

Cas pratique transverse Qlik et Talend

Comment tirer profit au mieux des outils pour améliorer mon activité ?

Perolle Maxim - mperolle@acssi.fr
Aurélien Thève - atheve@acssi.fr

Sommaire

Sommaire.....	2
Contexte.....	4
Projet de mise en place d'une base de données de ventes sur Talend.....	5
Modèle de données cible.....	5
Fichier source.....	6
Taux de changes.....	7
Attendus en sortie. Indices.....	7
Fichier Client. Indices.....	8
Produit. Indices.....	8
CommandeLIG. Indices.....	9
Pour aller plus loin (à faire après la partie Qlik). Indices.....	10
Création d'une application Qlik à partir des données Talend.....	11
Chargement des données. Indices.....	11
Dashboard. Indices.....	11
Focus France.....	11
Un comparatif des produits. Indices.....	12
Détails commande. Indices.....	12
Indices.....	14
Import de la table source. Sujet.....	14
Fichier client. Sujet.....	14
Fichier Produit. Sujet.....	14
Ajout de la famille.....	14
Ajout de la gamme.....	15
Suppression des produits invendus.....	15
Fichier CommandeLIG. Sujet.....	15
Récupération de la devise.....	16
Ajout de la devise.....	16
Séparateur virgule.....	16
Pour aller plus loin. Sujet.....	16
Données aberrantes.....	16
Avant de lancer le flux.....	17
Fichiers temporaires ou suppression de fichiers.....	17

Chargement des données. Sujet	19
Dashboard. Sujet	20
Chiffre d'affaires.....	20
Focus France. Sujet.....	20
Comparatif produit. Sujet.....	20
Différence de filtre entre partie gauche et droite.....	20
Graphique lors du survol	20
Echelle commune	21
Couleur la plus vendue.....	21
Détails commande. Sujet.....	21

Contexte

Créée en 2010, la société Cars & Co' est un revendeur international de voiture. Cars & Co' souhaite implémenter un outil de business intelligence capable d'analyser les statistiques de ventes, de coûts, de marges et d'expéditions.

Ce projet est conçu pour évaluer vos connaissances sur les outils Talend et Qlik Sense. Il vous sera demandé de lier les deux outils pour créer un flux de données Talend, puis une application de restitution sur Qlik Sense. Pour cela, vous avez à disposition un fichier de données regroupant l'historique du groupe. De plus, un dictionnaire de données est disponible pour vous aider à prendre la donnée en main.

Le groupe réalise depuis le 2 janvier 2020 des ventes sur 3 pays (France, Etats-Unis et Japon) pour une centaine de clients différents. Ces ventes concernent 13 modèles de voitures, proposés en 8 coloris, chacun lié à une gamme et une famille produit.

Des indices sont disponibles pour certaines questions. Ils sont représentés par les mots "Indices" et "Sujet" et peuvent vous éviter de rester bloquer. Ces indices vous aiguillent sur une façon (parmi d'autres) de répondre à chaque problématique. Pour faire apparaître l'indice, copiez-le sur un bloc-notes.

Projet de mise en place d'une base de données de ventes sur Talend

Modèle de données cible.

Le modèle de données à obtenir à la fin de cette partie sur Talend est le suivant :

