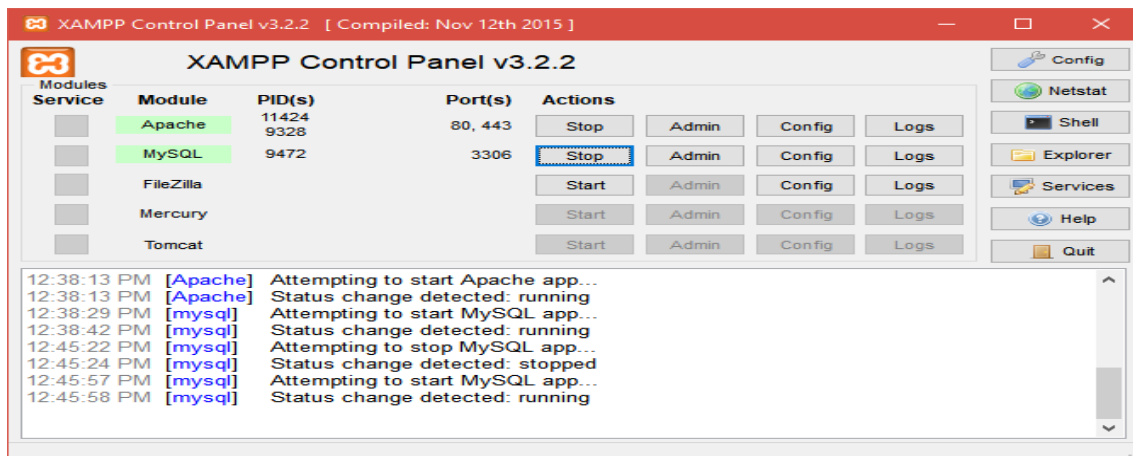


1.2 XAMP(MySQL)

XAMPP est un ensemble de logiciels permettant de mettre en place un serveur Web local, un serveur FTP et un serveur de messagerie électronique.

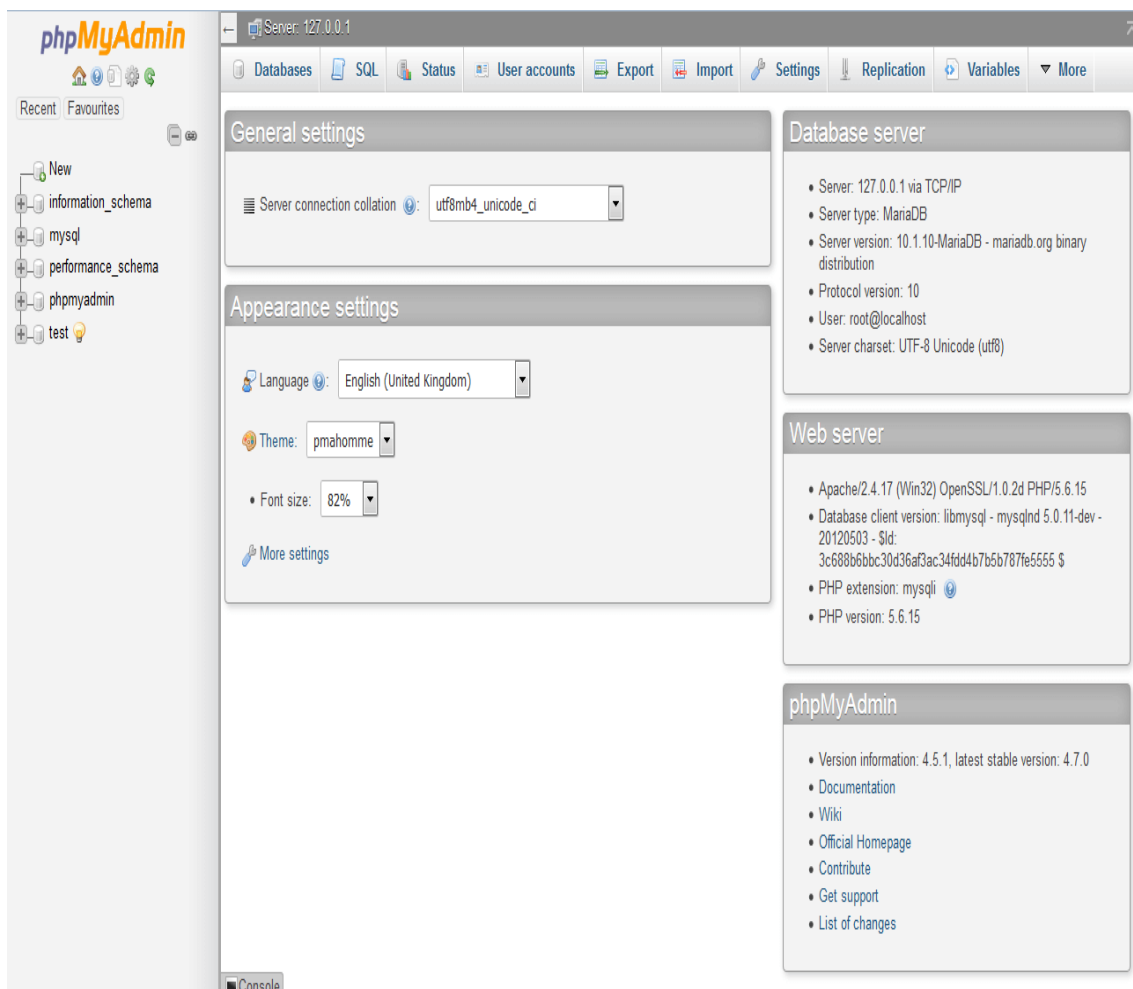
On peut lancer MySQL à partir de XAMP en lançant les serveurs nécessaires.

