

REPUBLIQUE DU SENEGAL

UN PEUPLE-UN BUT-UNE FOI



Ministère de l'économie, des finances et du plan

Agence Nationale de la Statistique et de la Démographie (ANSD)

Ecole Nationale de la Statistique et de l'Analyse Economique

NOTE EXPLICATIVE DATAWAREHOUSE

MODELISATION D'UN ENTREPOT DE DONNEES DES COMMANDES, VENTES, LIVRAISONS ET VERSEMENTS POUR UNE ENTREPRISE COMMERCIALE

Rédigé par :

Mohamed NIANG, élève ingénieur des travaux statistiques

Sous la supervision de :

M. Ibrahima DIOP, Ingénieur informaticien

Année 2018/2019

Introduction

L'objectif de ce projet est de mesurer à travers un cas pratique les connaissances acquises en travaux pratiques de datawarehouse (MySQL, SQL Server, Talend Open Studio et Power BI Desktop). Son but est la modélisation d'un entrepôt de données des commandes, ventes, livraisons et versement pour une entreprise commerciale.




Cette entreprise vend principalement du ciment en deux types (32.5 et 42.5), du sucre et du béton. Ainsi une base de données a été fournie par l'entreprise et contenant diverses tables.

La table « commande » enregistre l'ensemble des commandes effectuées par l'entreprise auprès des fournisseurs qui se trouvent dans la table « fournisseurs ». Certaines commandes sont décomposées (subdivisées selon les livraisons à faire) avant d'être livrées aux clients. Au cas échéant, les différentes parties sont enregistrées dans la table « part_client_commande » avant d'être livrées.

Les livraisons faites par l'entreprise à ses clients sont enregistrées dans la table « livraison ». Cette table contient la quantité livrée, le coût unitaire, le client ainsi que la monnaie utilisée pour la transaction.

Les informations des clients se trouvent dans la table « clients » et le fichier Excel « infos_clients » qui contient l'adresse de chaque client. Les versements ou paiements effectués par les clients se trouvent dans la table « versement ». Le responsable de l'entreprise souhaite disposer de différents outils qui l'aideront à mieux gérer les commandes, les livraisons, les versements. Il voudrait aussi partager quelques-uns des rapports avec ses collaborateurs.







Afin de pouvoir créer ces différents rapports, nous allons suivre la démarche suivante :

-  Créer l'entrepôt de données sur Microsoft SQL Server ;
-  Charger les données sources qui sont dans une base MySQL et un fichier Excel vers l'entrepôt de données en utilisant le logiciel Talend Open Studio ;
-  Créer les différents rapports sur Power BI Desktop.

PRELIMINAIRE

Afin de pouvoir concevoir les rapports, il va falloir d'abord effectuer modéliser l'entrepôt de données qui doit être utiliser dans SQL Serveur.

Le dossier du projet se présente comme suit :

Nom	Modifié le	Type	Taille
 Backup du DW et format SQL	03/03/2019 12:30	Dossier de fichiers	
 Données Sources utilisées	03/03/2019 12:31	Dossier de fichiers	
 Projet de Chargement TALEND	03/03/2019 12:49	Dossier de fichiers	
 MONPROJET_POWERBI.pbix	03/03/2019 18:54	Document Micros...	1 977 Ko
 Note explicative.pdf	03/03/2019 18:47	PDF Document	878 Ko
 PROJET_DW.xml	03/03/2019 16:18	Document XML	52 Ko




D'abord, il y a un sous dossier nommé « Backup du DW et format SQL » qui contient le backup de la base de données correspondant à l'entrepôt de données au format (*.bak) et au format SQL compatible avec SQL Server 2012.

Le backup de la base de données (*.bak) permet de générer à partir de SQL Server la base de données complète, les données, les tables et les relations existantes dans la base de données. Il y a aussi dans le même sous dossier le backup de la base de données (format SQL compatible avec SQL Server 2012) qui peut être exécuter directement sous SQL Server 2012.

Ensuite, il y a le second sous dossier nommé « Données Sources utilisées » qui contient les données sources utilisées (MySQL, Excel) ainsi que les éventuelles modifications que vous aurez apporté sur ces données. Le fichier de données nommé « db_its4_2018_modif » est le fichier de la base initiale qui a été modifié avec les nouvelles vues créés pour pouvoir les utiliser dans Talend.

Enfin, il y a un sous dossier nommé « Projet de Chargement TALEND » qui contient le Workspace du projet de chargement sous Talend qui se nomme « PROJET_DW ». Pour pouvoir l'utiliser directement sous Talend, il faut mettre le dossier « PROJET_DW » directement dans le dossier nommé Workspace du dossier d'installation de Talend Open Studio.

Aussi à ces dossiers s'ajoutent trois autres fichiers à savoir :

-  le projet crée sous Power BI Desktop pour les rapports ;
-  la note explicative sur la démarche utilisée ;
-  le schéma de l'entrepôt de données créés sous DB Designer Fork.