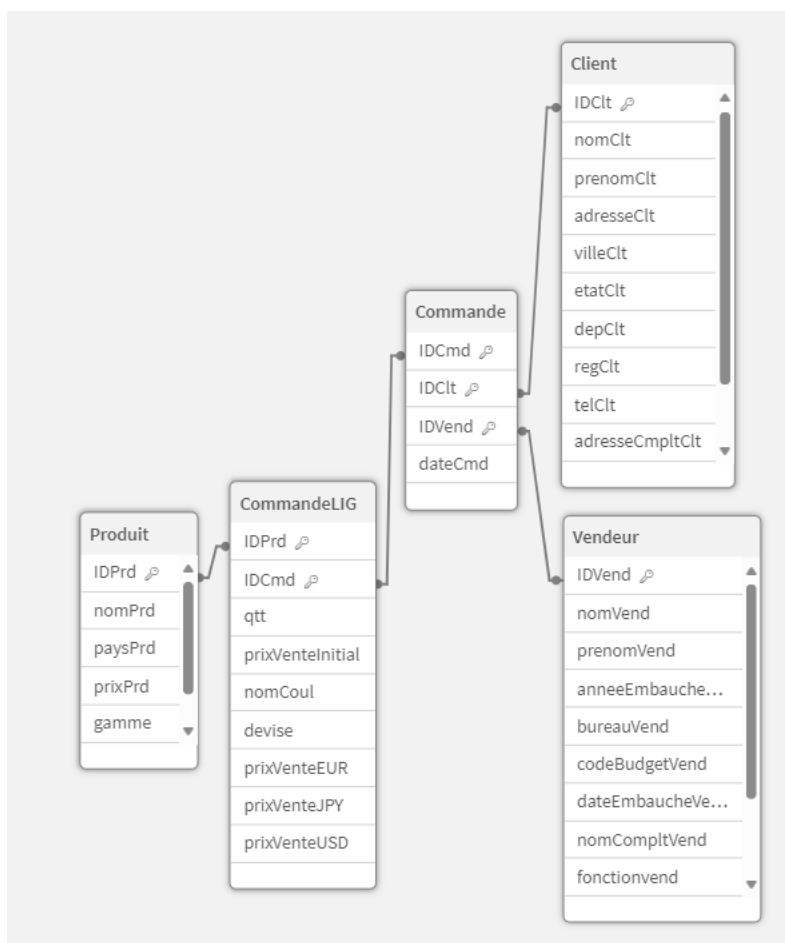


Figure 1 : modèle de données issue de Qlik

Les données principales sont les données de la table Commande, ce sont elles qui font le lien avec toutes les autres. Rattachées d'après l'IDCmd, on retrouve les lignes de commandes, liées au produit. On peut avoir pour une même commande plusieurs produits, dans plusieurs quantités et couleurs différentes. Chaque produit appartient à une gamme (luxe, grand luxe, 1er prix...) et une famille (sport, 4x4, citadine...). On retrouve aussi des informations sur les clients (notamment géographiques) et sur le vendeur.

Fichier source

Pour arriver à ce résultat, nous avons à disposition un seul fichier source. Il contient toutes les données en lien avec les tables ci-dessus. D'autres tables sont présentes pour enrichir les tables du modèle (visibles dans le dictionnaire de données). Ce fichier a la forme suivante :

```
"200","41","Ashkenaz","Mike","2000","01","52-1","11/30/2000","Ashkenaz Mike","Directeur Financier","-", "42","49500.00"
"200","42","Ranieri","Richard","1999","02","53-2","1/7/1999","Ranieri Richard","Directeur Financier","-", "23","49500.00"
"200","43","Richards","Sabrina","2003","03","54-3","4/3/2003","Richards Sabrina","Directeur Financier","-", "32","49500.00"
"200","44","Daw","Sehoon","2002","04","55-4","11/30/2002","Daw Sehoon","Directeur Financier","-", "37","49500.00"
"200","45","Lind","Stefan","1999","05","161-5","12/28/1999","Lind Stefan","Directeur Financier","-", "27","49500.00"
"300","1000","51","13","11/08/2020","","","","","","","",""
"300","1001","52","2","19/04/2021","","","","","","","",""
"300","1002","53","1","23/02/2022","","","","","","","",""
"300","1003","54","20","02/05/2021","","","","","","","",""
"300","1004","52","42","05/08/2022","","","","","","","",""
```

Figure 2 : lignes de vendeur et de commande du fichier MotorsTL

Pour les identifier, chaque ligne contient le champ typeLigne (première colonne), qui prend les valeurs suivantes :

typeLigne	tableCorrespondante
100	Client
200	Vendeur
300	Commande
301	Commande_LIG
302	Couleur
400	Produit
401	Famille
402	Gamme

Tableau 1 : correspondance des types de lignes

Cet extrait présente des données liées au client et à sa région. Il y a plus de sources données ici qu'on ne peut en voir dans le modèle final. Certaines devront donc être combinées.

