



PROJET DECISIONNEL

Licence professionnelle :
métiers du décisionnel et de la
statistique

L'équipe Totally Data :
KUNTZ Aurélie
DUEZ Nicolas

Sommaire

Installation des logiciels requis, planification des objectifs, et mise en place d'un répertoire GITHUB	2
Installation de Prestashop	2
Expression des besoins et maquette.....	3
Création du MCD (Modèle Conceptuel des Données).....	5
Création du Data Warehouse	6
Gestion des flux des données	7
1. Création des différents STG (ou DSA : Data Staging Area)	8
2. Exécution des STG.....	9
3. STG vers Datawarehouse	10
Conception du tableau de bord	11
Importation des données (connexion à la BD)	11
Mise en place des liaisons entre les tables de notre modèle	13
Transformation des données (création de nouvelles colonnes, etc...).....	13
Réalisation du tableau de bord	14
Clients.....	14
Commandes	14
Produits	15
Vue d'ensemble	15

Installation des logiciels requis, planification des objectifs, et mise en place d'un répertoire GITHUB

L'installation des logiciels requis a été détaillée dans la document intitulée **“Documentation installation”**

Nous avons utilisé l'interface Trello afin de lister l'ensemble des tâches à réaliser. (<https://trello.com/b/DQMdzOxT/projetprestashopbest>).

Trello est un outil de gestion de projet en ligne, lancé en septembre 2011 et inspiré par la méthode Kanban de Toyota. Il repose sur une organisation des projets en planches listant des cartes, chacune représentant des tâches.



De plus, afin de conserver nos scripts et fichier, nous avons également utilisé GITHUB. Il s'agit d'un service web d'hébergement et de gestion de développement de logiciels.

(<https://github.com/totallydata/PRESTASHOP/>).

Installation de Prestashop

Pas de projet Totally Data sans l'installation d'une boutique e-commerce en local, factice.

Ainsi, afin de pouvoir tester et développer à partir de données, nous avons installé l'application Prestashop et l'avons hébergé sur un serveur local qu'est Wamp Server. WampServer est une plate-forme de développement Web sous Windows pour des applications Web dynamiques à l'aide du serveur Apache2, du langage de scripts PHP et d'une base de données MySQL.

Wamp Server possède également PHPMyAdmin pour gérer les bases de données.

PhpMyAdmin va nous permettre de nous assurer, facilement, de la bonne intégration des données factices rentrées dans Prestashop.

C'est également grâce que nous avons pu modifier les dates de commandes entre autre pour les besoins en rendu sur la visualisation finale.

Une fois cette première étape réalisé, nous allons créer de fausses commandes afin d'avoir une base avec une, faible, volumétrie mais permettant la suite des opérations. Les commandes et les clients ont été générés depuis Prestashop que nous avons fait fonctionner en local.

