

Compte-Rendu NF26 TD1

Étape 1 : Analyse des données

Nous avons inspecté 4 types de fichiers :

1. Catalogue des livres
 2. Fichiers des ventes
 3. Fichiers des magasins
 4. Données géographiques sur les départements
- Pour le fichier 1, nous observons que les livres sont caractérisés par une référence, un ISBN, un titre, un auteur, une langue, une date de publication, un éditeur, des tags et un genre. Un livre peut posséder plusieurs auteurs et tags.

| REF | ISBN | TITLE | AUTHORS | LANGUAGE | PUBDATE | PUBLISHER | TAGS | GENRE |
|-----|----------|--------------------------------------|-----------------------|----------|-------------------|-----------|---------------------------------|-------------|
| 1 | 2778 721 | [Anita Blake-A]Narcisse Enchaîné | Hamilton,Laurell Kaye | fra | 2001-01-01T23:... | ? | Fantastique, Roman | Fantastique |
| 2 | 2833 779 | [La Communauté du Sud-1] Quand le... | Harris,Charlaine | fra | 2001-01-01T23:... | ? | Fantastique | Fantastic |
| 3 | 315 141 | [Assassin Royal-07] Le prophète b... | Hobb,Robin | fra | 2001-01-01T21:... | ? | Fantasy | Fantastic |
| 4 | 2883 828 | [Assassin Royal-08] La secte maudite | Hobb,Robin | fra | 2001-01-01T21:... | ? | Fantasy | Fantastic |
| 5 | 2770 712 | [Merry Gentry-1] Le baiser des om... | Hamilton,Laurell Kaye | fra | 2000-01-14T23:... | ? | Fantastique | Fantastic |
| 6 | 339 233 | [Anita Blake-9]Papillon d'Obsidienne | Hamilton,Laurell Kaye | fra | 2000-01-01T23:... | ? | Fantastique, Roman | Fantastique |
| 7 | 2882 827 | [Les Aventuriers de la mer-07] Le... | Hobb,Robin | fra | 2000-01-01T21:... | ? | Fantasy | Fantastic |
| 8 | 317 143 | [Les Aventuriers de la mer-08] Om... | Hobb,Robin | fra | 2000-01-01T21:... | ? | Fantasy | Fantastic |
| 9 | 2904 848 | [Les Aventuriers de la mer-09] Le... | Hobb,Robin | fra | 2000-01-01T21:... | ? | Fantasy | Fantastic |
| 10 | 259 214 | [Robert Langdon-1]Ange et démons | Brown,Dan | fra | 1999-12-31T21:... | ? | Policier, Thriller Scientifi... | Crime |

- Pour le fichier 2, il contient des informations comme le numéro de ticket, la date du ticket, le produit et le magasin. Mais il y a des données manquantes ou incorrectes.

```

-bash-3.2$ cat Fantastic
"216000001";"2014-08-05";"293";"M92"
"216000001";"2014-08-05";"9782702480526";"M92"
"289000002";"2014-10-17";"9782266035149";"M112"
"352000003";"2014-12-19";" "; "M9"
"022000004";"2014-01-23";"274";"M101"
"304000005";"2014-11-01";"9782221108871";"M135"
"304000005";"014-11-01";"9782290314944";"M135"
"0";"2014-12-23";"816";"0"
"227000007";"2014-08-16";"9782070439966";"M76"
"226000008";"2014-08-15";"9782266012263";"M146"
"226000008";"2014-08-15";"9782352940173";"M146"
"146000009";"2014-05-27";"9782709623155";"M119"
"146000009";"2014-05-27";"9782264044730";"11"

```

- Le fichier 3 donne les informations de magasin, il contient les numéros de département et magasin. Il existe 3 types de rayonnage : par Auteur, par Année et par Éditeur. Il enregistre aussi la présence du rayon Bestseller/Récent.