Taux de changes

Vous avez accès à un fichier répertoriant pour le yen, le dollar et l'euro les taux de changes moyens annuels entre 2020 et 2023.

```
Annee;Devise;EUR;JPY;USD
2020;EUR;1;121;0,879
2020;JPY;0,0082;1;0,0097
2020;USD;0,879;107;1
2021;EUR;1;129;1,18
2021;JPY;0,0077;1;0,0091
2021;USD;0,846;110;1
2022;EUR;1;138;1,05
2022;JPY;0,0072;1;0,0076
2022;USD;0,951;131;1
2023;EUR;1;152;1,08
2023;JPY;0,0060;1;0,0071
2023;USD;0,924;140;1
```

Figure 3 : taux de change de 2020 à 2023 pour 3 devises

Pour chaque année, ce fichier nous donne la valeur des trois monnaies selon celle des deux autres. En 2020, un euro vaut ainsi 121 yens. Le séparateur est le point-virgule.

Attendus en sortie. Indices

A partir du fichier initial, seuls les 5 fichiers suivants devront être importés par la suite sur Qlik :

- Fichier Client ;
- Fichier Commande ;
- Fichier CommandeLIG ;
- Fichier Produit ;
- Fichier Vendeur.

Pour cela le fichier source devra être séparé selon le typeLigne. Il y a 8 tables qui se transformeront en 5 finales.

Fichier Client. Indices

La table Client contient les champs suivants :

Client	
Champs	Définition
IDCl	Identifiant client
nomCl	Nom du client
preNomCl	Prénom du client
adresseCl	Adresse cliente (numéro de voie et rue)
villeCl	Ville cliente
etatCl	Etat de résidence pour les clients américains
depCl	Département de résidence pour les clients français
telCl	Téléphone client (plusieurs formats)
regionCl	Région de résidence
paysCl	Pays de résidence

Il est demandé d'ajouter le champ "adresseCmpltCl", concaténation de adresseCl, villeCl et etatCl pour les villes des Etats-Unis. La région Île-de-France est mal reconnue. Elle s'affiche comme "ÃŽle-de-France" et doit donc être modifié.

Produit. Indices

La table Produit a la forme suivante :

Produit	
Champs	Définitions
IDPrd	Identifiant produit
nomPrd	Nom du produit
paysPrd	Pays de production
prixPrd	Prix unitaire
IDFam	Famille de produit

Pour cette table, l'IDFam devra être remplacé par le nom de la famille associée. Il est aussi nécessaire d'ajouter le nom de la gamme liée au produit. Un produit

appartient à une gamme si son prix de vente est supérieur au prix minimum et inférieur au prix maximum de cette gamme.

Les produits n'ayant jamais été vendus (ceux non communs entre la table produit et la table commandesLIG) ne devront pas être gardés dans la table finale.

CommandeLIG. Indices

La table commandeLIG a la forme suivante :

CommandeLIG	
Champs	Définitions
IDCmd	Identifiant commande
IDPrd	Identifiant produit
Qtt	Quantité de produit acheté
IDCoul	Identifiant couleur
prixVente	Prix de vente total de la ligne

Deux transformations seront nécessaires à la table de ligne de commande. Le nom de la couleur devra remplacer l'ID dans la table finale. De plus, les ventes ont été effectuées dans plusieurs pays et les prix de vente sont donc donnés en valeurs locales. La table finale devra comprendre trois colonnes, avec les prix en euro, dollar et yen à chaque ligne pour permettre une analyse fiable du chiffre d'affaires dans la partie suivante.

Ce travail peut être décomposé en deux étapes. La première serait de récupérer depuis la ligne de commande le pays associé (du client) ainsi que l'année de la commande. Avec le pays il est alors possible d'associer une devise. Les ventes au Japon seraient alors en yen, les ventes en France en euro, et les ventes aux Etats-Unis en dollar (USD). Il n'est pas demandé de gérer les autres cas.

Lorsque la devise et l'année de commande est commune à une ligne du fichier Devise, il est ensuite possible de récupérer le produit du prix de vente avec chaque taux de change.