Voila la figure :



En ouvrant notre Localhost :

Ici on fait créer notre base de donnée qui va contenir les deux tables.



2 Extraction des données

L'extraction de donnée est une phase très importante. Dans notre travail j'ai choisi deux sites différents à partir des quelles j'extrai les données.

la première c'est un tableau qui contient les statistiques d'un groupe sur **facebook**(nombre de réaction, commentaire, j'aime, Love,...)

Voila le premier tableau :

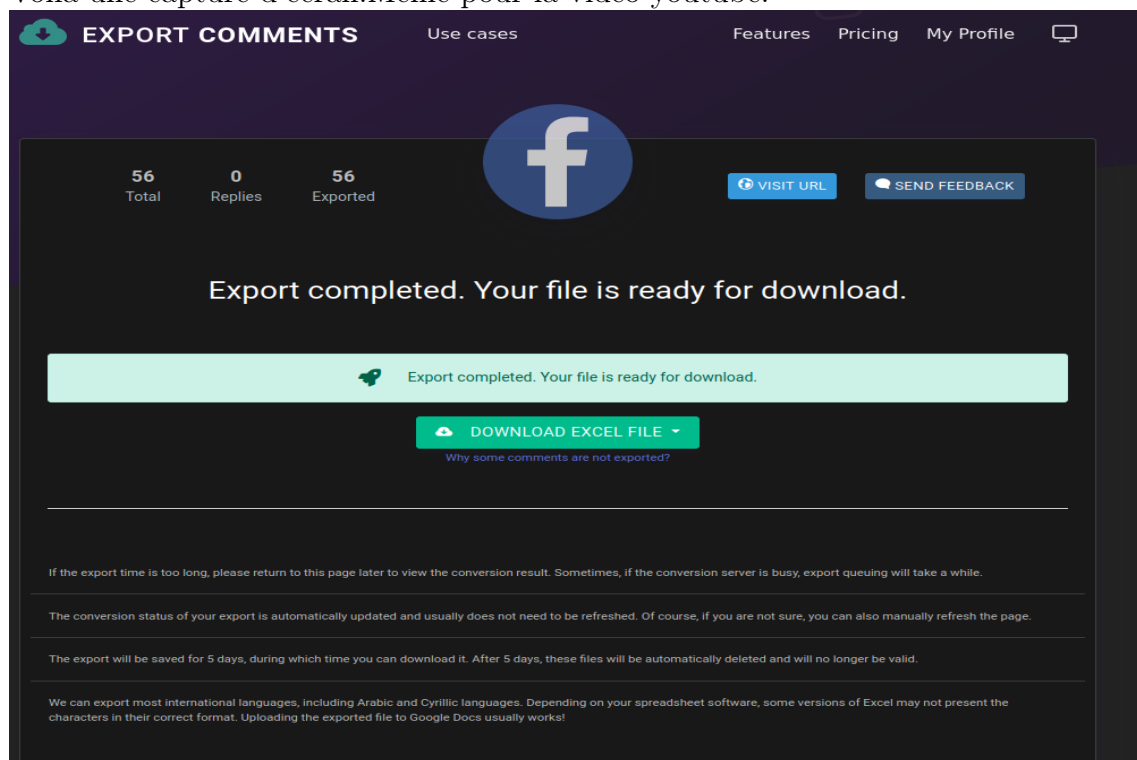
Post Url	(view source)	Comme	Reactio	Shares	Views	LOVE	LIK
https://www.facebook.com/105591702140900	view post	0	10	0	0	10	
https://www.facebook.com/105932745440129	view post	0	35	0	0	32	
https://www.facebook.com/105935142106556	view post	0	23	8	0	19	
https://www.facebook.com/106298335403570	view post	1	15	0	0	12	
https://www.facebook.com/106320845401319	view post	0	31	11	0	27	
https://www.facebook.com/106625758704161	view post	4	31	9	0	28	
https://www.facebook.com/106972412002829	view post	2	34	5	0	31	
https://www.facebook.com/107750975258306	view post	0	20	1	0	17	
https://www.facebook.com/109498588416878	view post	0	26	3	0	25	
https://www.facebook.com/109735401726530	view post	0	18	4	0	17	
https://www.facebook.com/109756888391048	view post	0	19	3	0	17	
https://www.facebook.com/110055651694505	view post	2	23	3	0	23	
