

NF26 – TD1 – Introduction au Datawarehouse et au Décisionnel

Projet Fantastique : Modélisation du datamart

Jingyi HU, Julien Jerphanion

Exercice 1 - Analyse des données

1.1 Schéma de la table CATÉGORIE

Cette table contient le catalogue complet de l'entreprise que chaque magasin à sa disposition. Elle comporte 1443 lignes et 9 colonnes.

NOM	TYPE
REF	NUMBER(38,0)
ISBN	VARCHAR2(13 BYTE)
TITLE	VARCHAR2(255 BYTE)
AUTHORS	VARCHAR2(255 BYTE)
LANGUAGE	VARCHAR2(2 BYTE)
PUBDATE	VARCHAR2(25 BYTE)
PUBLISHER	VARCHAR2(255 BYTE)
TAGS	VARCHAR2(255 BYTE)
GENRE	VARCHAR2(255 BYTE)

1.2 Schéma du fichier Fantastic (ventes)

C'est un fichier CSV qui contient une consolidation de l'ensemble des ventes de l'année passée réalisées dans chaque magasin.

Ce fichier possède 512019 lignes et les 4 colonnes suivantes :

- Numéro de ticket : identifiant de la vente
- Date de ticket: date de la vente
- Produit : référence, correspond à la référence ISBN du livre vendu.
- Magasin : "MXXX" (X est entier), correspond à l'identifiant du magasin dans lequel a eu lieu la vente du livre

Remarque :

Un fichier de données brutes ne possède pas forcément d'extension, particulièrement sous Unix. C'est le cas ici avec ce fichier qui ne comporte pas l'extensions .csv.

1.3 Schéma du fichier marketing.ods contient les magasins

C'est un fichier ODS géré par la direction marketing contient pour chaque magasin l'organisation des rayonnages.

Il possède 153 lignes sur les 6 colonnes suivantes :

- Magasin : l'identifiant du magasin
- Département : le département dans lequel est situé le magasin.
- Rayonnage : l'identifiant de rayonnage selon lequel est organisée magasin (A, Y, E)
- 4^{ème} colonne (sans en-tête) : la traduction du code de rayonnage (Auteur, Année, Éditeur)
- Rayon Bestseller : présence ou non d'un rayon bestseller
- Rayon Récent : présence ou non d'un rayon de livre récents

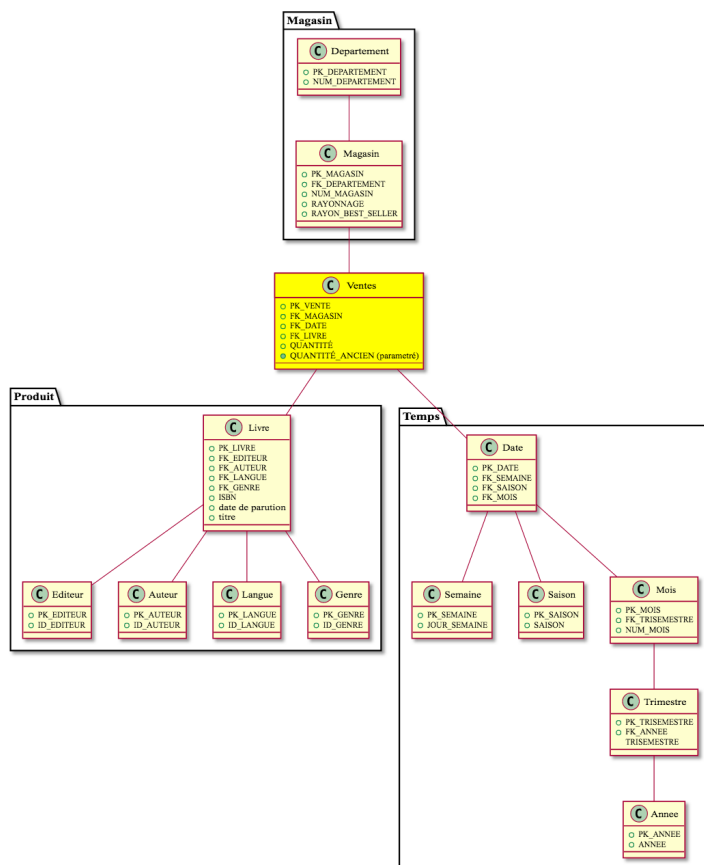
1.4 Fichier INSEE des départements et leur population

Trois colonnes sans en-tête, 95 lignes :

- Numéro de département
- Nom de département
- Population du département en 2003

Exercice 2 - Étude des besoins utilisateurs

2.1 Schéma en flocon – étudié dans la suite



2.2 Requêtes

Exercice 3 – Expression des besoins sous forme de requêtes

Voici les requêtes identifiées pour chacun des questions :

Question 1 : Besoin de la direction marketing :

Volumes des ventes

/ Magasin (rayonnage, rayon_best_seller)

« *L'organisation du rayonnage des magasins* »

/ Semaine

« *Des jours de la semaine* »

/ Trimestre

« *Certaines périodes de l'année* »

