

GESTION DES DONNÉES & BUSINESS
INTELLIGENCE

EXPERTISE INFORMATIQUE ET SYSTÈME
D'INFORMATION

Conception, modélisation & implantation d'une solution B.I

Soumis par :

Jimmy BOUMANS

Mathieu DORVILLE

Florian GAUTIER

Jérémie LAERA

Arthur TOULOUMOND

Destiné à :

EPSI Bordeaux

Année universitaire 2020-2021



l'école d'ingénierie
informatique

Table des matières

1	Introduction & Contexte	2
2	Conception d'un Data Warehouse ou Datamart	3
2.1	Étude des données produites par l'intermédiaire des applications actuellement en place	3
2.1.1	Calcul de la méthode Wilson	4
2.1.2	Optimisation du stock de sécurité :	5
2.1.3	Taux de disponibilité : mesure quelle quantité de nos produits sont en vente	5
2.1.4	Taux d'occupation : mesure la place disponible dans les hangars	6
2.1.5	Prédire les effets de mode	6
2.1.6	Eviter les livraisons partielles	6
2.1.7	Evitez le blocage de l'activité	7
2.1.8	Résumé des analyses préalables	8
2.2	Étude des données provenant de sources externes	10
2.3	L'apport du Data Mining & la visualisation des données de So Emballage	12
3	Conception du Datamart afin de répondre à la problématique de l'entreprise SO-EMBALLAGE	14
3.1	Solution A	15
3.2	Solution B	15
3.3	Choix de la solution	15
3.4	Les indicateurs proposés	17
3.5	Démonstration des résultats obtenus	17
3.6	Alimentation du DataWareHouse	18
3.7	Préconisations	18
4	Sources	20

1 Introduction & Contexte

La direction de SO-EMBALLAGE est décidée à lancer un projet de grande envergure. Au vu des réflexions déjà menées par les différents responsables des services, il apparaît que l'organisation toute entière devra sans doute subir quelques modifications. Il en suivra une refonte complète du système d'information.

Son besoin est concentré sur la nécessité de rester un acteur important dans le domaine de l'emballage, c'est pourquoi SO-EMBALLAGE souhaite répondre à différentes problématique au travers des décisions suivantes.

- La mise en place d'outils pour une meilleure sélection des fournisseurs et de produits sur la base d'un ensemble de critères.
- Une meilleure gestion des stocks pour moderniser son organisation et assurer un service clientèle de qualité.

C'est pourquoi notre équipe intervient dans l'accompagnement de SO-EMBALLAGE et propose au travers de ce dossier une première solution ainsi que des recommandations pour y arriver.

Nous allons par l'intermédiaire de cet écrit vous décrire et présenter notre ébauche de réalisation.

2 Conception d'un Data Warehouse ou Data-mart

2.1 Étude des données produites par l'intermédiaire des applications actuellement en place

Le point suivant résume notre analyse primaire sur le modèle de données de SO-EMBALLAGE. Un résumé synthétique est disponible dans la section 2.1.8 en fin de section.

Lorsque nous étudions la base de données existante nous pouvons constater les éléments suivants :

- Les tables suivantes inutilisées : LIVRAISON_C / DETAIL_LIVR_C / MOUVEMENT_STOCK
- Les tables *Client* et *Fournisseur* sont séparables en tables *Adresse* et *Contact*
- Il n'y a pas d'indication de capacité de stockage des hangars
- La table *Emplacement* est séparable en tables *Hangars* et *Site* :
 - Hangar : adresse, capacité, type de stockage (produit humide etc.)
 - Emplacement : Allée, Rayon, capacité en m³
- Pour la table *Achat* :
 - Le champ *comment_achat* contient des informations résumables en ENUM
 - * Urgent
 - * Opération promotionnelle
 - * Normal
- Dans la table *Produit*
 - Il y a un manque d'informations sur le conditionnement :
 - * Quel volume prend un lot ?
 - * Conditionner par lot ou à l'unité

- Il n’y a pas d’indication de taille, uniquement à l’unité
- L’indication de quantité par lot est contenue dans un champ texte libre (`comment_produit`)
- Aucun suivi des pertes dégradation ou vol
- Aucun lien entre ce qui a été livré et commandé
- La table *mouvement stock* peut sembler inutile (non-alimentée et non-utilisée)

Afin de pouvoir définir les changements à appliqués aux SGBD il est bon de rappeler les besoins de l’entreprise :

Les besoins principaux relevés sont les suivants :

- - Évitez le surstock/sous-stock
- Prédire les effets de mode
- Évitez les livraisons partielles
- Évitez le blocage de l’activité

Afin de pouvoir identifier les points de blocages et de répondre au mieux aux besoins nous devons évaluer les performances de fonctionnement de l’entrepôt. Ces performances peuvent être mesurées à l’aide d’indicateurs. Certains d’entre eux dépendent de données de la SGBD. En identifiant ces indicateurs nous pourrons déduire les données manquantes dans le SGBD d’origine.

