

Informatique décisionnelle 2 – TD 2

Inégration de données / ETL

Au cours des TD d'informatique décisionnelle 2, on s'intéresse à un cas d'étude simple et à la réalisation d'une partie des étapes nécessaires à la mise en place d'un système d'information décisionnel. Un **rapport court** sera à rendre sur les résultats obtenus sur les trois TD, **par groupe de 3**. Assurez vous donc de bien conserver vos notes sur les différentes tâches par groupe.

Description du cas d'étude : On s'intéresse à une entreprise (on va l'appeler **ScaMo**) qui vend des modèles réduits de véhicules (voitures, moto, bateau, etc.) par commande. **ScaMo** a donc un ensemble de produits dans son catalogue, plusieurs bureaux, des employés, du stock, des commandes, etc. Tous les produits ne sont pas populaires de la même façon, n'ont pas les mêmes coûts, les mêmes prix, etc. Les nombres de ventes peuvent aussi varier beaucoup dans le temps généralement, et par produit.

Ayant un espace de stockage limité, **ScaMo** voudrait mettre en place un système d'information décisionnel afin d'en optimiser l'utilisation, c'est-à-dire d'essayer autant que possible de stocker les produits dans des quantités proportionnelles à leur ventes à venir dans un future proche.

Le système d'information opérationnel de l'entreprise contient plusieurs sous systèmes qui sont autant de sources de données, incluant:

La base de données centrale de l'entreprise : Contient des tables pour les employés (nom, rôle, site, lien hiérarchique), les sites (ville, pays, territoire couvert), les classes de produits (code, description) et les produits (code, nom, classe de produit, échelle, fournisseur, coût d'achat, prix de vente, quantité en stock).

La base de données client : Contient des tables pour les clients (numéro client, nom et contact, adresse, employé représentant), commandes (numéro de commande, date, client, statuts), détails de commandes (numéro de commande, produit, quantité, prix à l'unité), paiements (client, date, montant).

Les données des deux bases sont disponibles à <https://mdaquin.github.io/d/scamo> au format CSV (on ne va pas utiliser les logs du site web pour des raisons de temps limité). Le logiciel Talend Open Studio peut être obtenu à <https://sourceforge.net/projects/talend-studio/> (et un tutoriel est disponible sur Arche. Il y en a beaucoup d'autres en lignes et des documentations).

En vous regroupant dans les groupes de 3 de la fois dernière, finissez le modèle dimensionnel que vous aviez commencé et mettez en place le job dans Talend Open Studio pour transformer les données des bases de **ScaMo** en tables correspondant à ce modèle dimensionnel (au minimum, la table de mesure, sauvegardées au format excel ou csv).