



REPUBLIQUE DU SENEGAL



Un Peuple - Un But - Une Foi

MINISTERE DE L'ECONOMIE DU PLAN ET DE LA COOPÉRATION

AGENCE NATIONALE DE LA STATISTIQUE ET DE LA
DÉMOGRAPHIE

ECOLE NATIONALE DE LA STATISTIQUE ET DE L'ANALYSE
ÉCONOMIQUE



Projet TP Data Warehouse, ITS4

MODELISATION D'UN ENTREPOT DE DONNEES, CHARGEMENT
DE DONNEES ET CREATION DE TABLEAUX DE BORD

Rédigé par :

Danis Rikel JIOGUE TAMATIO

Supervisé par :

M. Ibrahima DIOP

Chargé du cours de TP DataWarehouse à l'ENSAE - Dakar

Mars 2020



Table des matières

1	Création de l'entrepôt de données	2
1.1	Chargement de la base de donnée sql server via l'ETL Talend Of Studio	2
1.1.1	Paramétrage de la Construction de l'ETL	2
2	Mise en place de la solution décisionnelle via l'outil POWER BI	5

Chapitre 1

Création de l'entrepôt de données

* **NB :** Dans le document j'utiliserai le premier pronom personnel du pluriel en l'occurrence le nous ceci juste dans l'optique de continuer avec l'esprit de travail en groupe qui est fort demandé dans le milieu professionnel.

Pour construire notre entrepôt de donnée, nous allons tout abord identifier les sources d'informations autrement dit les inputs. ces derniers sont ainsi constitués :

- **Le fichier de donnée Excel :** les données contenues dans le tableur Excel sont composées des informations relatives à la livraison et au versement ;
- **La base de données MySQL :** cette base de donnée contient l'ensemble des information de l'entreprise relatives à ses activités.

Ces deux bases de données constitues les inputs pour la mise en place de notre solution décisionnelle.

Comme output, nous avons la Base de donnée SQL server qui sera construite grâce au script *sql server* obtenu du logiciel BD Design.

A ce niveau, nous disposons des différentes bases de données suivantes avec leur statut :

➡ **Input :**

- **fichier Excel**
- **Base de données My SQL "its4prod2020"**

➡ **Output :**

- **Base de données sql server "GestEntCom"**

1.1 Chargement de la base de donnée sql server via l'ETL Talend Of Studio

Afin de charger notre entrepôt de donnée, nous allons tout d'abord paramétrer les fichiers sources (input) le l'output pour que la compilation se fasse en toute aisance. Le procédé de paramétrage se fait comme suit :

1.1.1 Paramétrage de la Construction de l'ETL

1. Configuration du fichier Excel :

Nous avons opté pour sectionner la base en deux. Autrement dit, pour l'importation

des données dudit fichier, nous avons créé 2 *tfileinputexcel* (un livraison et l'autre versement). L'avantage de ce disséquement est qu'il nous permettra de debugger plus vite notre application lors de l'exécution.

2. Configuration de connexion MySQL :

La configuration à la base de donnée via *tDBInput(MySQL)* s'est faite sans problème.

3. Configuration de la connexion SQL SERVER :

Elle s'est faite via le *tDBOutput(Microsoft SQL Server)*. Cependant, afin de s'assurer de la vraisemblance des données remontées, il a été nécessaire de d'ajouter en un filtre sur les inputs. Ainsi, nous avons utilisé un *tfilter* pour s'assurer que les informations remontées ne sont pas supprimé. De ce fait la condition utilisée est la suivante :

Opérateur logique utilisé pour combiner des conditions OR

Conditions	Colonne d'entrée	Fonction	Opérateur	Valeur
	est_supprime	Vide	Vaut	'0'
	est_supprime	Vide	Vaut	null

Aussi, dans les options de création de table, nous avons choisi :

Action sur la table Créer la table si elle n'existe pas ☐ Activer les insertions Identity Action sur les données Update ou insert

Après toutes ces manipulations préalables, le un **tmap** est utilisé pour synchroniser et vérifier la cohérences des données remontées. Cependant, lors de l'exécution, nous sommes confrontés à la difficulté suivante :

```
Demarrage du job Job_DanisJIOGUE a 13:49 16/03/2020.
[statistics] connecting to socket on port 3743
[statistics] connected
Mon Mar 16 13:49:11 GMT 2020 WARN: Establishing SSL connection without server's identity verificat
[statistics] disconnected
```

Après plusieurs recherches, nous avons obtenu la solution suivante (surligner dans la sortie) :

Chaîne de caractères de connexion :mysql://localhost:3306/its4prod2020?noDatetimeStringSync=true&useSSL=false

Lors du chargement des tables, il est important de noter que c'est processus séquentiel. Ainsi, à fin de respecter ce pré-ordre, nous utiliserons différents job :

1. **Job_DanisJIOGUE** : à pour but de procéder aux chargement des tables n'ayant de clé étrangère c'est-à-dire les **tables de dimensions** à savoir : TypeClient, Conversion, Monnaie, etc.. Ceci permet de pallier à l'erreur suivante :

```
[statistics] connecting to socket on port 3964
[statistics] connected
L'instruction INSERT est en conflit avec la contrainte FOREIGN KEY "FK_acl_
```

Cette erreur survient principalement lors du chargement d'une table ayant une clé étrangère qui n'existe pas (n'a pas encore été chargée).