Afin d’éviter le surstock/sous-stock , nous souhaitons parcourir les axes suivants dans les prochaines sous-sections.

2.1.1 Calcul de la méthode Wilson

- Permet de déterminé la quantité optimale pour un réassort
- Calcul les délais optimaux entre chaque de commande
- Optimise les coûts de stock
- Optimise les coûts de passage de commande

- Limites :
 - * Fonctionne mal pour une demande instable
 - * Fonctionne mal avec des délais inconstants ou imprévisibles
- Données :
 - * Coût de passage de la commande :
 - Comporte le prix HT de la ressource à rajouter dans la table Achat
 - Charges transporteur
 - * Le coût de stock
 - * La demande pour ce produit

2.1.2 Optimisation du stock de sécurité :

- Complète la méthode de Wilson
- Permet de couvrir la demande incertaine
- Permet de couvrir les délais incertains
- Plusieurs méthodes de calcul à adapter en fonctions des besoins
- Données :
 - Les ventes pour un produit représenté par la somme des détails commande

2.1.3 Taux de disponibilité : mesure quelle quantité de nos produits sont en vente

- Données : produits/quantité
- Calcul :
 - Pour chaque produit en stock :
 - * Disponibilité : Si quantité $> 0 = 100\%$
 - Faire la moyenne : Somme disponibilité / nombre produits

2.1.4 Taux d'occupation : mesure la place disponible dans les hangars

- Calcul : $\text{Place occupée} / \text{place total}$
- Données :
 - Somme de la capacité de tous les emplacements d'un même hangar
 - Données de volume des produits/conditionner

2.1.5 Prédire les effets de mode

- Fiabilité de la prévision : Permet de d'ajuster et d'anticiper la mobilisation de moyens et de ressources. Nous pourrions en déduire par produits la quantité moyenne consommée et les pics de consommation, afin d'en produire un lot minimum de fonctionnement.
 - Calcul :
 - * Définir une période
 - * Pour chaque article
 - * Calculer l'écart absolu entre la vente réelle et la vente effectuée
 - * Diviser par la vente réelle
 - * Prendre le résultat et le soustraire à 1
 - * Faire le pourcentage de toutes les prévisions

2.1.6 Eviter les livraisons partielles

- Taux de service : Permet de mesurer notre capacité à livrer, et d'évaluer nos prestataires de livraisons. Deux informations peuvent être extraites.
 - Calcul :
 - * Récupérez la liste complète des commandes reçues sur la période souhaitée
 - * Récupérez ensuite sur ces même articles le nombre de commandes livrées dans les temps. Si celles-ci sont en retard, alors entrez 0
 - * Divisez les commandes livrées par les commandes reçues pour connaître votre taux de service.

- Données :
 - * Commande/date
 - * Details_Commande/quantité
 - * Livraison/date_départ
 - * Livraison/date_arrivée
 - * Livraison/date_estimée
 - * Coût livraison
 - * Details_Livraison/quantité
 - * Livreur
 - Permet d'identifier quel transporteur auxquelles nous avons fait appel pour cette livraison.

2.1.7 Evitez le blocage de l'activité

- Livraison dans les temps :
 - Permet d'identifier les retards et leur raisons
 - Calcul de pourcentage entre livraison respecté sur livraison total
 - Calcul de pourcentage entre livraison en retard sur livraison total
 - * Pourcentage d'une raison du retard sur le total de livraison retard
 - Données :
 - * Livraison/date_départ
 - * Livraison/date_arrivée
 - * Raison du retard
- Rotation de stock : Permet de mesurer en combien de temps met votre stock à s'épuiser.
 - Calcul :
 - * Déterminé une période
 - * Stock moyen : Moyenne de produit vendu sur cette période
 - * Calculer la valeur HT du stock moyen
 - * Multiplier le nombre de vente total par le nombre de jour de la période