https://www.facebook.com/110072245026179	view post	1	20	1	0	19	
https://www.facebook.com/111762081523862	view post	0	21	6	0	21	
https://www.facebook.com/111975018169235	view post	7	51	19	0	50	
https://www.facebook.com/112519604781443	view post	0	16	3	0	16	
https://www.facebook.com/112952761404794	view post	0	29	1	0	28	
https://www.facebook.com/113008971399173	view post	0	17	3	0	17	
https://www.facebook.com/113423284691075	view post	0	18	4	0	17	
https://www.facebook.com/113735974659806	view post	0	14	1	0	14	
https://www.facebook.com/113844111315659	view post	0	10	0	0	10	
https://www.facebook.com/113932911306779	view post	0	7	2	0	7	
https://www.facebook.com/113933691306701	view post	1	5	2	0	5	
https://www.facebook.com/114012824632121	view post	0	13	2	0	12	
https://www.facebook.com/114029691297101	view post	0	11	1	0	10	
https://www.facebook.com/114403104584084	view post	0	0	0	0	7	

Le deuxième tableau représente les réaction sur une vidéo sur **Youtube**. voila le tableau :

[illegible]

2.1 L'outil utilisé pour extraire les données

J'ai utilisé le site **exportcomment.com** pour extraire les données.
On faisant entrer le lien de la page qu'on veut extraire les données et il nous génère des
fichier **CSV**.
Voila une capture d'écran.Même pour la video youtube.



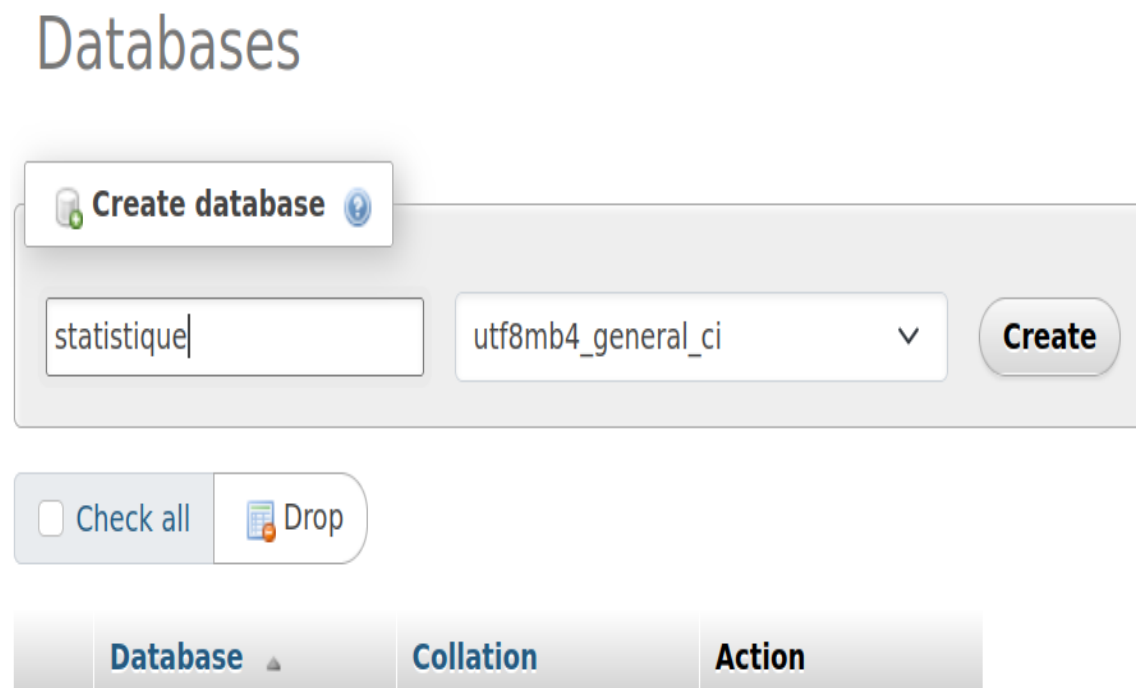
2.2 Modification manuelle des bases de donnée