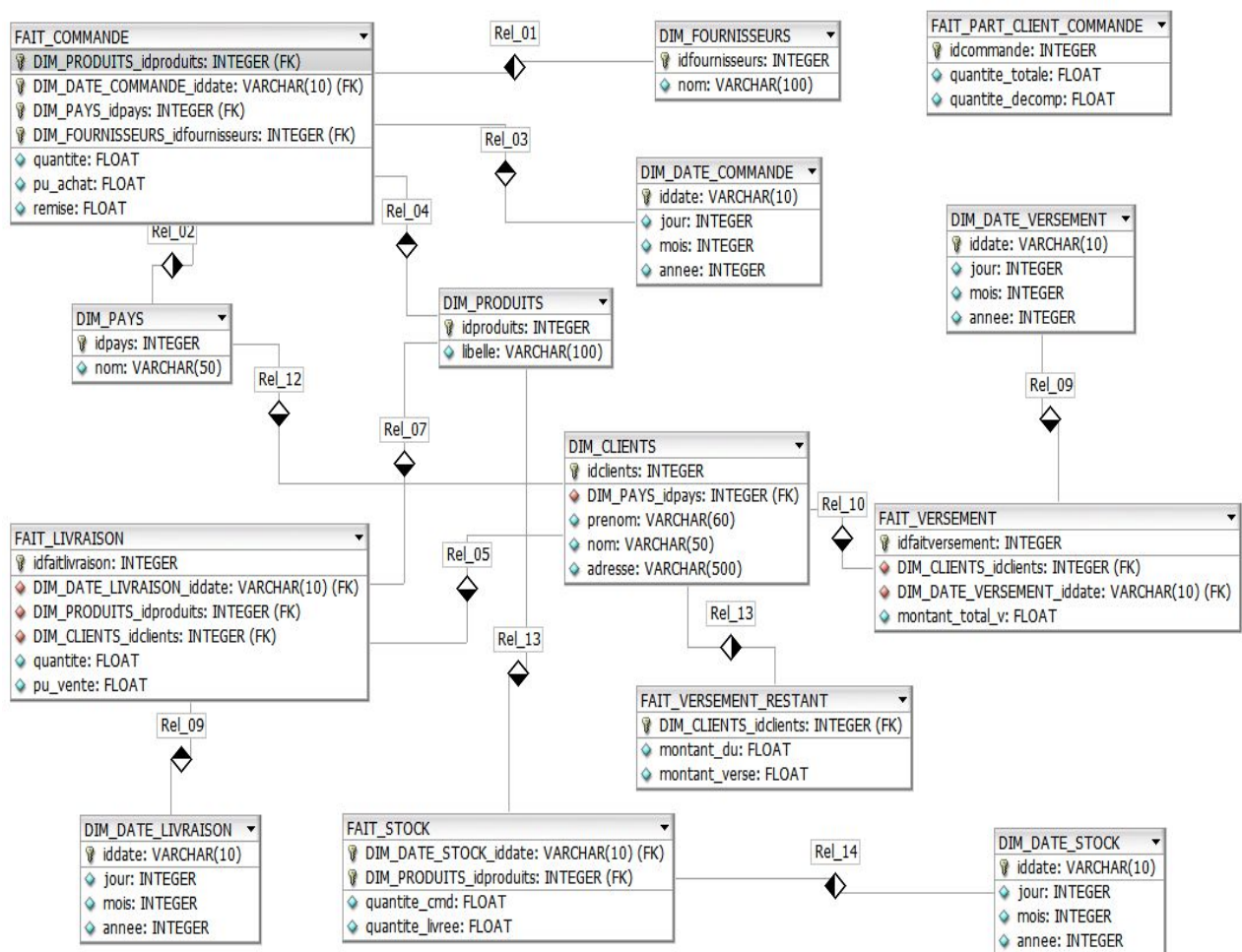
NB : le schéma de l'entrepôt de données doit être créé afin de répondre aux besoins des décideurs c'est-à-dire les rapports attendus dans Power BI.

I. Création de l'entrepôt de données

Pour créer l'entrepôt de données, il d'abord modéliser l'entrepôt de données afin de pouvoir exporter le schéma sous format SQL Serveur pour avoir le schéma et les tables de la ladite base.