- * Diviser la valeur du stock moyen par le résultat de la multiplication
- Données
 - * Commande détails / Quantité
- Taux de démarque : Permet de calculer le nombre de perte/vol/remise etc...
 - Calcul :
 - * $\text{Valeur stock après inventaire} / \text{Valeur du stock théorique} \times 100$
 - Données :
 - * Somme de (Produit Quantité \times Prix HT)
 - * Somme de (Produit Compté \times Prix HT)
 - Table *Inventaire* : champs quantité et date_inventaire
- Coût logistique et transport : Représente la part de la logistique et du transport sur le CA , exprimé en pourcentage du chiffre d'affaire total(1). Représente également le coût moyen par article du coût de la logistique et du transport, exprimé en pourcentage du chiffre d'affaire total(2).
 - Calcul :
 - * Sur l'année : $(\text{cout total d'entrepasage} + \text{Coût total des transporteurs} + \text{Charge de service}) / \text{CA} \times 100$ (1)
 - * Sur l'année : $(\text{cout total d'entrepasage} + \text{Coût total des transporteurs} + \text{Charge de service}) / \text{Valeur total de la marchandise vendu} \times 100$

2.1.8 Résumé des analyses préalables

Pour résumé les changements supplémentaires à apporter aux SGBD :

- Table *Produit* :
 - Il n'y a pas de distinction de quantité entre ce qui est disponible et ce qui attend d'être livré

- Ajout Quantité en stock
 - Ajout Quantité disponible
 - Ajout Poids
 - Ajout Quantité_lot
 - Possède 0 ou plusieurs inventaires
 - Possède 0 ou plusieurs conditionnements
- Création Table *Inventaire* :
 - Quantité
 - Date
- Création table *Conditionnement* :
 - Largeur
 - Hauteur
 - Longueur
 - Poids
 - Quantité
- Table *Emplacement* :
 - Possède 1 hangar
 - Ajout Espace_disponible_m³
 - Ajout Rayon
- Table *Livraisons* :
 - date_départ
 - date_arrivée
 - date_estimée
 - Coût livraison
 - A obligatoirement un Transporteur
- Création Table *Transporteur*
 - Nom

- Délai vendu
- Garantie
- Création table énumération *Origine du retard* :
 - Délai non respecté par le fournisseur
 - Délai non respecté par le transporteur
 - Erreur de stock
 - Qualité
 - Regroupement pour expédition
- Table Detail_achat :
 - Indiquez le prix de l'unité acheté en HT

2.2 Étude des données provenant de sources externes

Notre recherche de types de données externes nous a conduit à trouver les sources de données suivantes :

- **PackPlus** (<https://www.packplus.in/>) : Propose une newsletter tous les 15 jours avec des informations sur le monde de l'emballage (Ces données sont à exploiter manuellement puisqu'il n'y a pas de newsletter avec des statistiques, des tendances de marchés et des actualités).
- **Japan Packaging Institute (JPI)** (<https://www.jpi.or.jp/english/>) : Le JPI propose énormément d'informations intéressantes puisqu'il intervient dans les domaines suivants :
 - Réalisation d'enquêtes et de travaux de R&D liés à l'emballage
 - Collecte de données relatives au conditionnement, à la production de statistiques, etc. et partage de ces informations
 - Développement de concepts d'emballages
 - Éducation sur l'emballage et formation du personnel
 - Consultation et orientation concernant les emballages
 - Création et promulgation de normes d'emballage, et conduite d'activités éducatives à leur intention au Japon et à l'étranger
 - Échange d'informations et collaboration avec des organisations liées à l'emballage au Japon et à l'étranger

- Mener des activités mondiales liées à l’emballage, organiser des expositions sur l’emballage
- Publication et vente de livres sur l’emballage.

Il existe un onglet statistique regroupant énormément d’informations sur le domaine de l’emballage, ainsi qu’une section regroupant les dernières publications sur ce domaine.

- **L’université du Michigan** (<https://www.canr.msu.edu/packaging/research/>) : L’université dispose d’un centre de recherche sur le packaging. La périodicité de parution des données est aléatoire mais reste une bonne source de d’informations.
- **Transparency Market Research** (<https://www.transparencymarketresearch.com/packaging-market-reports-35.html?page=2>) : Il s’agit d’un site internet qui regroupe énormément d’études axés sur le domaine de l’emballage. Il est possible d’y retrouver des graphes, des tableaux, des statistiques, des indicateurs d’opportunités ou menaces... Il est cependant important de signaler que cette source de données est payante, cela signifie donc un investissement de la part de SO-EMBALLAGE.
- **Syskevasia** (<https://syskevasia-expo.gr/en/>) est une société Grecque axée sur les expositions ayant pour but de promouvoir les techniques les plus actuelles en matière d’emballage de tout type. “Packaging at its best” comme l’indique son slogan accrocheur, la société se veut innovatrice et convie chaque année divers entreprises spécialisées dans la production d’emballage par le biais d’expositions. Plus d’information sont disponibles sur le site officiel, traduit en anglais depuis la langue grecque.