On fait chercher les valeurs NULL et on les attribut une valeur nulle(0) pour que notre table soit claire et présentable.

En suite, on supprime les lignes inutiles qui ne portent pas une information .

3 Creation d'une base de donnée avec XAMP(phpmyadmin)

On doit créer maintenant une base de donnée nommée **Statistique**



Databases

Create database

statistique utf8mb4_general_ci Create

Check all Drop

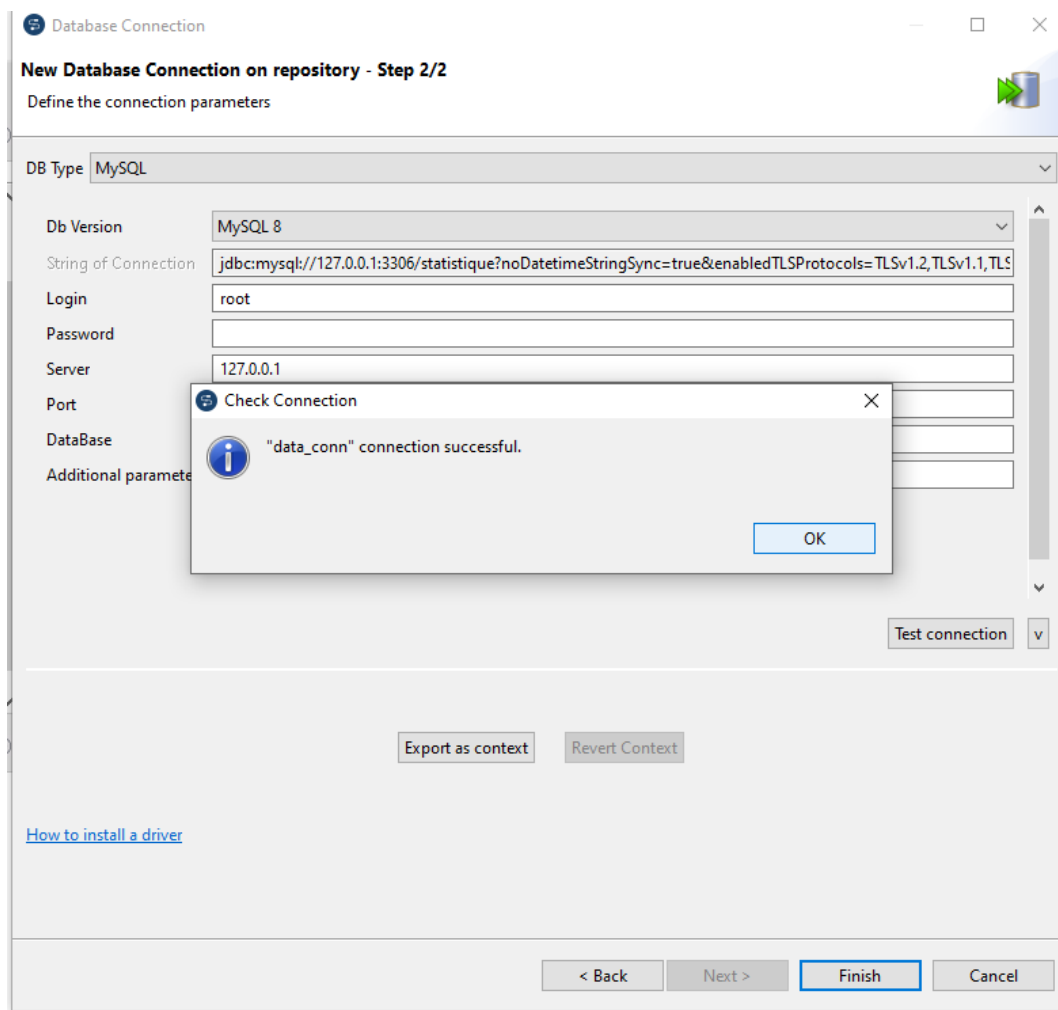
Database	Collation	Action
----------	-----------	--------