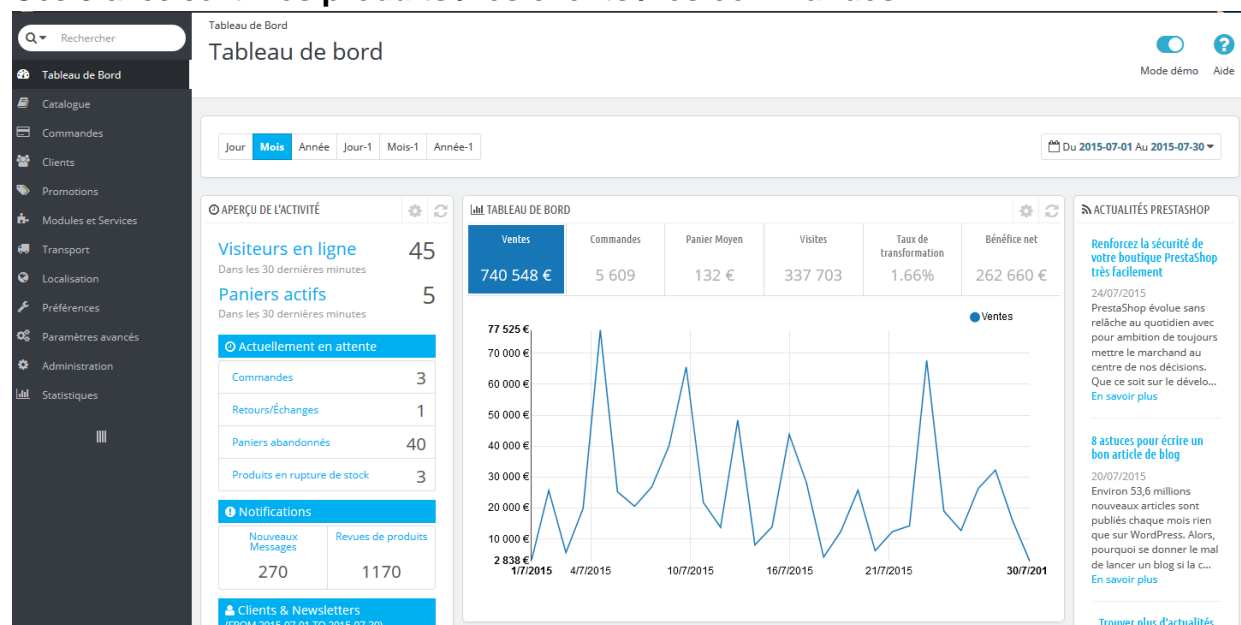


Logiciels utilisés lors de cette étape : Prestashop, Wamp, PhpMyAdmin

Expression des besoins et maquette

Nous avons, par la suite, réalisé une expression des besoins. Ce qui en est ressorti est un manque de visibilité sur prestashop. En effet, Prestashop contient quelques tableaux de bords permettant une vue simplifiée et rapide sur ses ventes. Cependant, ces tableaux de bords trop succincts ne correspondent pas à l'utilisation du client. Nous avons eu 3 grands axes à analyser et sur lequel il fallait pouvoir avoir vision tant macro que micro.

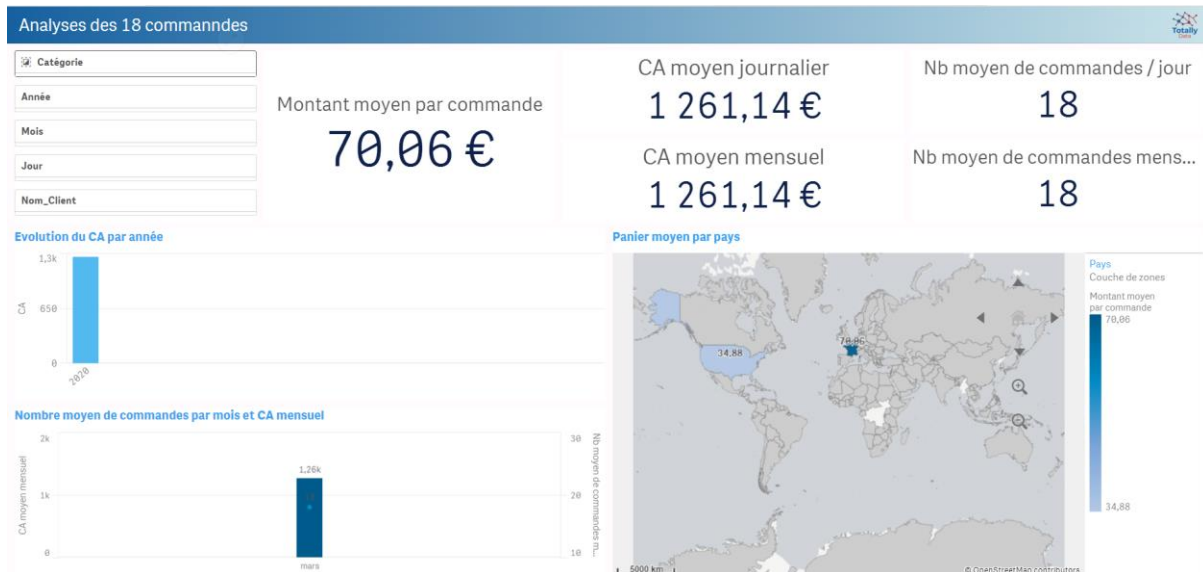
Ces 3 axes sont : **les produits / les clients / les commandes.**



Afin de valider l'expression de besoin, nous avons utilisé **Qlik Sense Desktop** pour concevoir une maquette permettant au client de visualiser, avec de fausses données, les possibilités.

Qlik Sense Desktop est une application bureau d'analyse de données, et de visualisation de données.

Cette maquette a permis aussi de se rendre compte des possibles difficultés sur les données à traiter lors de la création du datawarehouse (champs vides, mauvaise syntaxe du champs date....). Qlik Sense permet, sur une faible volumétrie, de remplacer un ETL en codant directement dans le chargement des modifications. Pour créer une maquette fonctionnelle et assez rapidement, Qlik était LA solution.