L’extraction, l’uniformisation, l’analyse et l’exploitation de ces données nécessitent une intervention humaine. Bien-sûr, une partie de ces processus peuvent être automatisés mais la plupart, telle que la lecture des newsletters pour identifier les informations importantes, restent à la charge d’une personne. Il peut donc être intéressant pour la société SO-EMBALLAGE de former une personne en interne ou alors de recruter une personne disposant déjà des compétences nécessaires.

Conjointement aux possibilités multiples d’observer et de travailler avec les sources de données externes, SO-EMBALLAGE devrait orienter ses actions vers un recrutement spécifique. Les ressources possédant les compétences en lien à ces activités sont des ressources cruciales : développeur senior ou data scientist, il conviendra d’effectuer les choix nécessaires afin de

s'adapter au nouveau projet imminent. Nous mentionnons le poste de data scientist car les masses de données que la société collecte peuvent devenir volumineuses et nécessiter des analyses mathématiques plus approfondies : nous souhaitons mettre en avant le domaine du Data Mining.

2.3 L'apport du Data Mining & la visualisation des données de So Emballage

Data Mining signifie “exploration de données”. L'idée principale est de détecter des motifs, les visualiser en les explosant, les modéliser (normalisation, classification) et enfin de les expliquer. Il s'agit d'effectuer de la description à l'aide de statistiques descriptives dans le but de produire des résultats graphiques.

Le cas de figure de la société française SO-EMBALLAGE par le biais de son désir d'implémenter une solution Business Intelligence apporte une problématique touchant le domaine du Data Mining. Cependant, il ne faut pas mélanger BI et Data Mining, la dernière ne produisant pas de données de sortie. Dans l'idéal, l'exploration des données de SO-EMBALLAGE pourrait subir l'ajout d'algorithmes de Machine Learning afin d'effectuer des prédictions.

Exemple : Nous disposons de données telles que les quantités de produits disponibles ainsi que du nombre de commandes par mois. Nous avons donc dans un premier temps choisi de créer un affichage sur le nombre de produits restant pour chaque mois effectif. Ensuite, étant donné que nous possédons des données sur les commandes échelonnées sur plusieurs mois, nous recommanderions de mettre en place un système de prédiction afin d'avoir une visibilité sur les commandes pour les mois à venir. Ainsi l'entreprise et ses décideurs auraient une nouvelle force : ils pourraient revoir les stocks en fonction des résultats.

So Emballage pourrait tout d'abord avoir recours au modèle de processus **CRISP-DM (Cross Industry Standard Process for Data Mining)**¹. Il s'agit d'une approche séparant le processus de Data Mining en six étapes au sein d'un projet IT :

- La compréhension du métier
- La compréhension des données
- La préparation des données

1. Conçu en 1996 et élue méthode la plus utilisée par les Data Miners depuis 2002.

- La modélisation des données
- L'évaluation
- Le déploiement du produit

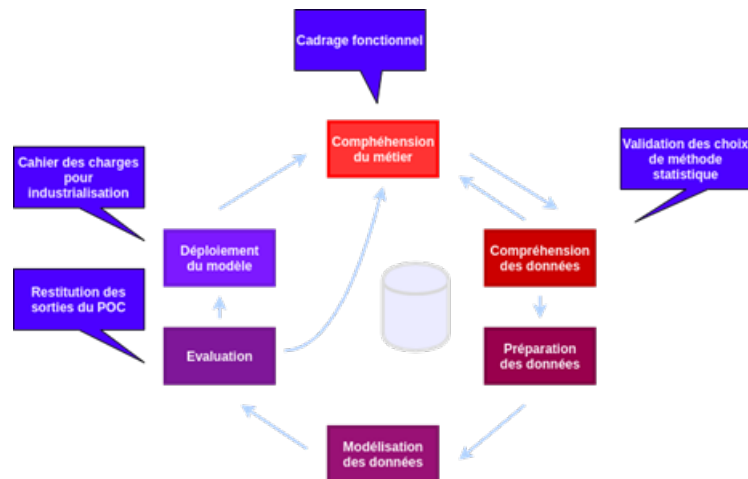


FIGURE 1 – La méthode CRISP-DM représentée pour un processus lambda au sein d'un projet informatique

Lors du recours à cette démarche, il convient d'établir une communication accentuée entre les experts en Data Mining et les ressources du métier.