Pour aller plus loin (à faire après la partie Qlik). Indices

- Certaines lignes de commande ont des données aberrantes, filtrez les pour ne garder que celles dont le prix de vente initial est supérieur à 10 000. Ce filtre peut se faire avec deux composants différents. Il serait aussi intéressant d'avoir de façon visuel (dans la console Talend ou dans un fichier externe) les lignes qui n'ont pas passé le filtre. Récupérez aussi le nombre de ligne rejetés.

- Dans le cas où le fichier de données serait manquant, Talend ne peut charger les données et stop son exécution. Vérifiez si le fichier est présent avant de lancer la partie principale du flux. Si ce n'est pas le cas, il faudrait alors renvoyer un message d'erreur.

- Le fichier de données concerne 10 tables mais seulement 5 sont demandés pour l'application Qlik. Les 5 restantes peuvent être créées et supprimées depuis Talend, il est aussi possible de créer des fichiers qui ne soient que temporaires dans lesquels on viendrait ajouter les données pour s'en servir le temps du flux.

- Les numéros de téléphone ne sont pas harmonisés entre les pays et plusieurs formats existent aussi pour les numéros français, rendant leur exploitation compliquée. Il serait utile de tous les harmoniser sur un format "+33 649767862"

Création d'une application Qlik à partir des données Talend

Chargement des données. Indices

Grâce à notre travail le chargement des données est très simple. Les 5 tables ne nécessitent pas de transformation pour être chargées. Pour la suite, nous aurons cependant besoin d'un calendrier qui devra contenir année, mois, trimestre (T1,T2,T3,T4) et annéeMois à partir de la date de commande.

La démarche ici est plus libre. 4 feuilles vous sont demandées mais leur contenu sera en partie décidé par vous. Soyez créatifs !

Vous avez accès aux prix dans trois devises, pouvoir passer de l'une à l'autre de façon libre serait intéressant. Cet exercice est à faire à la fin, en attendant, basez-vous sur le prix en dollar.

Dashboard. Indices

La première feuille demandée sera celle sur laquelle arrivera le client en ouvrant l'application. Elle doit contenir des indicateurs de haut niveau (indicateurs larges).

Elle devra entre autres contenir le chiffre d'affaires de l'année, du trimestre et du mois (3 indicateurs). Il devra être possible de choisir la période du KPI (choisir l'année, le trimestre ou le mois). Lorsque l'on sélectionne un mois, cela ne devra par contre pas avoir d'influence sur l'indicateur trimestriel et annuel. Si possible, ces indicateurs renverront les dernières données disponibles par défaut.

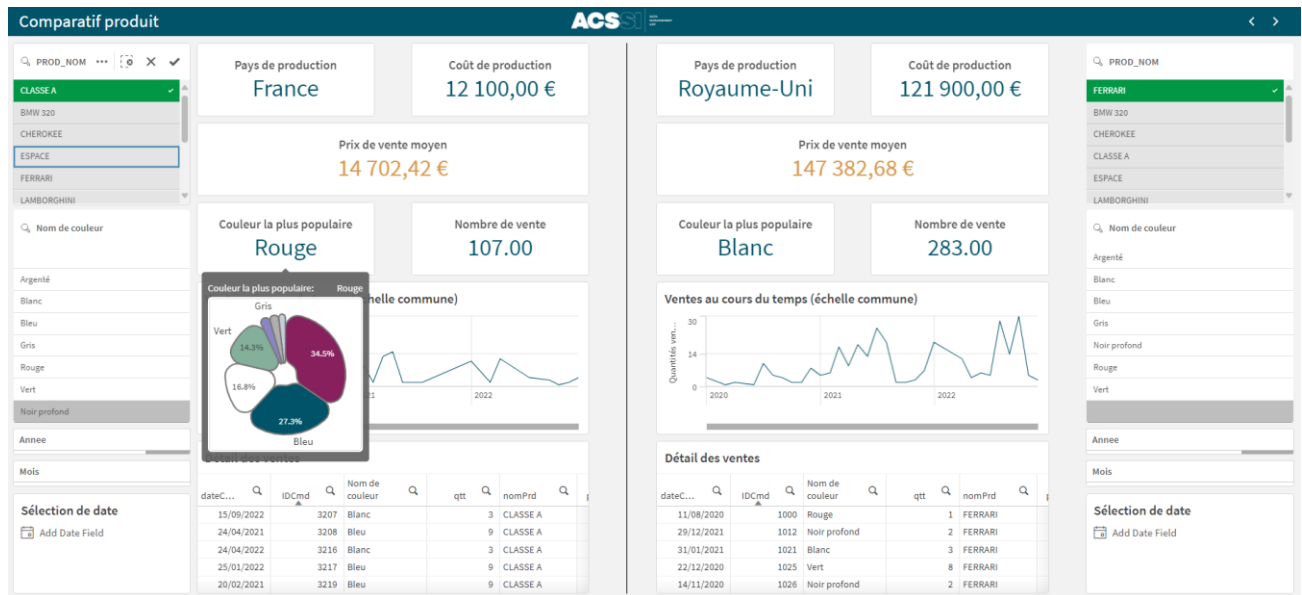
Focus France.