4 Transformer les données :Travailler avec Talend

4.1 Liaison avec la base de donnée

On fait tout d'abord la création d'une liaison avec notre base de donnée **statistique**. Cette base de donnée sera notre cible.

Voila la capture d'écran qui mentionne la liaison.



4.2 Importer les fichier CSV

On fait créer des fichier excels qui vont contenir vos deux fichiers excels.
On fait ajouter nos fichier en appliquant les modification qu'on veut et tout en conservant les données.
On fait sélectionner la première ligne comme header.
Dans la capture d'écran suivante l'exemple :

New Excel File

File - Step 3 of 4
Add a Metadata File on repository
Define the setting of the parse job

File Settings
Encoding: UTF-8
☐ Advanced separator(for number)
Thousands separator: ,
Decimal separator: .

Rows To Skip
If any rows must be ignored, specify the following parameters
Header: ☒ 1
Footer: ☐

Metadata column setting
First column: 1
Last column:

Limit Of Rows
If the number of lines must be limited, specify this number
Limit: ☐

Preview
☒ Set heading row as column names
Refresh Preview

Column 0	Column 1	Name (click to view profile)	Date	Likes	isHearted	isPinned	Comment
1		Brian McCarthy	Wed Nov 09 09:59:23 WAT 2022	4	no	no	Humans are the same way. Humans are e
2		Flo Miller	Fri Nov 11 02:44:53 WAT 2022	2	no	no	Amen!
3		Initiate Change Spark Peace	Sat Nov 12 02:35:17 WAT 2022	2	no	no	Powerful Video!!!
4		Christopher's Fidgitty friends	Sun Nov 13 02:57:42 WAT 2022	1	no	no	WOW 🤔🤔🤔 Amen everybody just clim
5		BALI CULTURE AND NATURE	Tue Nov 15 09:08:45 WAT 2022	1	no	no	Thanks for sharing ,very helpful,see you!!
6		Mwiine Abel Wilson	Mon Dec 12 11:00:04 WAT 2022	2	no	no	This is really Good!!!!, thanks Steve ❤️