Logiciels utilisés lors de cette étape : Qlik Sense, Google Drive

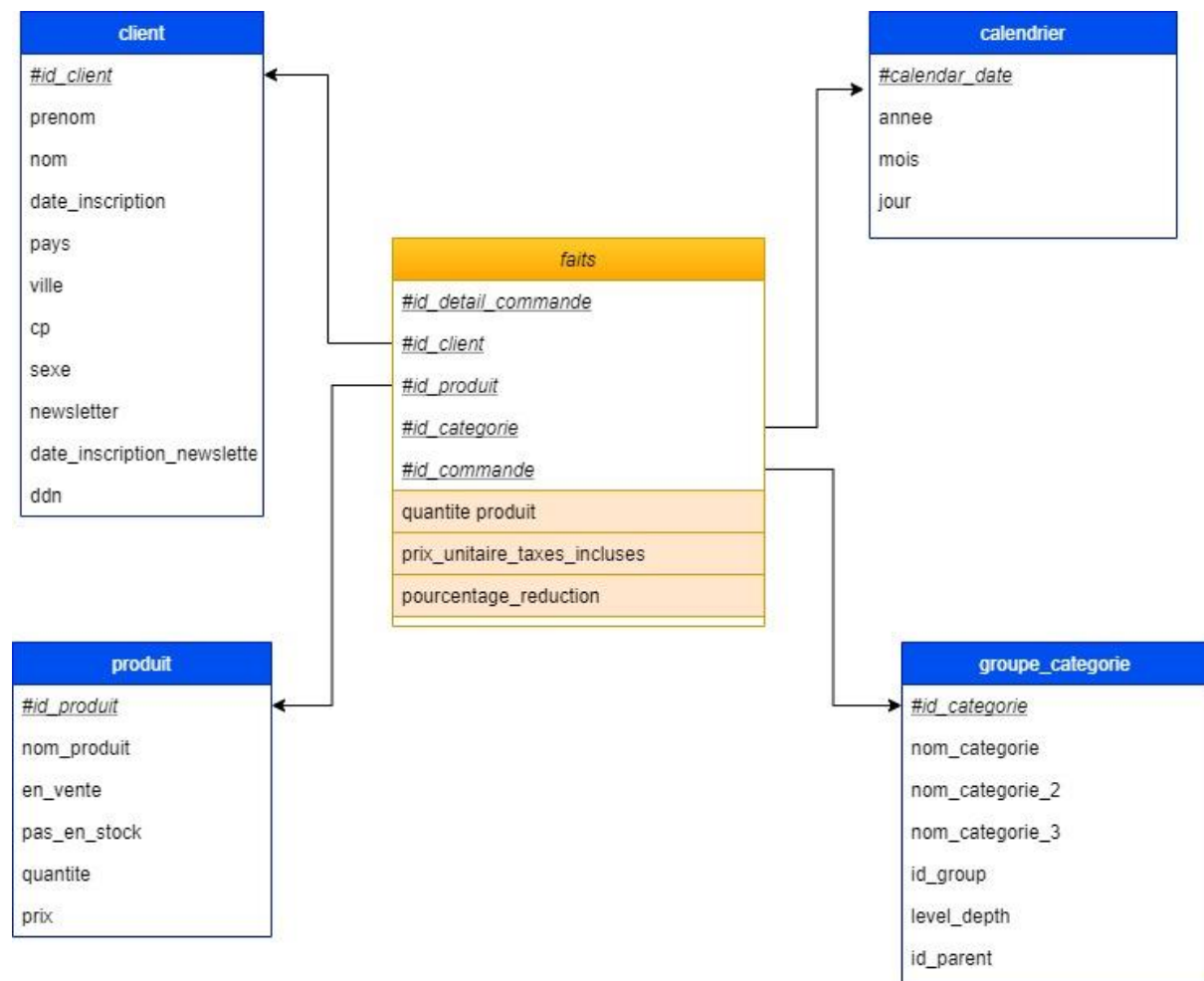
Création du MCD (Modèle Conceptuel des Données)

MCD est une abréviation qui signifie modèle conceptuel de données. Il s'agit d'une représentation logique de l'organisation des informations et de leurs relations.

La construction du modèle MCD s'avère nécessaire.

Pour cela, nous avons utilisé l'outil en ligne **Draw**, disponible à l'adresse suivante :

<https://app.diagrams.net/>



Dans le même temps, l'utilisation de Qlik Sense pour la maquette a permis, de sélectionner les tables et les champs nécessaires à la réalisation du tableau de bord.

Création du Data Warehouse

Les Data Warehouse ou entrepôts de données sont des bases de données permettant de stocker des données historiques structurées non volatiles orientées sujet afin de les analyser.



Logiciel utilisé : MySQLWorkbench

Nous avons choisi d'utiliser le SGBD **MYSQL WORKBENCH** pour la création du Datawarehouse. Il s'agit de la base de données qui sera utilisé pour la création du tableau de bord.

MySQL Workbench (anciennement MySQL administrator) est un logiciel de gestion et d'administration de bases de données MySQL créé en 2004. Via une interface graphique intuitive, il permet, entre autres, de créer, modifier ou supprimer des tables, des comptes utilisateurs, et d'effectuer toutes les opérations inhérentes à la gestion d'une base de données. Pour ce faire, il doit être connecté à un serveur MySQL.

L'ensemble des scripts de création de table se trouvent dans notre répertoire GITHUB accessible à l'adresse suivante :

<https://github.com/totallydata/PRESTASHOP/>.

Cette étape de création est indispensable. En effet, c'est l'endroit où les flux de données termineront leur course.

C'est également depuis cette interface que nous avons généré la table "calendrier" qui recense toutes les dates du 1er janvier 2000 à la date d'aujourd'hui, avec la date, l'année, le mois et le jour.