Cette feuille sera consacrée à la France. Elle devra contenir au moins :

- Une carte des ventes selon le département ;
- Un graphique des ventes de produit selon leur famille

Un comparatif des produits. Indices

Vous devrez reproduire la feuille suivante :



Trois choses sont à noter :

- Les graphiques de la partie gauche ne sont sensibles qu'à l'état du filtre de gauche (le produit, la couleur, l'année, le mois et la date). Ceux de droite ne le sont qu'aux filtres sélectionnés à droite ;
- Un graphique précise le nombre de vente par couleur lorsque l'on survole le KPI de couleur la plus populaire. Ce graphique pourrait aussi apparaître au survol du nombre de vente ;
- **En bonus** lorsque vous avez finis la partie Qlik : L'échelle entre les deux graphiques de vente est commune ce qui permet de mieux les comparer. Elle va de 0 au maximum entre les deux courbes.
- **En bonus** lorsque vous avez finis la partie Qlik : Un KPI montre la couleur la plus vendue ;

Détails commande. Indices

La dernière feuille devra montrer un tableau donnant des détails à la maille commande de l'activité. L'utilisateur devra être en mesure de choisir lui-même les dimensions et mesures qu'il souhaite voir dans ce tableau de façon dynamique (pour de l'analyse ou un potentiel export sur Excel).

Indices

Import de la table source. Sujet

Indice 1 :

Indice 2 :

Indice 3 :

Indice 4 :

Fichier client. Sujet

Indice 1 :

Indice 2 :

Fichier Produit. Sujet

Ajout de la famille

Indice 1 :

Indice 2 :

Ajout de la gamme

Indice 1 :

Indice 2 :

Indice 3 :

Suppression des produits invendus

Indice 1 :

Indice 2 :

Fichier CommandeLIG. Sujet

Récupération de la devise

Indice 1 :

Ajout de la devise

Indice 1 :

Séparateur virgule

Indice 1 :

Indice 2 :

Pour aller plus loin. Sujet

Données aberrantes

Indice 1 :

Indice 2 :

Avant de lancer le flux

Indice 1 :

Indice 2 :

Indice 3 :

Fichiers temporaires ou suppression de fichiers

Indice 1 :

Indice 2 :

Format téléphone

Indice 1 :

Indice 2 :

Indice 3 :

Chargement des données. Sujet

Solution du calendrier :

Dashboard. Sujet

Chiffre d'affaires

Indice 1 :

Indice 2 :

Indice 3 :

Focus France. Sujet

Comparatif produit. Sujet

Différence de filtre entre partie gauche et droite

Indice 1 :

Indices 2 :

Graphique lors du survol

Indice 1:

Indice 2:

Echelle commune

Indice 1:

Indice 2 :

Indice 3 :

Couleur la plus vendue

Indice 1 :

Indice 2 :

Indice 3 :

Détails commande. Sujet

Indice 1 :