1. Modélisation de l'entrepôt de données sous DB Designer

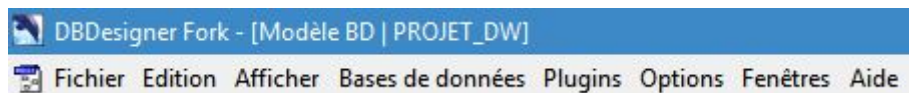
Comme il est dit plus haut, il va falloir d'abord modéliser l'entrepôt de données. Le schéma de la base obtenue sous DB Designer est le suivant :



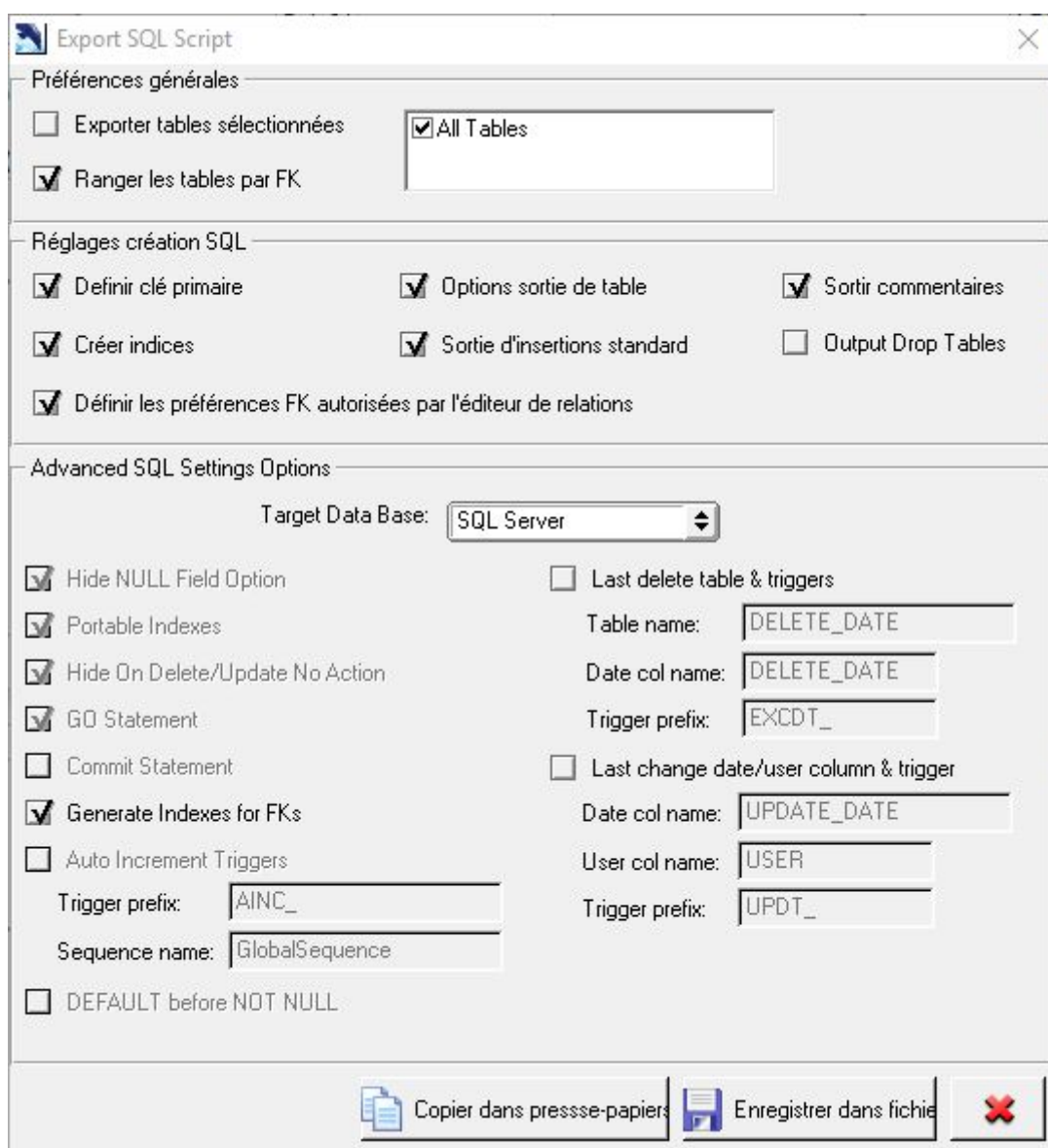
L'entrepôt de données a été constitué en utilisant des tables de dimensions et des tables de faits. Les tables de sert à stocker les indicateurs que l'on veut mesurer afin de produire les rapports. Toutes les tables de faits et de dimensions seront chargées à partir de Talend Studio soit en utilisant les tables de la base MySQL « tp_its4_2018 » ou bien en utilisant des requêtes sur ces tables.