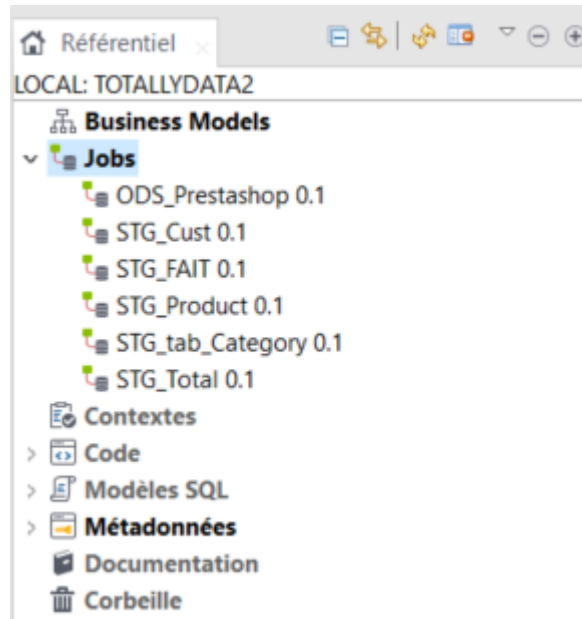
Cette solution nous a semblée plus pertinente que de générer le calendrier depuis Microsoft Power BI, car elle existe déjà et peut être utilisée directement par d'autres outils de visualisations.

	calendar_date	annee	mois	jour
►	2000-01-01	NULL	NULL	NULL
	2000-01-02	2000	1	2
	2000-01-03	2000	1	3
	2000-01-04	2000	1	4
	2000-01-05	2000	1	5
	2000-01-06	2000	1	6
	2000-01-07	2000	1	7
	2000-01-08	2000	1	8
	2000-01-09	2000	1	9
	2000-01-10	2000	1	10
	2000-01-11	2000	1	11
	2000-01-12	2000	1	12
	2000-01-13	2000	1	13
	2000-01-14	2000	1	14
	2000-01-15	2000	1	15
	2000-01-16	2000	1	16
	2000-01-17	2000	1	17
	2000-01-18	2000	1	18

Gestion des flux des données



ETL utilisé : Talend



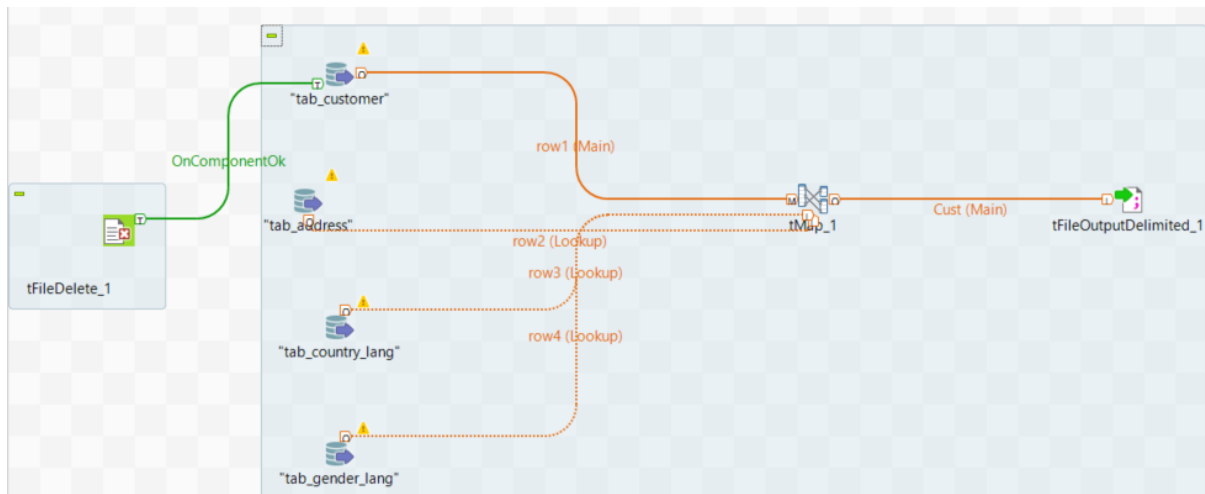
1. Création des différents STG (ou DSA : Data Staging Area)

Exemple pour les clients

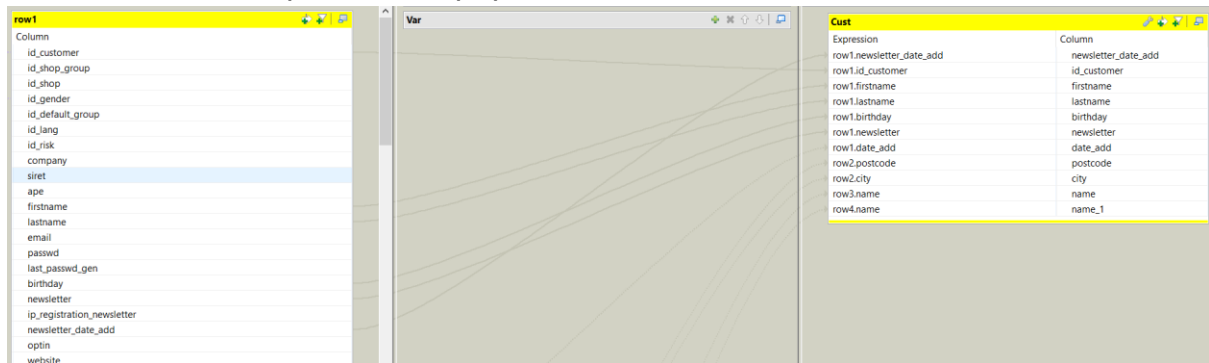
Le DSA est un espace de stockage temporaire. Son contenu peut être vidé après chaque chargement du Datawarehouse. Contrairement à l'ODS les données sont stockées telles quelles, sans nettoyage ni contrôles.