2. **Job_Table_dep1** : permet le chargement des tables ayant une ou deux clés étrangères pour lesquelles les tables ont été précédemment chargées (table `acl_client`, `acl_commande`). Ces tables sont généralement reconnues sous l'appellation **table de fait**.
3. **Job_Table_dep2** : utiliser pour le chargement des tables ayant un foreign key **idaclclient**. En effet, la table client qui est une table de fait, se charge dans le second job il faut un autre job qui s'exécute après lui pour prendre en compte.
4. **Job_chargement** : Utiliser pour mettre de l'ordre dans l'exécution des jobs précédents. Ainsi, il utilise les **TRunjob** avec l'option *trigger*, **Onsubjob**. Il se présente ainsi :



A ce niveau, notre entrepôt de donnée est fin prêt, nous allons maintenant passer à la phase suivante à savoir mise en place d'une solution décisionnelle pour répondre aux requêtes du gérant de l'entreprise.

Chapitre 2

Mise en place de la solution décisionnelle via l'outil POWER BI

Pour la mise en place de la solution décisionnelle, l'outil de publication utilisé est **Power bi**.

Power bi est un ensemble de services logiciels, applications et de connecteurs qui œuvrent ensemble pour transformer les sources de données disparates. Il permet de se connecter facilement aux différentes sources de données, de visualiser (ou découvrir) les informations importantes et de les partager avec les personnes de votre choix. Ce logiciel présente diverses fonctionnalités qui seront présentées dans la suite avec une application dans notre cas pratique :

⇒ **Collecter, transformer et préparer les données :**

Cette phase est effectuée via l'éditeur **Power Query** de power bi. Dans notre cas d'application, nous fait appel à cette fonctionnalité pour la réalisation des tâches suivantes :

- **Suppression de l'attribut "est_supprimé"** dans toutes les entités. Cette action est motivée par le fait que cette attribut n'apporte aucune information pour notre jeux de donnée. Ceci permet entre autre de gagner en espace. Mais compte tenu de la faible taille de nos données, ceci n'a pas réellement d'avantage mais il est important de le garder en mémoire.
- **Transformation de l'attribut date_commande** : lors de l'importation de l'entrepôt de donnée, cet attribut de l'entité *acl_commande* est formaté en texte. Donc il est nécessaire de la re-convertir en format Date.
- **Transformation du format des attributs date_livraison, date_versement** : Le traitement est le même que celui précédemment décrit.

⇒ **Création du modèle de données et des indicateurs :**

Cette étape est matérialisée par la création d'un certain nombre d'indicateurs nécessaires pour la réalisation de la solution décisionnelle. Les indicateurs créés sont les suivants :

- **Prix total des commandes en Fcfa** : la formule de calcul de cet indicateur prend en input les attributs suivants : *id_acl_monnaie* ; *Taux1* ; *Taux2* ; *tot_prixm3*. La formule utilisée est donnée par la figure suivante :

```
tot_cmd(cfa) = IF([acl_commande][id_monnaie] = 1;acl_commande[prix_tot];(acl_commande[prix_tot]*acl_commande[taux_2])/acl_commande[taux_1])
```

- **Prix total des commandes en dalaxy** : la formule de calcul de cet indicateur prend les mêmes inputs que le précédent :

```
Prix_tot_dalaxy = IF([acl_commande[id_monnaie] = 2; acl_commande[prix_tot];(acl_commande[prix_tot]*acl_commande[taux_1])/acl_commande[taux_2])
```

- **Solde en Frcfa** : Obtenu comme étant la différence entre le prix total des livraisons en cfa et des versement en cfa.
- **Sole en Dalaxy** : Obtenu comme étant la différence entre le prix total des livraisons en dalaxy et des versement en dalaxy.
- ⇒ **Création des rapports et dashboards interactifs** : suivant les différents objets de visualisations et les mises en formes allouées, les différents graphiques ont été établis suivant nos appréciations.
- ⇒ **Partager et diffuser en mode web ou mobile** :

Un clicher de l’affichage mobile est illustrer par la figure suivante :