2. Exportation du schéma vers MS SQL Serveur

Pour exporter le schéma sous format SQL Serveur, il faut aller sur l'onglet « fichier » de DB Designer comme le montre la figure suivante :



Puis cliquer sur « exporter » puis sur « Script de création SQL... ». Ensuite il sera ouverte une fenêtre comme le montre la figure suivante :

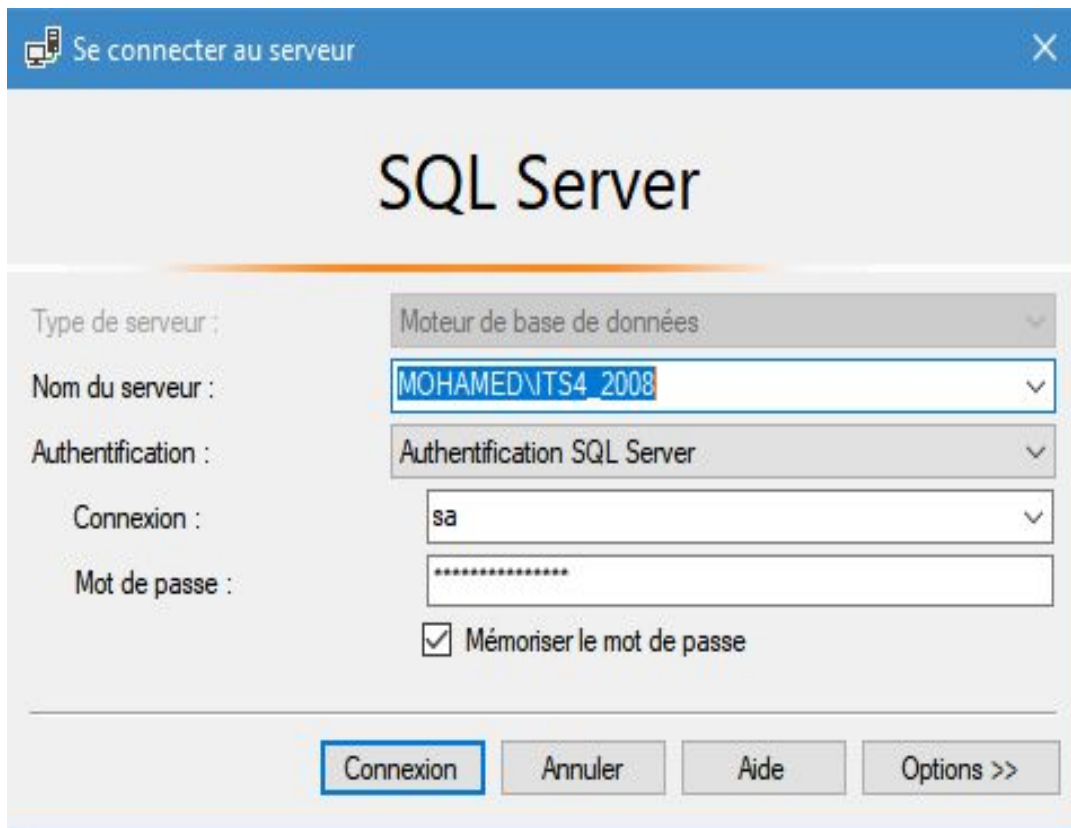


Enfin après avoir configuré les options comme montrées ci-dessus, il faut cliquer sur « Enregistrer dans fichier » puis sauvegarder le fichier dans le répertoire voulu.

3. Génération de l'entrepôt de données dans MSSQL

Pour générer la base dans MS SQL Serveur, il faut utiliser le fichier de format SQL récemment enregistré dans le répertoire voulu et l'exécuter dans SQL Serveur. Avant cela, si l'utilisateur souhaite utiliser cette application, il va falloir d'abord ouvrir MS SQL Serveur avec les **paramètres de connexion définis dans le projet**. Ce dernier fait référence aux paramètres de connexion utilisés dans MS SQL Serveur et dans le projet de chargement sous Talend. Donc il faut respecter les paramètres de connexion pour utiliser ce ladite projet. Sinon, il est bien possible de suivre la démarche de ce projet pour créer un entrepôt de données et l'alimenter sous SQL Serveur par le biais de Talend Studio.