| Département | Magasin | Rayonnage | Year | Rayon Bestseller | Rayon Recent |
|-------------|---------|-----------|--------|------------------|--------------|
| 02 | M146 | Y | 0 | 1 | |
| 06 | M17 | A | Author | 1 | 1 |
| 06 | M91 | A | Author | 1 | 1 |
| 06 | M100 | A | Author | 1 | 1 |
| 06 | M121 | A | Author | 1 | 1 |
| 06 | M124 | A | Author | 1 | 1 |
| 06 | M141 | A | Author | 1 | 1 |
| 06 | M152 | A | Author | 1 | 1 |
| 08 | M51 | E | Editor | 1 | 1 |
| 08 | M103 | E | Editor | 1 | 1 |
| 09 | M105 | A | Author | 1 | 1 |
| 11 | M128 | E | Editor | 1 | 0 |
| 13 | M9 | E | Editor | 0 | 0 |

- Le fichier 4 donne la population de chaque département.

```
"01"; "Ain"; "529378"
"02"; "Aisne"; "552320"
"03"; "Allier"; "357110"
"04"; "Alpes-de-Haute-Provence"; "144809"
"05"; "Hautes-Alpes"; "126636"
"06"; "Alpes-Maritimes"; "1022710"
"07"; "Ardèche"; "294522"
"08"; "Ardennes"; "299166"
"09"; "Ariège"; "142834"
"10"; "Aube"; "301388"
"11"; "Aude"; "319611"
"12"; "Aveyron"; "277779"
"13"; "Bouches-du-Rhône"; "1861068"
"14"; "Calvados"; "663408"
```

Étape 2 : Étude des besoins des utilisateurs

Ensuite, nous avons fait l'analyse de l'expression des besoins des utilisateurs. Pour la question 1, avec l'expression comme "une influence sur les volumes de ventes", nous pouvons définir que le processus analysé est la vente (nombre de ventes). Pour la question 2, selon l'expression comme "si certains livres se vendent mieux", nous pouvons définir que c'est la même mesure que la question 1.

Dans la question 1, la direction marketing veut savoir l'influence de l'organisation des rayonnage, du temps et du lieu des magasins sur le nombre de ventes. La requête associe les 3 entités suivantes : "Magasin", "Département" et "Date". Le résultat demandé est "Nombre de ventes".

Vue 1 : NB_Vente

/Magasin (Rayonnage)

/Département

/Date (jour, mois, semaine, année)

Dans la question 2, la direction éditoriale s'intéresse à l'influence du temps, du lieu, des caractéristiques d'un livre (auteur, éditeur et ancienneté) sur le nombre de vente. La requête associe les 3 entités suivantes : "Magasin", "Département", "Date", "Livre", "Auteur" et "Éditeur".

Vue 2 : NB_Vente

/Date (jour, mois, semaine, année)

/Magasin

/Département

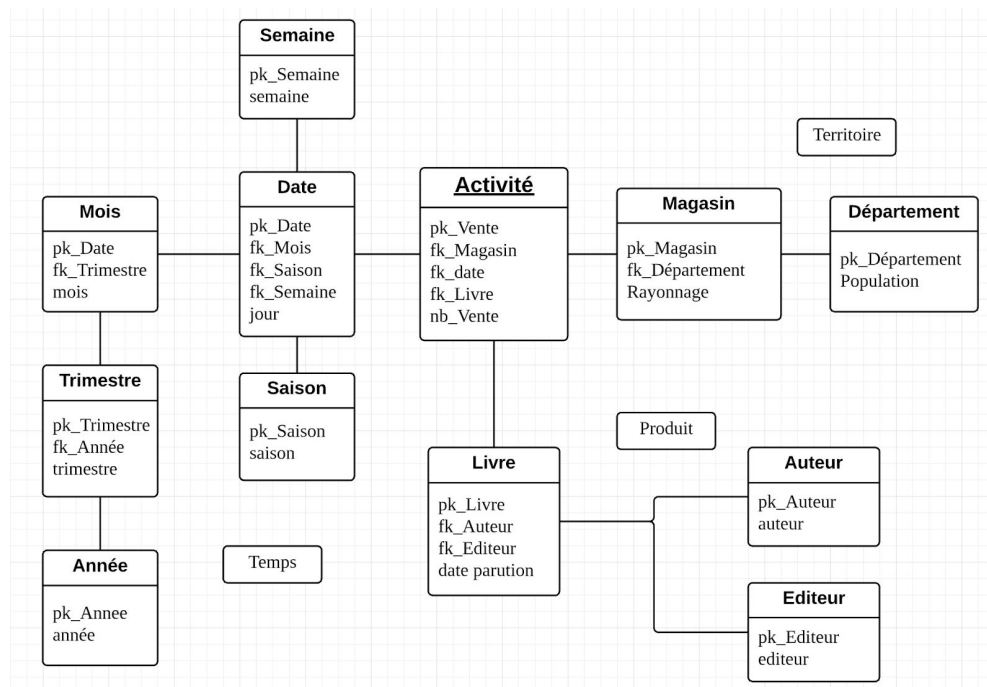
/Livre (date de parution)