Talend offre la possibilité de glisser différents composants au sein d'un « job », permettant de réaliser des actions diverses.

Le composant «tFileDelete» permet de remplacer le fichier à chaque exécution. Le composant «tDBInput» nous permet de récupérer les différentes tables nécessaires à la création du STG. Ensuite, le composant «TMap» nous permet de relier les tables entre elles, de sélectionner les champs qui nous intéressent, et de réaliser des conditions afin d'obtenir en sortie une table unique.



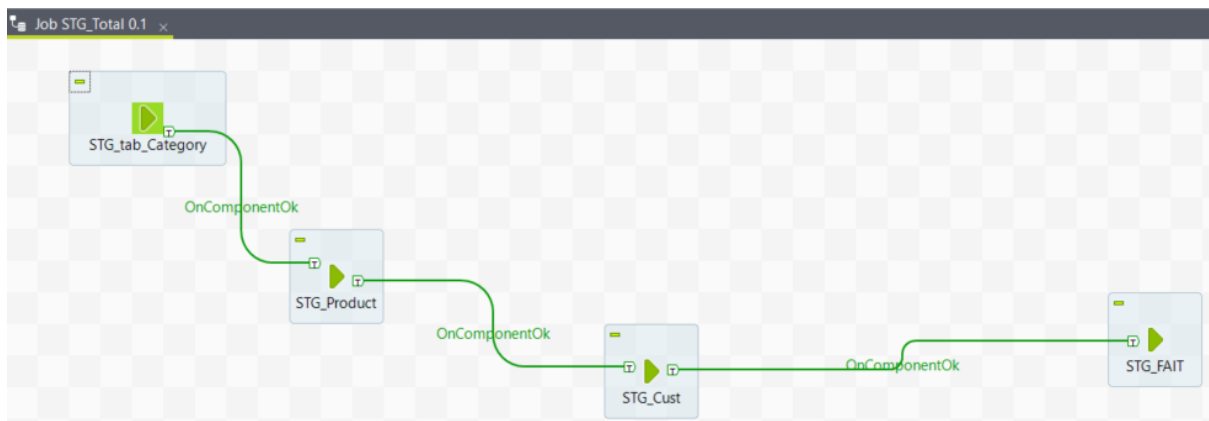
Paramètres du composant “tMap” pour le STG client.



Enfin, la table en sortie du “tMap” est stockée dans un fichier délimité localement sur le pc.

2. Exécution des STG

L'utilisation du composant “tRunJob” nous permet d'exécuter successivement les différents Jobs liés à la création des différents STG.



3. STG vers Datawarehouse



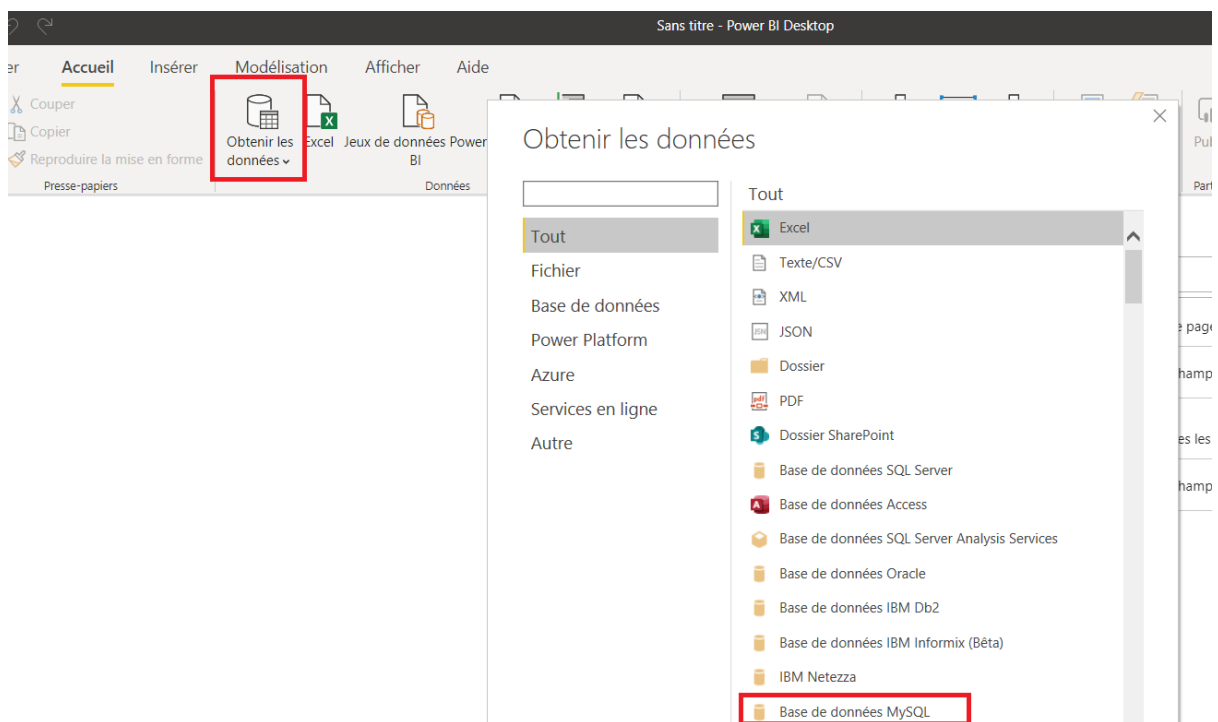
Il s'agit ensuite de récupérer les STG créés précédemment afin d'envoyer les données dans le DataWarehouse "dw_totallydata2" grâce au composant "tDBOutput".