Actuellement, la solution que nous proposons utilise des composantes de ce domaine par le biais de l'outil Google Data Studio. D'autres solutions sont envisageables et nous les mentionnerons dans notre section dédiée à cet effet.

3 Conception du Datamart afin de répondre à la problématique de l'entreprise SO-EMBALLAGE

Si nous reprenons le besoin initial, nous pouvons schématiser une première architecture permettant de visualiser ce vers quoi nous souhaitons nous orienter :



FIGURE 2 – Schéma de la solution générale pensée

Pour répondre à cette problématique nous sommes tout d'abord parti sur une solution moderne et innovante autour d'une technologie appelée « Suite Elastic ». Cette terminologie regroupe en réalité plusieurs outils dont trois en particulier qui s'avèrent très intéressants dans la situation de SO-EMBALLAGE :

La conception d'un Datawarehouse peut être réalisée par le biais de trois modèles :

- Elastic Search
- Logstash
- Kibana

Il s'agit d'une suite fiable et sécurisée, qui vous permet de collecter n'importe quel format de données depuis n'importe quelle source, puis d'interroger, d'analyser et de visualiser vos données en temps réel.



FIGURE 3 – Schéma de la solution portant sur la suite ELK

Après une étude approfondie des impacts, du temps nécessaires à la mise en place d'une telle solution et de par sa complexité, nous n'avons pas souhaité conserver cette dernière (qui reste tout de même intéressante pour une mise en place à plus grande échelle). Dès lors, nous nous sommes orientés vers des solutions alternatives moins chronophage et qui nous permettent de vous présenter des solutions viables à vos problématiques.

3.1 Solution A

Pour cette première solution, la base de données SO-EMBALLAGE est retravaillée en base de données MariaDB dans un souci de compétences techniques et de simplicité d'utilisation, ce qui signifie que si cela est nécessaire il est tout à fait possible de rester sur une base de données Oracle. Ensuite, un script javascript faisant office d'ETL permet d'extraire les données utiles pour notre situation, de faire des modifications dessus et de les ajouter dans une autre base de données MariaDB qui fait office de DataWarehouse. Enfin, nous utilisons un logiciel afin de visualiser sous forme de tableaux et graphiques les données de notre DataWarehouse, tel que Google Data Studio ou encore Microsoft Power BI².

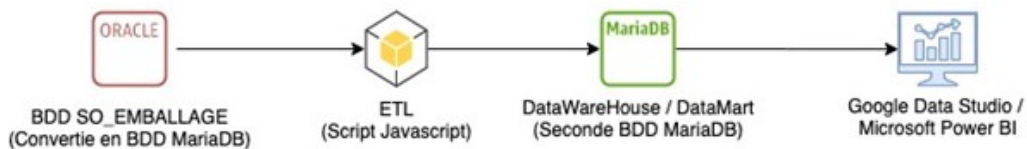


FIGURE 4 – Schéma de la première solution proposée

3.2 Solution B

Une alternative à la première solution est d'oublier l'étape du script permettant d'extraire les données de notre première base de données, mais plutôt d'utiliser des vues SQL qui permettent d'obtenir les données voulues grâce à la bonne combinaison de requêtes. La différence réside dans le fait que les vues SQL sont moins performantes qu'un ETL, car ce dernier permet de traiter les données et de les charger automatiquement vers le DataWarehouse correspondant.

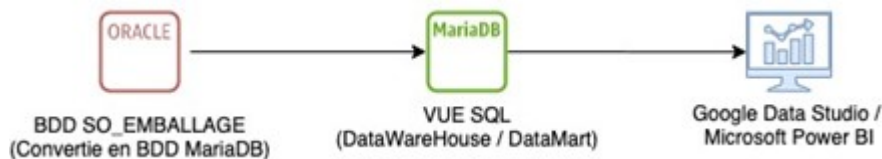


FIGURE 5 – Schéma de la solution alternative proposée

3.3 Choix de la solution

Nous avons donc fait le choix d'implémenter la solution A pour démontrer l'étendu des possibilités qui s'offre à SO-EMBALLAGE. Pour ce faire, nous

2. Power BI offre un panel de fonctionnalités gratuites ainsi qu'une version en ligne avec d'autres caractéristiques

avons converti la base de données Oracle en base de données MariaDB, puis à l'aide d'un script écrit en Javascript nous récupérons les données dans ce schéma de données, nous les modifions et nous les injectons dans un second schéma de données faisant office de Datawarehouse : « SO_WAREHOUSE ». C'est sur ce schéma que nous connectons Google Data Studio pour visualiser des tendances.