/Auteur

/Éditeur

Ensuite, on fait l'intégration des deux vues. On peut mettre les deux entités "Magasin" et "Département" en hiérarchie comme chaque magasin appartient à un département. On met aussi les informations du temps en hiérarchie et on ajoute deux entités "Semaine" et "Saison" pour créer les hiérarchies multiples. On met également "Auteur", "Éditeur" après "Livre" comme chaque livre contient des informations d'auteur et éditeur. Si le directeur s'intéresse aux autres caractéristiques des auteurs et éditeurs, on peut créer deux nouvelles dimensions.

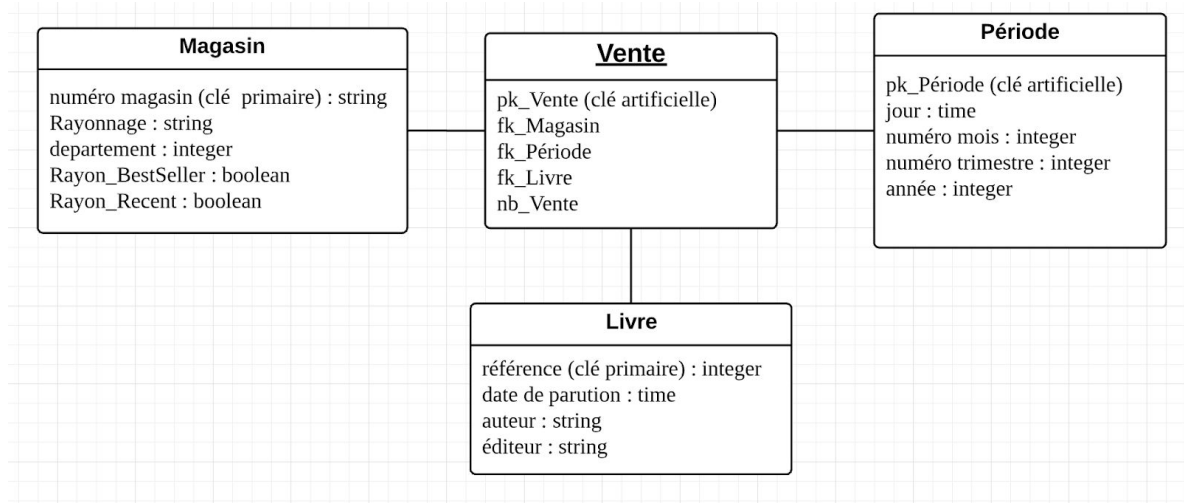
On peut fusionner les deux vues et on obtient un contexte comme ci-dessous :



Ensuite, on vérifie la normalisation du contexte. Il n'existe pas de dépendance fonctionnelle entre deux entités appartenant à des dimensions différentes. Il y a une seule mesure (Nb_Vente) à mesurer. Le grain du contexte est défini par la combinaison Date / Magasin / Livre. Le nombre de ventes est défini pour le grain. Le graphe de chaque dimension est acyclique.

Étape 3 : Modélisation du datamart

Dans cette étape, on fait la modélisation avec le schéma en étoile. Bien qu'il y ait une forte redondance, c'est plus efficace. Les types de données et les clés primaires sont précisés dans le schéma en étoile.



- Pour l'entité "Magasin", la clé primaire est un numéro de magasin unique qui commence par une lettre 'M', Rayonnage précise le type de rayon (Y par année, A par auteur, E par éditeur), département donne le numéro de département. les Rayon_BestSeller et Rayon_Recent sont booléens pour indiquer si le rayon existe.
- Pour l'entité "Période", il est commode d'utiliser une clé artificielle pour identifier les dates.
- Pour l'entité "Livre", on choisit la référence comme la clé primaire car c'est unique pour chaque livre, la date parution est en forme JJMMAAAA. Auteur et éditeur contiennent les nom et prénom des personnes.