Conception du tableau de bord

Logiciel utilisé : Power BI



Importation des données (connexion à la BD)



Base de données MySQL

Serveur

localhost:3306

Base de données

dw_totallydata2

Options avancées

Délai de commande en minutes (facultatif)

Instruction SQL (facultatif, nécessite une base de données)

☒ Inclure des colonnes de relation

☐ Naviguer avec la hiérarchie complète

OK

Annuler

Navigateur

Options d'affichage

localhost:3306: dw_totallydata2 [5]

☒ dw_totallydata2.calendrier

☒ dw_totallydata2.client

☒ dw_totallydata2.faits

☒ dw_totallydata2.groupe_categorie

☒ dw_totallydata2.produit

dw_totallydata2.calendrier

Aperçu téléchargé le vendredi 26 juin 2020

calendar_date	annee	mois	jour
01/01/2000	null	null	null
02/01/2000	2000	1	2
03/01/2000	2000	1	3
04/01/2000	2000	1	4
05/01/2000	2000	1	5
06/01/2000	2000	1	6
07/01/2000	2000	1	7
08/01/2000	2000	1	8
09/01/2000	2000	1	9
10/01/2000	2000	1	10
11/01/2000	2000	1	11
12/01/2000	2000	1	12
13/01/2000	2000	1	13
14/01/2000	2000	1	14
15/01/2000	2000	1	15
16/01/2000	2000	1	16
17/01/2000	2000	1	17
18/01/2000	2000	1	18
19/01/2000	2000	1	19
20/01/2000	2000	1	20
21/01/2000	2000	1	21
22/01/2000	2000	1	22
23/01/2000	2000	1	23

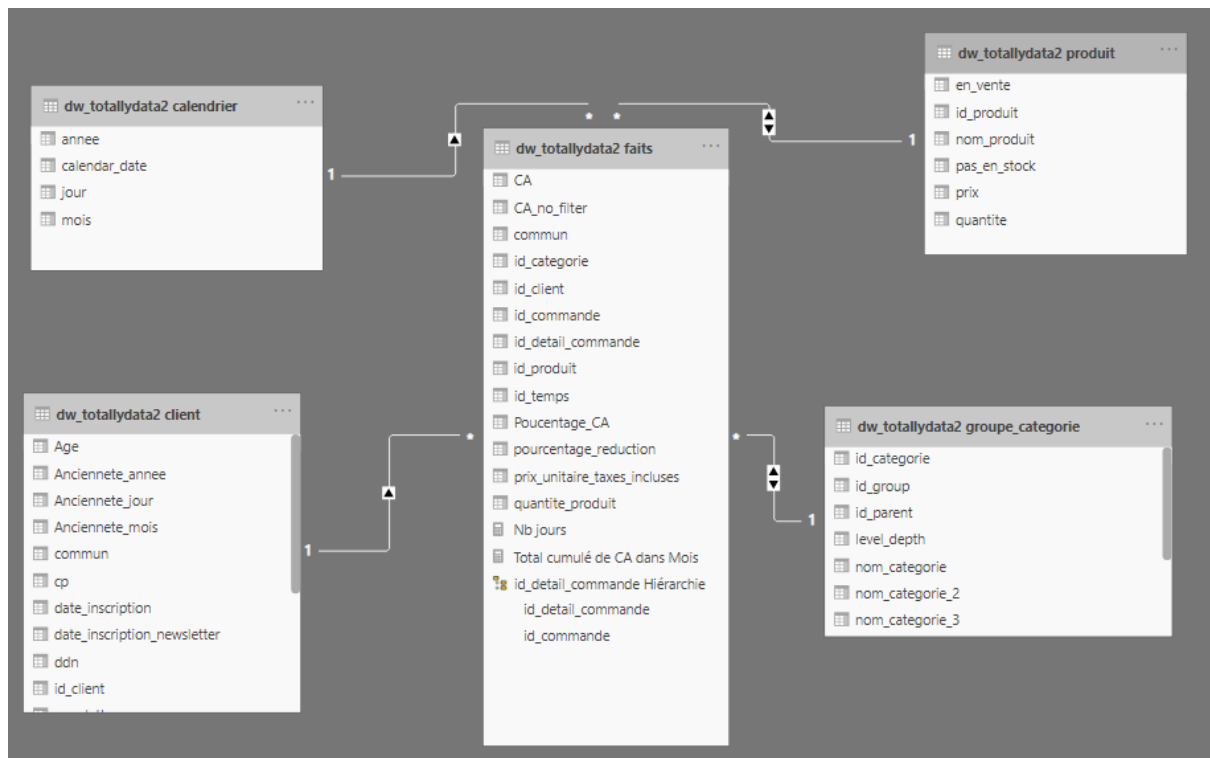
Sélectionner les tables associées

Charger

Transformer les données

Annuler

Mise en place des liaisons entre les tables de notre modèle



Transformation des données (création de nouvelles colonnes, etc...)

