

# Entrepôts de données et Big Data

Cette feuille de TD/TP se fait sur 2 séances : partie 1 le 30/09 et partie 2 le 07/10. Le rendu est facultatif.

## PARTIE 1

### 1) Les interrogations : requêtes transactionnelles vs analytiques

Considérons le cas d'un système d'information d'une chaîne nationale de cinémas multisalles. Pour chacune des requêtes suivantes, identifiez s'il s'agit d'une requête transactionnelle ou analytique. Justifiez vos réponses, en faisant des hypothèses si nécessaire.

- 1) Restent-ils des billets à Montpellier pour la séance de 20 heures du film "Logan" ?
- 2) Aurait-on éventuellement pu proposer des plus grandes salles et plus de séances pour le dernier film de Star Wars ?
- 3)  

```
SELECT Film.titre, Cinema.nom, Date.mois, COUNT(Place.placeID)
FROM Film, Ventes, Cinema, Place, Temps, Date
WHERE Ventes.filmID = Film.filmID AND Ventes.cinemaID = Cinema.cinemaID
AND Ventes.tempsID = Temps.tempsID AND Place.cinemaID = Cinema.cinemaID
AND Ventes.dateID = Date.dateID
GROUP BY Film.titre, Cinema.nom, Date.mois
```
- 4)  

```
SELECT Temps.creneau, COUNT(*)
FROM Ventes, Temps
WHERE Ventes.tempsID = Temps.tempsID
GROUP BY Temps.creneau
```
- 5)  

```
INSERT INTO Ventes
VALUES ('film1','cinema24','date2','temps3','place44','7.50')
```

### 2) Un entrepôt de données pour [Amazon](#)

1. Proposez un modèle en étoile pour un entrepôt de données permettant l'analyse des ventes dans Amazon. Le modèle doit pouvoir permettre une analyse temporelle des ventes en fonction des *produits*, des *utilisateurs*, ainsi que des *promotions* mises en place.
2. Comment feriez vous pour intégrer les commentaires client dans ce modèle ?
3. Proposez trois requêtes analytiques pour le modèle de données en étoile que vous avez conçu.

### 3) Requêtes Analytiques

Considérez la table de faits (simplifiée) suivante, enregistrant les ventes journalières chez Monoprix (attention : un script de création et remplissage de la base est à disposition dans la page Moodle du cours).

```
CREATE TABLE ventes_monoprix (  
  id_date VARCHAR(10) NOT NULL,  
  id_produit VARCHAR(10) NOT NULL,  
  id_magasin VARCHAR(10) NOT NULL,  
  id_ville VARCHAR(10) NOT NULL,  
  montant_journalier NUMBER(10,2) NOT NULL );
```

Exprimez en SQL les interrogations suivantes, à l'aide de l'opérateur GROUP-BY.

1. Donner le montant total des ventes par produit
2. Donner le montant total des ventes par produit et par ville
3. Donner le montant total des ventes par produit et par jour
4. Donner la moyenne du montant des ventes par magasin et par jour
5. Donner le montant total des ventes par ville par jour
6. Donner le montant total des ventes par produit, ville et jour

Enfin, testez les options ROLLUP et CUBE et comparer les résultats. Pourrait-on regrouper les interrogations grâce à ces options ?

## PARTIE 2

### 4) Classification des faits

Considérez la liste de faits ci-dessous, chacun correspondant à un tuple de la table des faits de l'entrepôt.

- a) Précisez s'il s'agit d'un fait transactionnel ou d'un snapshot. Justifiez votre réponse.
  - b) Précisez si (l'éventuelle) mesure correspondante est additive, semi-additive ou non-additive. Justifiez votre réponse.
1. Un fait  $(j, p, c, m, x)$  existe lorsqu'un produit  $p$  est acheté par un client  $c$  le jour  $j$  au magasin  $m$ . La mesure  $x$  correspond au prix total.
  2. Un fait  $(j, p, m, x)$  existe lorsqu'un produit  $p$  est acheté le jour  $j$  au magasin  $m$ . La mesure correspond au chiffre d'affaires.
  3. Un fait  $(j, p, m, x)$  existe pour chaque combinaison de produit  $p$ , magasin  $m$  et jour  $j$ . La mesure  $x$  correspond au stock de  $p$  en  $m$  le jour  $j$ .
  4. Un fait  $(j, p, m, x)$  existe pour chaque combinaison de produit  $p$ , magasin  $m$  et jour  $j$ . La mesure  $x$  correspond au nombre de ventes de  $p$  en  $m$  cumulées depuis le début de l'année jusqu'au jour  $j$ .
  5. Un fait  $(c, e, j)$  existe lorsqu'un appel du client  $c$  le jour  $j$  est traité par l'employé  $e$ . Aucune mesure n'existe.
  6. Un fait  $(c, j, x)$  existe lorsqu'un client  $c$  le jour  $j$  laisse une note sur un produit acheté. La mesure  $x$  est la note donnée par le client.
  7. Un fait  $(c, e, j, x)$  existe lorsqu'un appel du client  $c$  le jour  $j$  est traité par l'employé  $e$ . La mesure  $x$  est la durée de l'appel en secondes.
  8. Un fait  $(m, b, j, x)$  existe lorsque la monnaie  $m$  est changée à la banque  $b$  le jour  $j$ . La mesure  $x$  est le montant total de la monnaie changée en euros.
  9. Un fait  $(m, b, j, x)$  existe lorsque la monnaie  $m$  est changée à la banque  $b$  le jour  $j$ . La mesure  $x$  est le cours de change moyen de  $m$  en euros pour toutes les transactions du jour  $j$ .