

Informatique décisionnelle 2 – TD 2

Inégration de données / ETL

Au cours des trois prochains TD, nous allons nous intéresser à un cas d'utilisation simple et réaliser en partie les 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 **BuyStuff**) qui vend différents articles de maisons (meubles, fournitures de bureau, technologie) en ligne. **BuyStuff** a donc un ensemble de produits dans son catalogue, des stocks, des commandes, etc. Tous les produits ne sont pas populaires de la même façon, n'apportent pas les mêmes revenus, etc. Les nombres de ventes peuvent aussi varier beaucoup dans le temps généralement, et par produit.

A l'heure actuelle, les stocks sont gérés directement par les trois départements de **BuyStuff** (*furnitures, office supplies, technology*) dans trois entrepôts différents. L'entreprise souhaite revoir cette stratégie et voir si l'espace de stockage pourrait être mieux utilisé (i.e. pour s'assurer d'avoir du stock pour les produits les plus avantageux proche de la demande).

Ayant un espace de stockage limité, **BuyStuff** 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:

La base de données centrale de l'entreprise : Contient des tables sur les départements (ville, capacité de stockage), les produits (nom, catégorie), etc.

La base de données client : Contient des tables sur les clients (nom, catégorie) et leurs adresses.

La base de données des commandes : Contient des tables sur les commandes (date, livraison, client) et les produits commandés (quantité, prix, bénéfices).

Les tables de ces trois sources de données sont fournies sur Arche au format CSV et aux liens suivants : `addresses.csv`, `customers.csv`, `departments.csv`, `order_details.csv`, `orders.csv` and `products.csv`.

Pentaho Data Integration (Community Edition) peut être obtenu à <https://www.hitachivantara.com/en-us/products/dataops-software/data-integration-analytics/pentaho-community-edition.html> et Talend Open Studio peut être obtenu à <https://www.talend.com/products/talend-open-studio/>.

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 Pentaho Data Integration ou Talend Open Studio pour transformer les données des bases de **BuyStuff** en tables correspondant à ce modèle dimensionnel (au minimum, la table de mesure, sauvegardées au format excel ou csv).