La conception d'un Datawarehouse peut être réalisée par le biais de trois modèles :

- Le modèle en étoile
- Le modèle en flocon de neige
- Le modèle en galaxie

Ces modèles comportent deux composants : les dimensions et les faits. Lors de la conception de modèles de données (MCD/MLD), les notions de tables et relations sont mises en avant. Néanmoins, au sein de l'informatique décisionnelle, les termes de dimensions et faits les remplacent.

Une dimension est un axe sur lequel nous souhaitons réaliser une analyse. Un fait est une table sur laquelle portera l'analyse en question et qui reflète l'état de l'entreprise de par ses données opérationnelles.

Notre solution actuelle contient le Datawarehouse SO_WAREHOUSE, solution contenant un total de quatre tables en l'état présent. De ce fait, nous nous rapprochons du modèle en étoile sans pour autant l'atteindre car il nous faudrait améliorer notre conception et basculer toutes les clés vers une table de fait, dans notre cas la table *SITE*, qui nous permet actuellement d'étudier les caractéristiques des produits et des stocks. Notre nombre de tables étant bas, il n'est pas nécessaire de partir sur un modèle en flocon de neige. Notons enfin que nous n'évoquons pas le modèle en galaxie car celui-ci est de fait, le moins pratiqué en entreprise. Ainsi, en améliorant notre conception, il serait possible pour la société d'envisager des analyses concernant les stocks et tendances sur plusieurs sites. Notre solution actuelle demeure principalement axée sur les produits et stocks. Notons que la conception devrait idéalement refléter les 3V³ du Big Data afin d'être au top de la performance.

Les tendances permettront à toute ressource en besoin de consultation (manager, directeur, analyste, CFO, RH) d'être le plus à même de prendre des décisions concernant les critères de leur système d'information. L'objectif final de notre POC (Proof Of Concept) est d'apporter à SO-EMBALLAGE

3. Volume, Vitesse, Variété : Combinaison utilisée afin de traiter des volumes de données conséquentes.

les outils qui lui octroieront la force de frappe pour rester leader du marché tout en modernisant ses mécaniques internes (travail managérial, travail d'ingénierie informatique...).

3.4 Les indicateurs proposés

Pour vous présenter un POC de notre solution, nous avons fait le choix de nous axer sur des indicateurs qui reflètent les problématiques de SO-EMBALLAGE et qui sont les indicateurs suivants :

- Le total des commandes par mois sur une année
- Les références produit les plus vendus par mois
- Les quantités restantes des produits en stock.

3.5 Démonstration des résultats obtenus



FIGURE 6 – Dashboard créés à l'aide de Google Data Studio

La figure 6 montre les représentations graphiques que nous avons mises en place. Elles nous permettent d'entre-apercevoir les possibilités qui s'offrent

à la société SO-EMBALLAGE avec une mise en place de ces outils sur une période de temps plus importante.

Les deux premiers graphes nous permettent d'établir des tendances des ventes sur une année pour alimenter les stocks en conséquence au moment opportun. Le troisième schéma nous permet de visualiser les références produit les plus vendus par mois, nous pouvons ainsi éviter des problématiques de «sur-stock» sur des références qui se vendent moins. Enfin, le dernier tableau nous permet de suivre numériquement les stocks de l'entreprise par référence.

En fonction de l'agrandissement des stocks et des espaces de stockages, il sera facile d'implémenter un suivi par entrepôt à ce niveau.

3.6 Alimentation du DataWarehouse

Le data warehouse est alimenté via plusieurs sources. Tout d'abord, les données produites par SO EMBALLAGE sont envoyées à l'ETL de façon régulière. Ceci est commandé par une Cron task sur une base journalière et durant la nuit afin de ne pas charger les serveurs de base de données durant les heures de travail. Celle-ci est configurable afin de changer l'heure de l'envoi. De plus il est nécessaire d'aller chercher des données venant de sources extérieures à SO EMBALLAGE afin d'avoir une vue globale du marché sur nombre de points différents. Cette tâche est faite manuellement sur une base trimestrielle, il n'est pas nécessaire de subir cette charge de travail supplémentaire plus régulièrement afin d'avoir une bonne vision de l'état du marché actuel. Ces données sont à récupérer dans un certain format (Excel, Json, CSV, etc) et importées directement dans Google Data Studio, où les informations devront manuellement être triées. Afin de visualiser les nouvelles données, on doit finalement penser à appuyer sur le bouton "Actualiser".