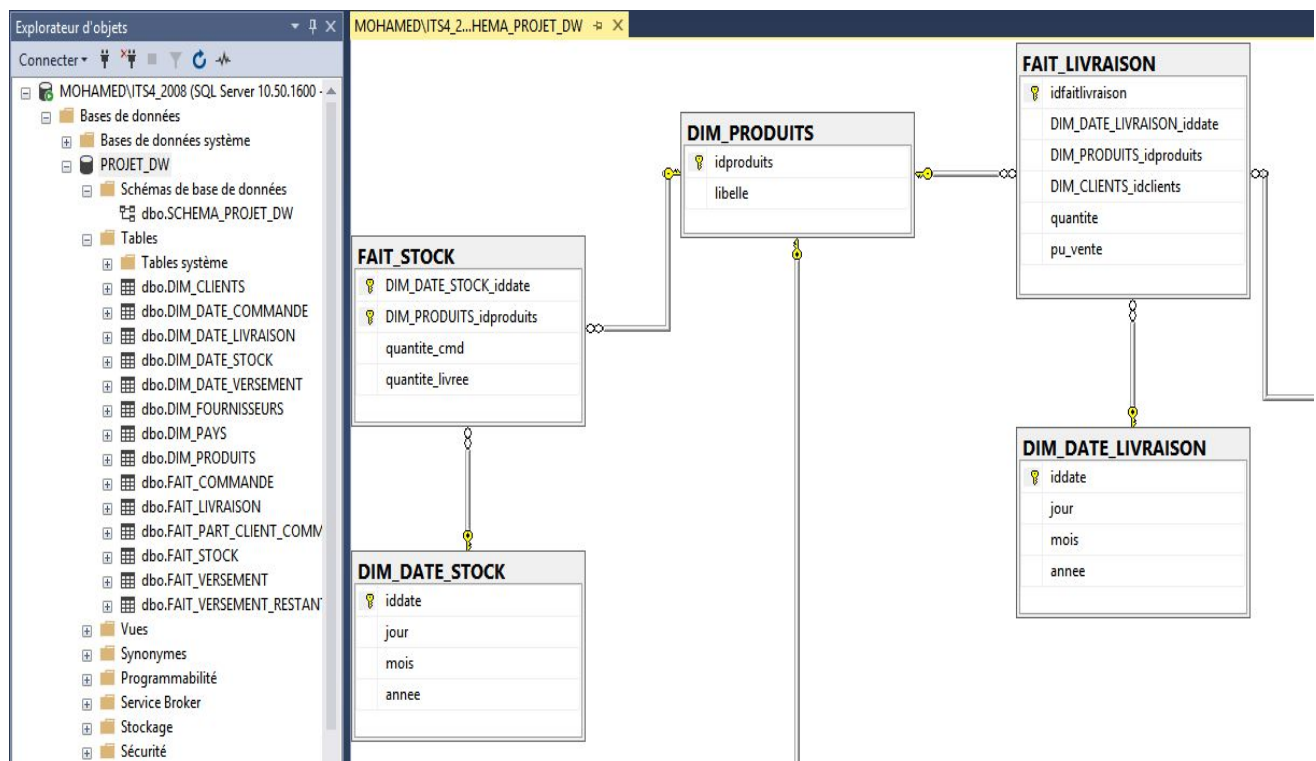
La connexion utilisée dans SQL Serveur est donnée par la figure suivante :



Le mot de passe utilisé ici est : LEAVEMEALONE1994.

Après que la connexion est établie avec succès, il faut créer une nouvelle base de données puis la donner un nom. Dans le cas de ce projet, la base de données s'appelle « PROJET_DW ». Ensuite il faut faire un clic droit sur la base puis cliquer sur l'option « nouvelle requête ». Après il suffit juste d'ouvrir le fichier SQL avec un éditeur puis de copier le contenu du code SQL dans SQL Serveur puis l'exécuter.

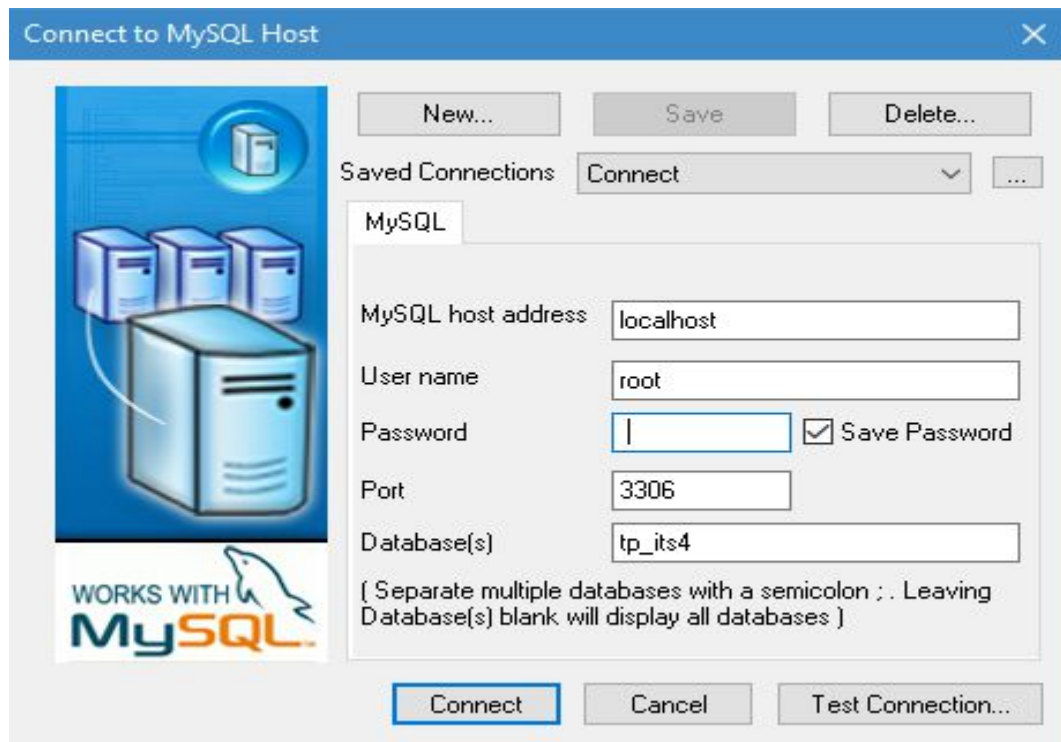
Si tout se passe bien, vous devez obtenir la figure suivante :



Après cette étape, il faut d'abord se connecter à la base de données de l'entreprise commerciale nommée « tp_its4_2018 » dans PhpMyAdmin ou SQLyog avec WampServer. Dans le cas de ce projet, c'est SQLyog qui a été utilisé pour faire quelques manipulations dans la base à savoir les vues créées afin de les utiliser dans Talend Open Studio. Il faut aussi savoir que ces dernières connexions vont être utiles pour charger les données dans Talend Studio.