Création de champs calculés dans l'onglet “**données**” tels que :

- l'âge
- l'ancienneté
- le CA (prix * quantité * (1-remise))

A l'aide du langage DAX.

Exemple pour l'âge :

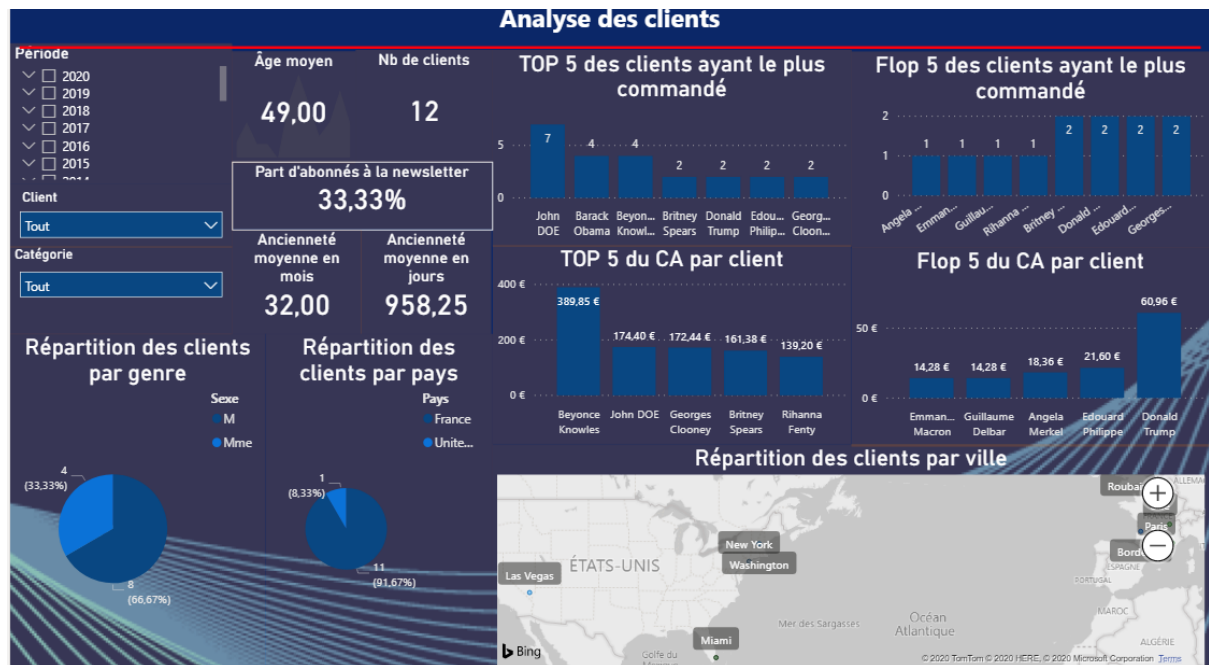
```
1 Age = DATEDIFF('dw_totallydata2 client'[ddn], TODAY(), YEAR)
```

ddn	Nom_prenom	Age	commun	Anciennete_annee	Anciennete_jour	Anciennete_mois
10/04/1980 00:00:00	Anonymous Anonymous	40	1	0	102	4
15/01/1970 00:00:00	John DOE	50	1	6	2294	76
04/09/1981 00:00:00	Beyonce Knowles	39	1	5	1929	64
02/10/1981 00:00:00	Britney Spears	39	1	5	1929	64
20/02/1988 00:00:00	Rihanna Fenty	32	1	4	1563	52
04/08/1961 00:00:00	Barack Obama	59	1	3	1198	40
14/06/1946 00:00:00	Donald Trump	74	1	2	833	28
06/05/1961 00:00:00	Georges Clooney	59	1	2	833	28
21/12/1977 00:00:00	Emmanuel Macron	43	1	1	467	16
28/11/1970 00:00:00	Edouard Philippe	50	1	0	101	4
17/07/1954 00:00:00	Angela Merkel	66	1	0	101	4
17/05/1971 00:00:00	Guillaume Delbar	49	1	0	101	4