3.7 Préconisations

Tout d'abord, il est important de mentionner que l'état des lieux et la structure de la base de données actuelle SO-EMBALLAGE (SGBD Oracle) ne permet pas une utilisation optimale du SI permettant entre-autre une gestion des stocks correcte. Il apparaît donc important de revoir la base de données ainsi que les processus d'alimentation de celle-ci.

Une fois fait, nous pourrions accompagner SO-EMBALLAGE dans la mise en place d'une solution plus pérenne d'analyse et d'exploitation des données. Pour ce faire, nous pouvons recommander les solutions suivantes :

- Utiliser la Suite Elastic comme recommandé initialement : Cette option

est très intéressante car la suite ELK, malgré son arrivée relativement récente sur le marché, remplit tout à fait les besoins d'informatique décisionnel et peut supporter un large volume de données.

- Une alternative d'implémentation par le biais de scripts (R, Python, Javascript...) qui feraient office d'ETL grâce aux bibliothèques présentes dans ces technologies. Python possède des bibliothèques fondées sur des wrappers C/C++ procurant plus de performance lors d'exécutions de scripts. Quant à R, il existe des bibliothèques natives ainsi que le package Shiny, très simple de prise en main et permettant d'implémenter des interfaces web. Une étude sur les performances serait à réaliser car il faut considérer la volumétrie grandissante des bases de données.
- L'utilisation de l'outil Talend Data Quality Studio pour la mise en place de critères de qualité des données et l'utilisation d'algorithmes et de calculs statistiques.
- L'utilisation d'outils fournis par les leaders du marché en matière de solution d'informatique décisionnel tel que Talend ou Microsoft Power BI. L'avantage pour Microsoft Power BI se trouve dans sa facilité de prise en main ainsi que les nombreuses fonctionnalités proposées dans la version gratuite. De plus, il peut être apprécié pour tout aficionado du logiciel Excel car il fonde ses calculs et génération de tableau sur les tableaux dynamiques croisés d'Excel. Les montées en compétence pour ce type d'outil serait moindre et le coût engagé serait réduit.

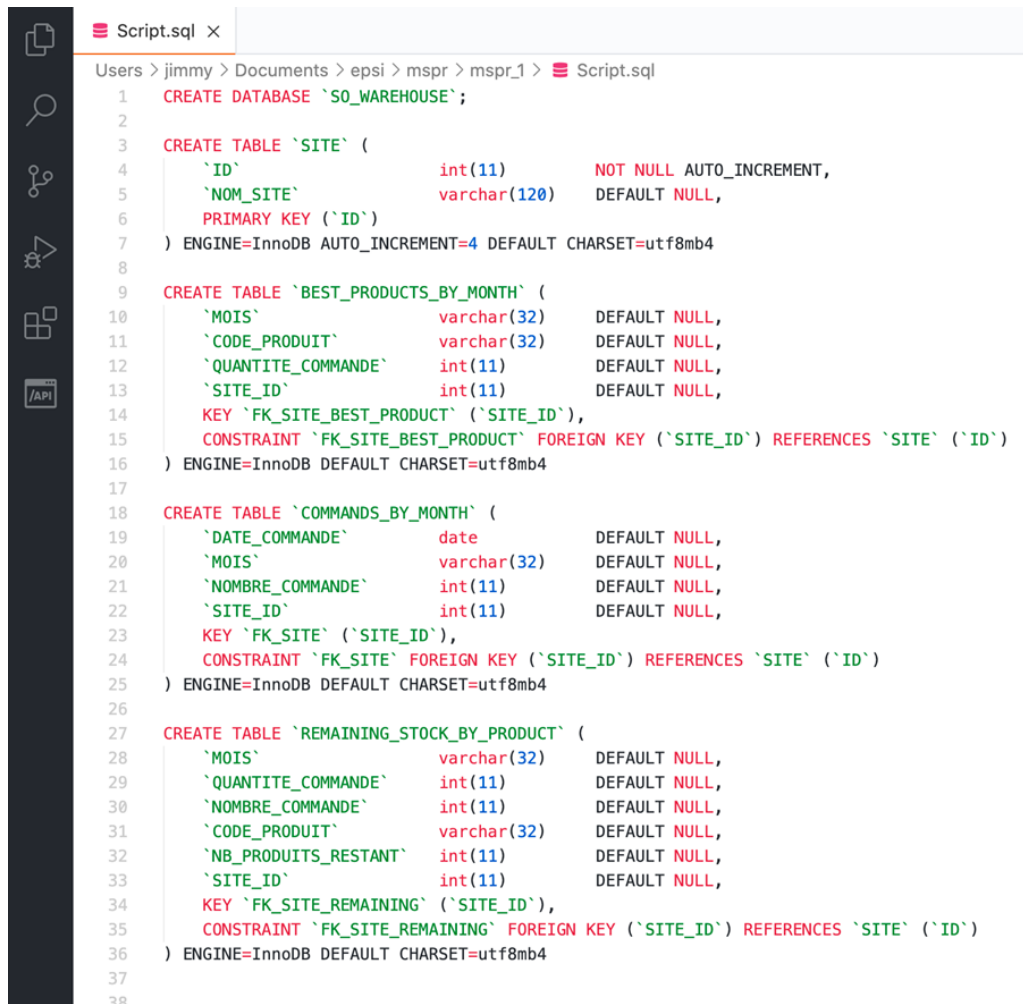
Enfin, une dernière préconisation que nous souhaiterions faire auprès de SO-EMBALLAGE serait de coupler la solution BI que nous proposons avec une solution orientée Big Data. En effet, les indicateurs pourraient être utilisés afin d'effectuer des calculs statistiques et de la prédiction afin d'accompagner la direction de la société dans sa prise de décision.