II. *Chargement des données dans Talend Studio*

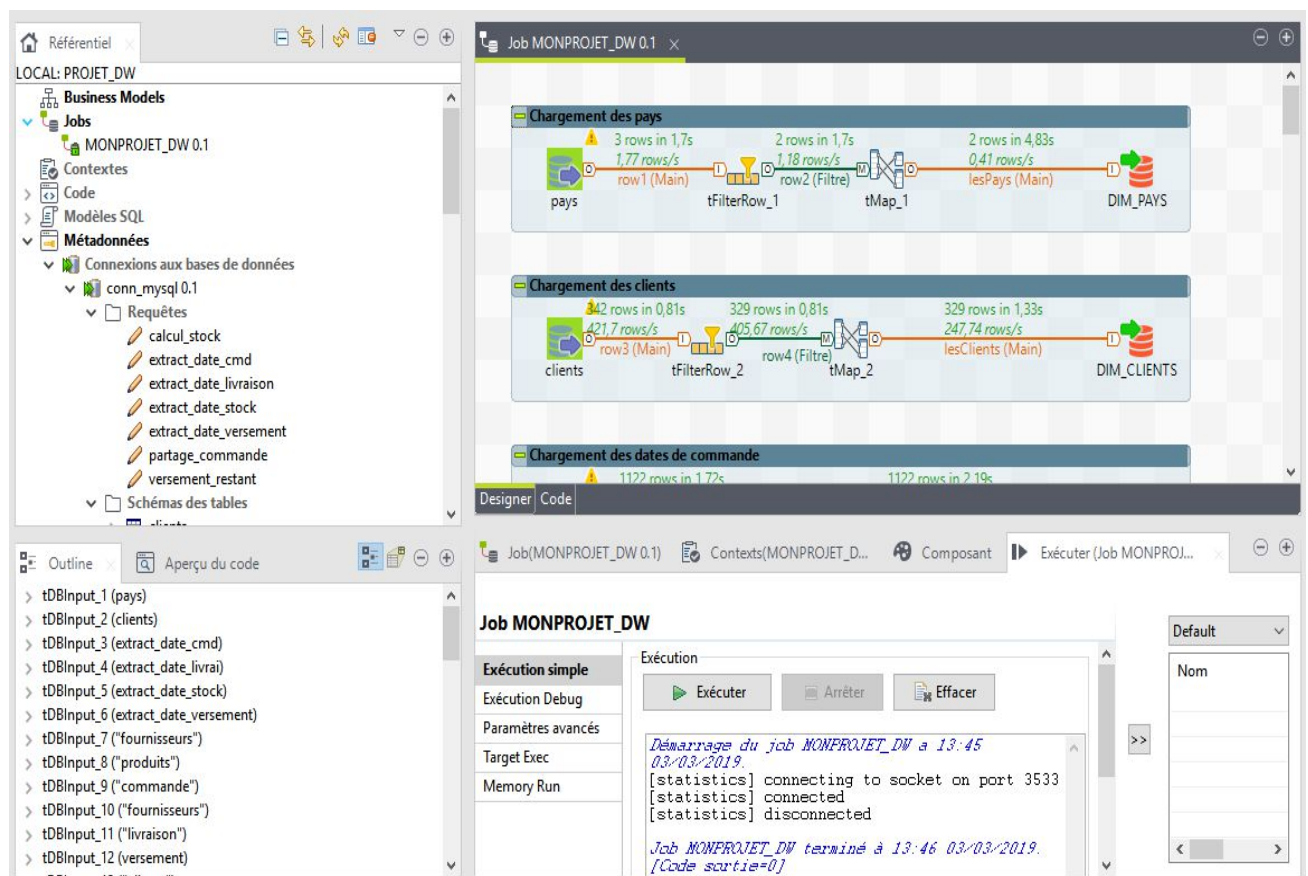
Comme il est dit plus haut, avant de créer le projet sous Talend, il d'abord conseiller de mettre en garde les deux connexions à savoir la connexion dans SQL Serveur et la connexion dans SQLyog. La connexion dans MS SQL Serveur a été déjà expliquée, reste à montrer la connexion dans SQLyog. Pour faire cette dernière, il faut ouvrir SQLyog et paramétrer la connexion comme suit en cas d'utilisation de ce projet :