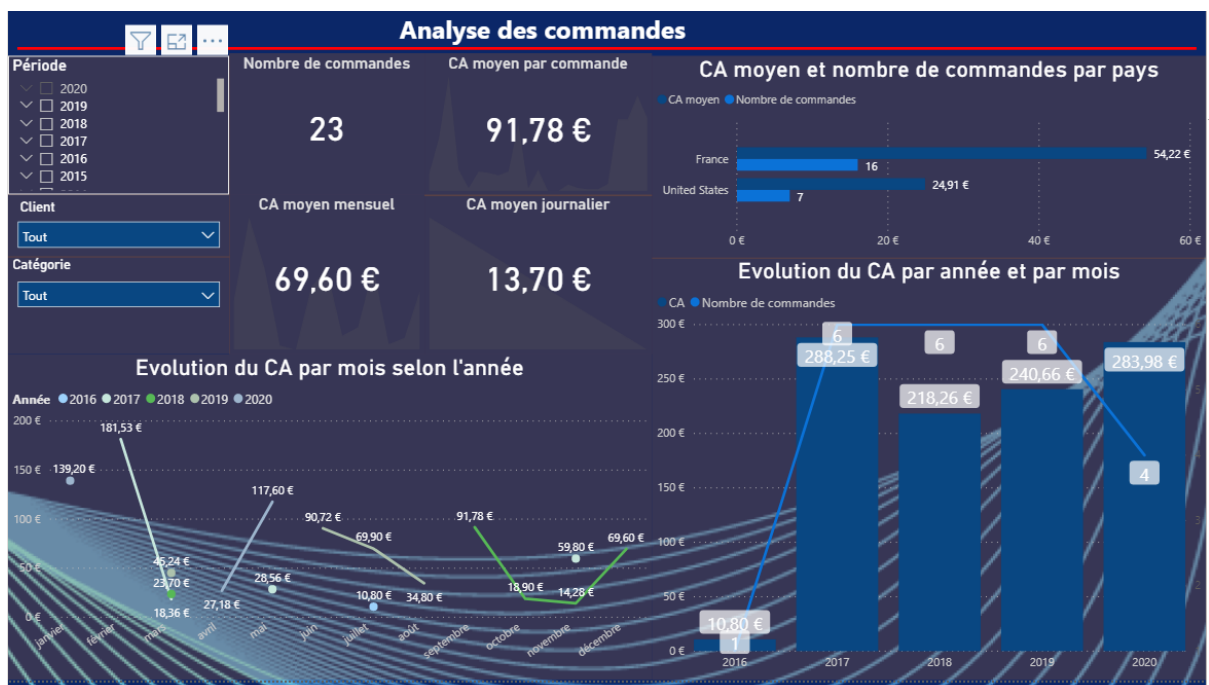
La création des champs calculés est également possible dans l'onglet “**Transformer les données**”

Réalisation du tableau de bord

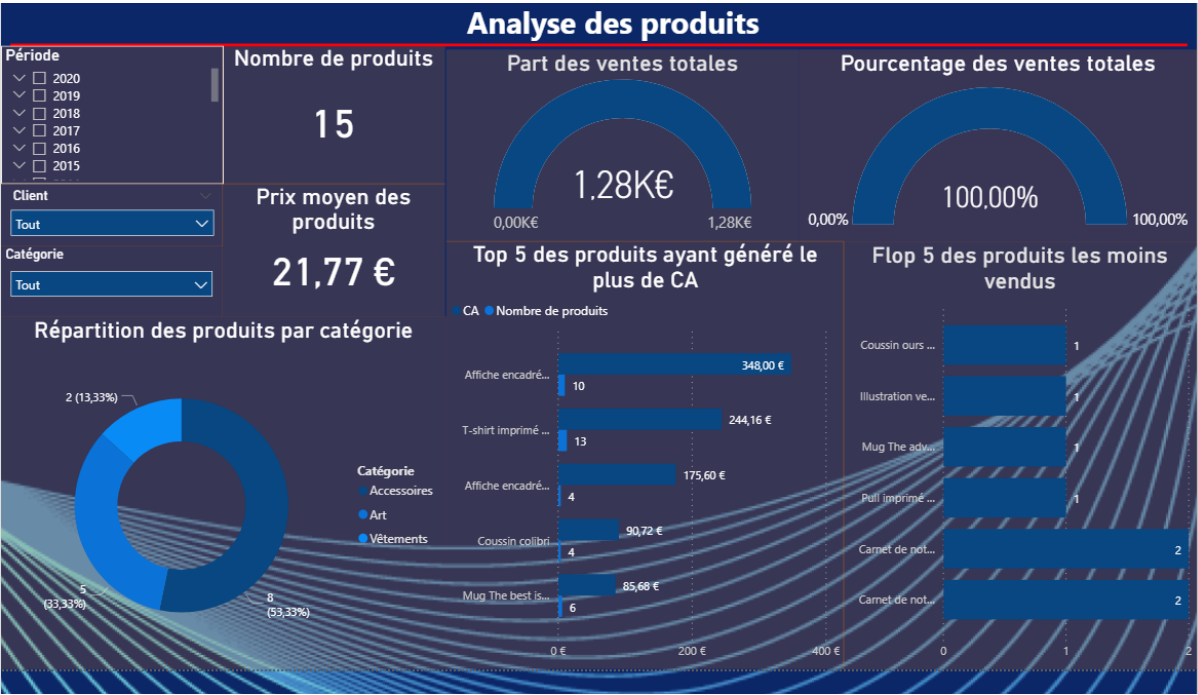
Clients



Commandes



Produits



Vue d'ensemble

