

Introduction au Datawarehouse et au Décisionnel

Projet Fantastique : Modélisation du datamart

NF26

Printemps 2017

Problème posé

Vous travaillez en tant qu'ingénieur spécialisé dans les systèmes décisionnels au siège de l'entreprise française "Fantastique".

L'entreprise "Fantastique" vend principalement des ouvrages de divertissement de type science fiction, thriller, policier ... Elle dispose pour cela de plusieurs magasins dans certaines grandes villes de France.

La direction de l'entreprise souhaite faire une étude large sur les ventes de l'année passée afin de prendre des orientations stratégiques nouvelles : ouverture de nouveaux magasins, fermeture ou transfert de magasins mal implantés, extension territoriale à de nouveaux départements français, réorganisation des directions, réorientation du marketing, élargissement ou réduction du catalogue, etc.

La question posée est donc : **quels sont les facteurs sur lesquels on pourrait jouer pour augmenter les ventes ?**

Elle vous charge dans ce cadre de mettre en place une solution logicielle permettant d'intégrer les données pertinentes et de pouvoir les interroger efficacement sous des angles divers.

Notez que bien entendu, la direction éclairée de l'entreprise ne compte pas se fier à ces seuls facteurs de ventes pour prendre ses décisions, mais bien privilégier les facteurs sociaux et territoriaux, en dialoguant avec ses salariés et ses clients, pour maintenir sa mission culturelle et son rôle d'entreprise citoyenne. Votre posture d'ingénieur est aussi de se préoccuper de ces dimensions fondamentales, même si elles seront largement ignorées dans le cadre de cet exercice à vocation essentiellement technique. Elle pourront néanmoins être brièvement abordées en marge de vos rapports d'analyse.

1 Données disponibles

Afin de réaliser votre travail, l'entreprise vous met à disposition les données suivantes :

- **Catalogue des livres** : une base Oracle contient le catalogue complet de l'entreprise que chaque magasin a à sa disposition.
 - Cette base, composée d'une seule table publique catalogue, est disponible sur le serveur Oracle `sme-oracle.sme.utc`, sous le schéma `nf26prof2`.
- **Fichier des ventes** : un fichier contient une consolidation de l'ensemble des ventes de l'année passée réalisées dans chaque magasin.
 - Ces données sont disponibles sous la forme d'un fichier CSV (*Fantastic*) dans un répertoire du serveur `sme-oracle.sme.utc:/home/nf26/data`.
 - La structure du fichier est : numéro de ticket, date de ticket, produit, magasin.
- **Fichier des magasins** : un fichier ODS (`marketing.ods`) géré par la direction marketing contient pour chaque magasin l'organisation des rayonnages.
 - Le responsable des ventes de chaque département décide de l'organisation des rayonnages des magasins de son département.
 - Il existe 3 types de rayonnage : par Auteur (A), par Année (Y), par Éditeur (E).
 - Le fichier est déposé dans un répertoire du serveur `sme-oracle.sme.utc:/home/nf26/fantastic`.
- **Données géographique sur les départements** :
 - Un stagiaire a trouvé sur Internet un fichier CSV (`departementsInsee2003.txt`) permettant de connaître la population de chaque département, un peu daté mais qui pourra suffire. Ce fichier est disponible sur le site de l'UV à l'adresse : <http://www4.utc.fr/nf26>.

Exercice 1. Analyse des données

Inspectez les données et pour chaque source donnez-en le modèle relationnel sous-jacent :

1. Connectez-vous à la base Oracle avec SQL Developer et inspectez le schéma de la table Oracle.
2. Ouvrez un terminal et connectez-vous en ssh au serveur (pour initier une connexion ssh : `ssh user@serveur`). Utilisez la commande Unix `more` ou tout autre outil pour inspecter les premières lignes du fichier `Fantastic` (vous pouvez également copier le fichier dans un répertoire sur votre compte en utilisant la commande : `cp /home/nf26/data/Fantastic .`, ou encore sans vous connecter en utilisant la commande : `scp nf26p0xy@sme-oracle.sme.uttc:/home/nf26/data/Fantastic .`)
3. Téléchargez le fichier `marketing.ods` et ouvrez-le avec un traitement de texte (même démarche que ci-dessus).
4. Téléchargez le fichier CSV `departementsInsee2003.txt` et ouvrez-le avec un éditeur de texte.

2 Etude des besoins des utilisateurs

Un de vos collègues ayant effectué une étude sommaire des besoins, vous les rapporte sous la forme des deux questionnements ci-dessous.

Le contexte de l'exercice ne permet pas de dialoguer réellement avec des utilisateurs, **en situation réelle il faudra développer cette phase de recueil des besoins des utilisateurs**. Vous pourrez amender l'indicateur d'utilité des données en fonction de cette étude.

— Question 1. :

« La direction marketing est en charge de l'implantation des magasins dans les départements et de l'organisation des rayonnages (type de rangement et présence de rayons spécifiques pour les best-sellers). Elle cherche à savoir si l'organisation du rayonnage des magasins a une influence sur les volumes de ventes, et si cela varie en fonction des jours de la semaine ou de certaines périodes de l'année. Elle voudrait également savoir si certains magasins ou départements sont plus dynamiques que d'autres. »

— Question 2. :

« La direction éditoriale se demande si certains livres se vendent mieux à certaines dates et/ou dans certains magasins ou départements. Elle aimerait également savoir si certains auteurs ou éditeurs se vendent mieux, et s'il existe un lien entre l'ancienneté des livres et les ventes. Elle se demande aussi si certaines périodes sont plus propices que d'autres à l'écoulement des livres les plus anciens. »

Exercice 2. Expression des besoins sous forme de requêtes

À partir de l'étude des besoins effectuée par votre collègue, et en fonction des données disponibles, exprimez les **requêtes cibles** de votre système pour chacune des questions.

Rappel : une requête s'exprime sous la forme :

```
mesure 1, mesure 2, mesure 3, ...  
/ entité 1 (propriété, propriété, ...)  
/ entité 2 (propriété, propriété, ...)  
/ ...
```

3 Modélisation du datamart

Exercice 3.1. Définition des processus

A partir des besoins exprimés à partir des deux questions énoncées ci-dessus, définissez le ou les processus analysés.

Exercice 3.2. Définition du grain

Après avoir analysé les fichiers disponibles définissez le grain *a priori* de chaque vue.

Exercice 3.3. Dimensions et hiérarchies

Définissez pour chaque vue les dimensions de votre modèle et précisez les hiérarchies éventuelles.

Exercice 3.4. Définition des faits

Définissez pour chaque vue le(s) fait(s) que vous allez mesurer.

Exercice 3.5. Modélisation du datamart

Intégrez les vues et proposez un schéma **en étoile du datamart**.

Exercice 3.6. Spécifications Établissez les métadonnées du modèle dimensionnel : décrivez chaque donnée (type précis, description...); identifiez la **clé primaire** de chaque dimension, ainsi que les informations descriptives.