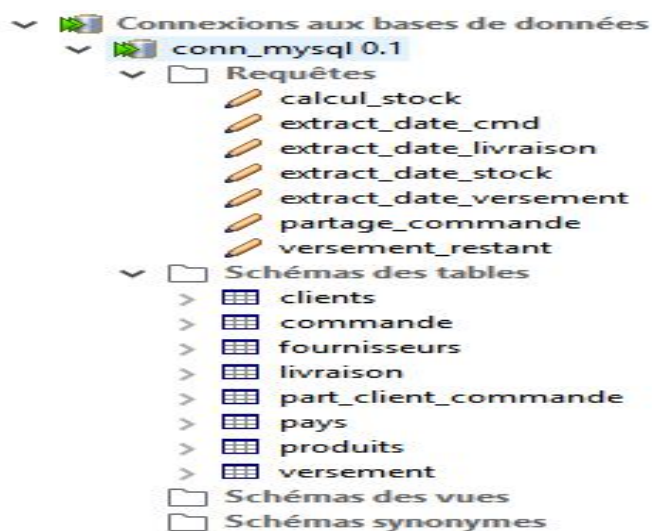
Ensuite il suffit juste de cliquer sur « Connect » pour pouvoir se connecter à la base de données MySQL tout en sachant qu'il faut nécessairement démarrer WampServer pour que ça marche. Il est aussi possible d'utiliser PhpMyAdmin pour faire le même travail. Après que la connexion s'est établie avec succès, il est possible d'aller ouvrir Talend Studio pour continuer le reste du travail à savoir : charger les données vers l'entrepôt de données.

Maintenant après avoir ouvert Talend, pour l'utilisateur qui décide d'utiliser ce projet, il va devoir prendre le dossier nommé « PROJET_DW » dans le dossier « Projet de Chargement TALEND » et le mettre dans le Workspace du dossier d'installation de Talend comme il est dit plus haut. Ainsi après cela, une fois que Talend sera ouvert, il sera possible de voir le projet sur les projets existants et de le charger directement. Sinon l'utilisateur doit créer un nouveau projet et constituer lui-même le job et les sous job nécessaires pour répondre aux questions posées par le décideur.

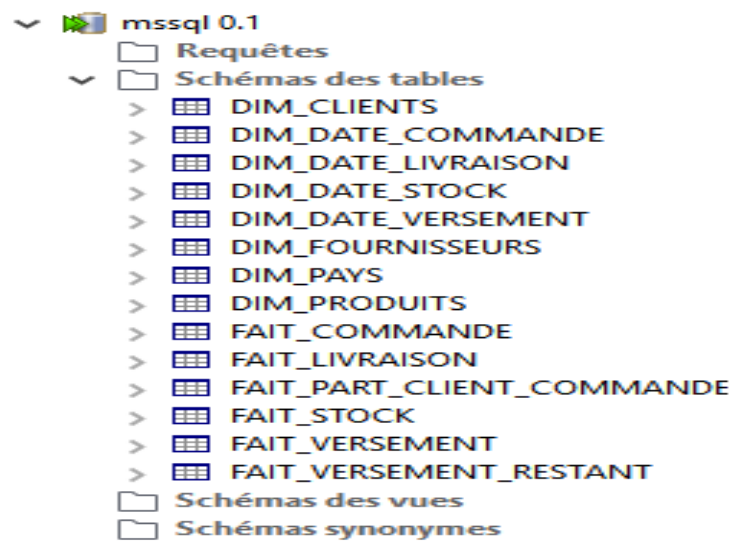
Si tout a été bien fait, le projet de chargement sous Talend prendra la forme suivante :