/ Département

« *Certains départements sont plus dynamiques que les autres* »

Question 2 : Besoin de la direction éditoriale :

Volumes des ventes

/ livre

/ date

/ magasin

/ département

« *Certains livres se vendent mieux à certaines dates et/ou dans certains magasins ou départements.* »

/ auteur

/ éditeur

« *Certains auteurs ou éditeurs se vendent mieux* »

/ ventes

« *Existe un lien entre l'ancienneté/ des livres et les ventes* »

/ trimestre

« *Certaines périodes...* »

Exercice 3 – Modélisation du datamart

3.1 - Définition des processus

Cette étude du sujet nous amène à identifier plusieurs processus, en particulier des processus de ventes : l'entreprise *Fantastique* souhaiterait savoir quels sont les facteurs influent sur le volume de vente mais aussi sur les types de produits vendus. On imagine qu'il s'agit bien pour l'entreprise d'étudier comme elle pourrait réaliser plus de profits. En ce sens, les questions plus précises qu'elle serait amené à se poser seraient :

- Dans le cas de la première question : l'ouverture ou la fermeture de magasins dans des départements, l'augmentation de l'activité de certains magasins ou la réorganisation de leurs rayonnages ;
- Dans le cas de la seconde question : une réorientation particulière de leur offre de produits vers les produits les plus plébiscités.

3.2 – Définition du grain

Le grain est représenté par les entités :

Date (le date du jour que livre est vendu),

Livre (le livre),

Magasin (le magasin dans lequel il est vendu)

3.3 – Dimensions et Hiérarchies

Ces questions peuvent s'étudier sur 3 dimensions :

- une dimension Temps,
- une dimension Produit,
- une dimension Magasin.

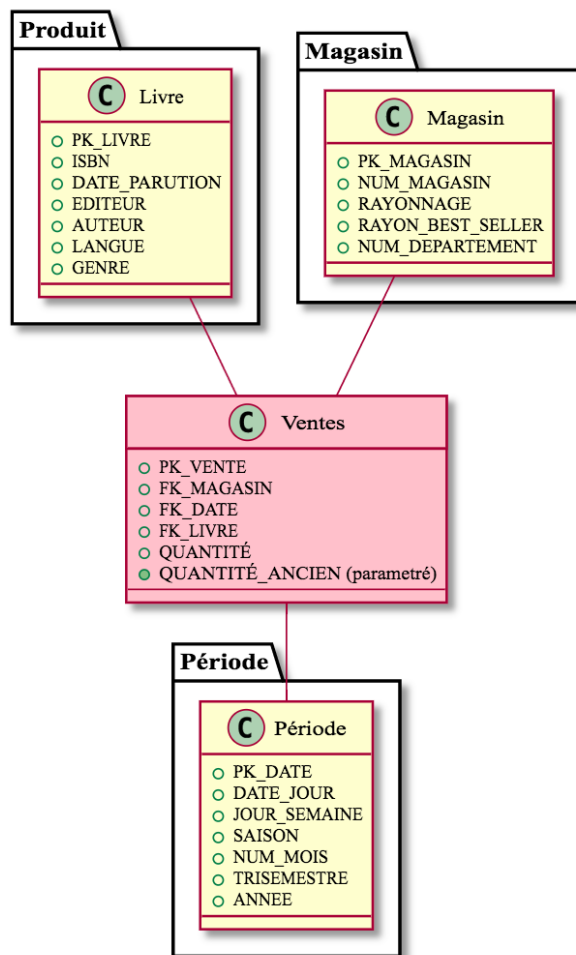
On se réfèrera au schéma en flocon pour une précision plus fine de ces dimensions.

3.4 – Définition des faits

Les faits à mesurer sont, pour la question :

1. le *Volume des ventes*, au regard des entités : Magasin, Semaine, Trimestre, Département
2. le *Volume des ventes*, au regard des entités : Livre, Auteur, Editeur, Magasin, Département, Trimestre

3.5 – Modélisation du datamart



3.6 – Spécification

3.6.1 Dimension Magasin

NOM	TYPE	DESCRIPTION
PK_MAGASIN	Int	La clé primaire
NUM_MAGASIN	Texte	Identifiant du magasin
RAYONNAGE	Caractère	Identifiant de rayonnage (A, Y, E)
RAYON_BEST_SELLER	Int	Présence ou non d'un rayon bestseller
NUM_DEPARTEMENT	Int	Le département dans lequel est situé le magasin

3.6.2 Dimension Temps

NOM	TYPE	DESCRIPTION
PK_PERIODE	Int	
DATE_JOUR	Date(jj/mm/année)	
JOUR_SEMAINE	Int	
SAISON	Texte	
NUM_MOIS	Int	
TRIMERSTRE	Int	
ANNEE	Int	

3.6.3 Dimension Produit

NOM	TYPE	DESCRIPTION
PK_LIVRE	Int	
ID_LIVRE	Int	ISBN
DATE_PARUTION	Date	
EDITEUR	Int	
AUTEUR	Texte	
GENRE	Texte	
LANGUE	Texte	