4 Sources

<https://www.packplus.in/>
<https://www.jpi.or.jp/english/>
<https://www.canr.msu.edu/packaging/research/>
<https://www.transparencymarketresearch.com/packaging-market-reports-35.html?page=2>
<https://syskevasia-expo.gr/en/>

Annexes

Script de création du Data Warehouse après étude de l'équipe.



```
Script.sql x
Users > jimmy > Documents > epsi > mspr > mspr_1 > Script.sql
1 CREATE DATABASE `SO_WAREHOUSE`;
2
3 CREATE TABLE `SITE` (
4   `ID` int(11) NOT NULL AUTO_INCREMENT,
5   `NOM_SITE` varchar(120) DEFAULT NULL,
6   PRIMARY KEY (`ID`)
7 ) ENGINE=InnoDB AUTO_INCREMENT=4 DEFAULT CHARSET=utf8mb4
8
9 CREATE TABLE `BEST_PRODUCTS_BY_MONTH` (
10  `MOIS` varchar(32) DEFAULT NULL,
11  `CODE_PRODUIT` varchar(32) DEFAULT NULL,
12  `QUANTITE_COMMANDE` int(11) DEFAULT NULL,
13  `SITE_ID` int(11) DEFAULT NULL,
14  KEY `FK_SITE_BEST_PRODUCT` (`SITE_ID`),
15  CONSTRAINT `FK_SITE_BEST_PRODUCT` FOREIGN KEY (`SITE_ID`) REFERENCES `SITE` (`ID`)
16 ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4
17
18 CREATE TABLE `COMMANDS_BY_MONTH` (
19  `DATE_COMMANDE` date DEFAULT NULL,
20  `MOIS` varchar(32) DEFAULT NULL,
21  `NOMBRE_COMMANDE` int(11) DEFAULT NULL,
22  `SITE_ID` int(11) DEFAULT NULL,
23  KEY `FK_SITE` (`SITE_ID`),
24  CONSTRAINT `FK_SITE` FOREIGN KEY (`SITE_ID`) REFERENCES `SITE` (`ID`)
25 ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4
26
27 CREATE TABLE `REMAINING_STOCK_BY_PRODUCT` (
28  `MOIS` varchar(32) DEFAULT NULL,
29  `QUANTITE_COMMANDE` int(11) DEFAULT NULL,
30  `NOMBRE_COMMANDE` int(11) DEFAULT NULL,
31  `CODE_PRODUIT` varchar(32) DEFAULT NULL,
32  `NB_PRODUITS_RESTANT` int(11) DEFAULT NULL,
33  `SITE_ID` int(11) DEFAULT NULL,
34  KEY `FK_SITE_REMAINING` (`SITE_ID`),
35  CONSTRAINT `FK_SITE_REMAINING` FOREIGN KEY (`SITE_ID`) REFERENCES `SITE` (`ID`)
36 ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4
37
38
```

FIGURE 7 – Script SQL de génération du Data Warehouse