Export as context
Revert Context

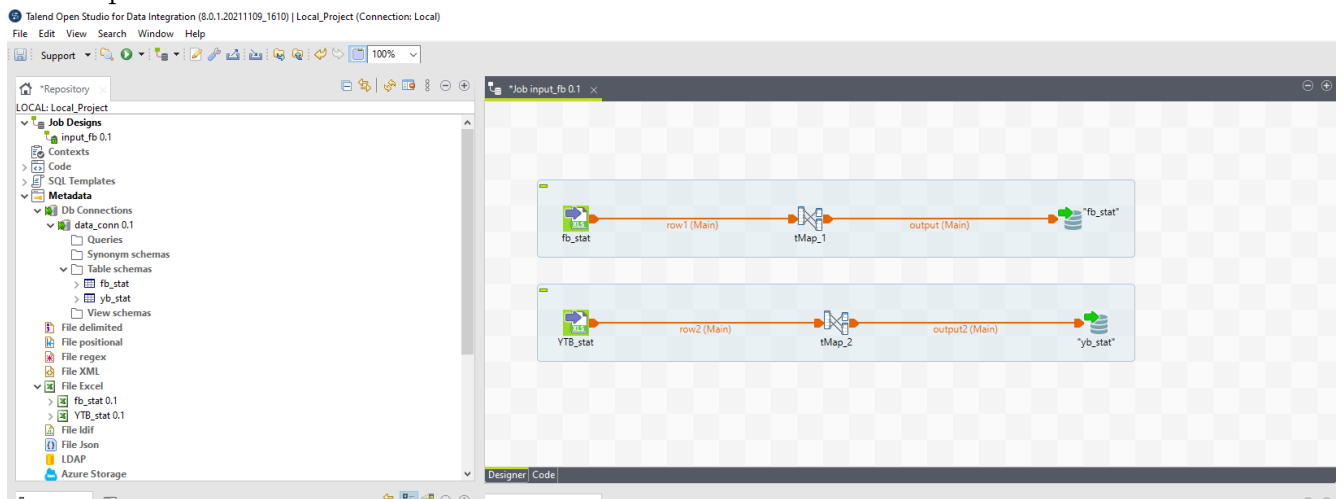
< Back
Next >
Finish
Cancel

4.3 Création des Jobs/Création des inputs/Création des outputs

On fait la création du job qui va contenir les étapes d'insersion des données dans notre base de donnée.

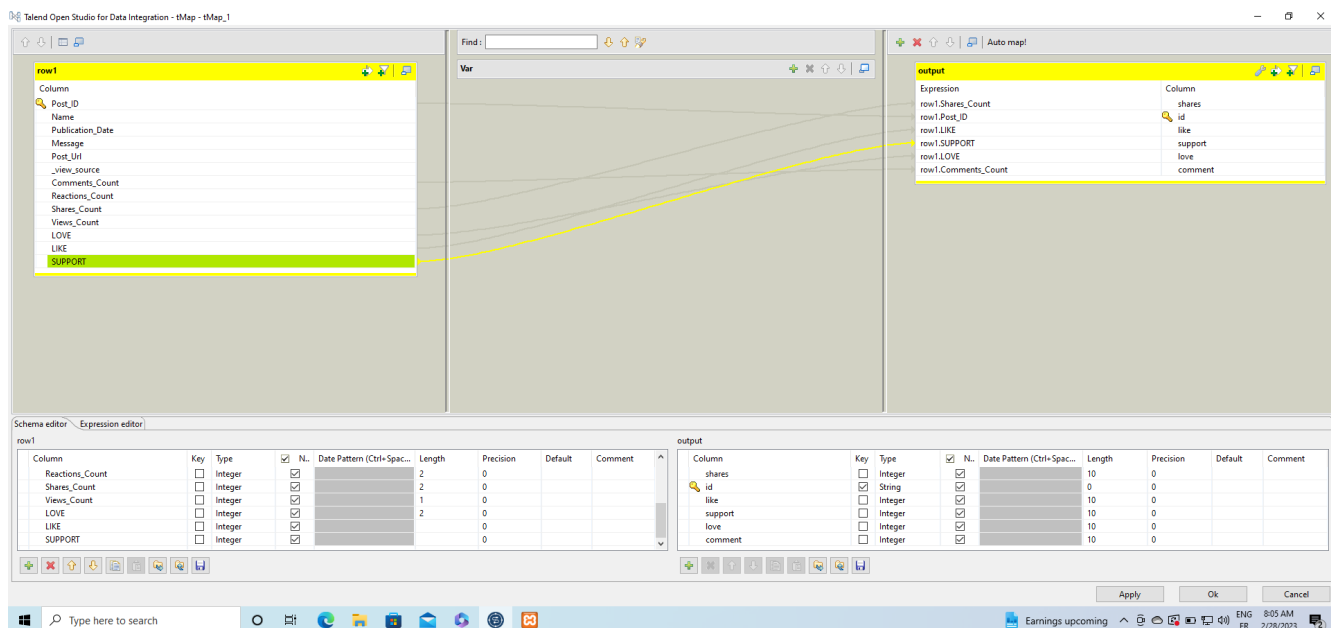
Les inputs sont nos fichiers excel à envoyer.

Les outputs sont nos tables vides dans la base de donnée.



Avec le **tMap** on fait la relation entre les inputs et les outputs

4.4 Configuration au niveau du tMap



5 Sauvegarde des données(Load)

5.1 Exécuter les Jobs

On fait cliquer sur **Exécuter** pour réaliser le job.
On obtient nos bases de données sur MySQL.

6 Modèle Multidimensionnel