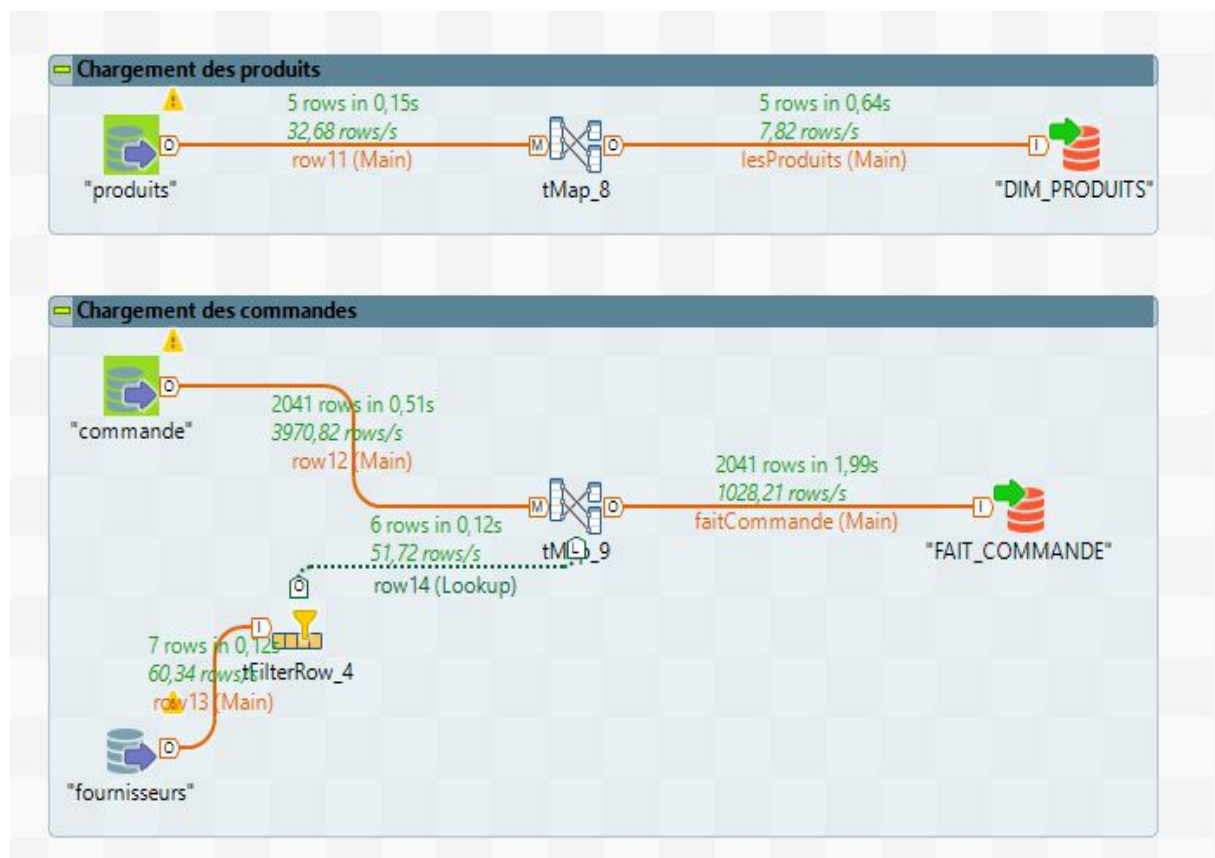
Cette figure est l'environnement obtenu après l'exécution des tous les sous job sans erreur. Autrement dit que toutes les données de l'entrepôt ont été bien chargées. Dans cet environnement, il y a deux connexions qui sont créées comme il est dit plus haut. Après avoir établie ces deux connexions, il va falloir charger les schémas des bases de données. En ce qui concerne la base MySQL, il faut seulement charger les tables dont on a besoin lors du chargement comme le montre la figure suivante :



Après cela, il faut refaire la même pour l'autre base en important toutes les tables comme le montre la figure suivante :



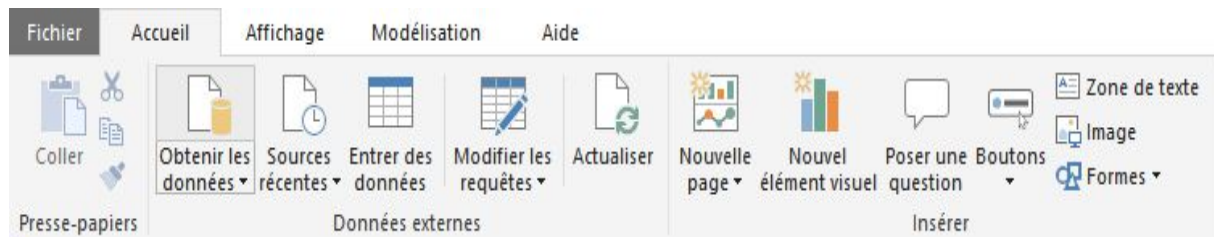
Après ces deux étapes, il faut commencer à construire les sous jobs. La figure suivante montre l'exemple de quelques sous jobs créés pour charger les données dans l'entrepôt de données :



Ces sous jobs ci-dessus représentent respectivement les « chargement des produits » et les « chargement des commandes (quantités commandés, remise) ». C'est le tMap qui va permettre de faire le chargement et le tFilterRow permet de filtrer les colonnes supprimées.

III. Création des rapports sur Power BI Desktop

Une fois que tout s'est bien déroulé et que les données ont été bien chargées dans l'entrepôt de données, le reste du travail se fait dans Power BI Desktop. Pour ce faire il faut ouvrir Power BI Desktop, ensuite créer un nouveau projet Power BI et le sauvegarder. Maintenant dans le menu de Power BI, il faut cliquer sur « obtenir les données » puis cliquer sur « SQL Serveur » comme le montre la figure suivante :



En utilisant les mêmes paramètres de connexion, il est possible de se connecter à l'entrepôt de données déjà alimenter. Le travail qui reste consiste à calculer certains indicateurs comme le « restant des versements par clients » et de construire les rapports sur les indicateurs demandés par le décideur. La figure suivante montre un exemple de rapport créé sous Power BI :

