



# HMIN122M

## Entrepôts de données et big-data

### Rendu TD/TP1 - EDBD

---

**Auteur :**

Gracia-Moulis Kévin  
Canta Thomas

Master 1 - AIGLE/DECOL  
Faculté des sciences de Montpellier  
Année universitaire 2020/2021



# Table des matières

<b>1. Les interrogations : requêtes transactionnelles vs analytiques.</b>	<b>2</b>
Question 1 . . . . .	2
Question 2 . . . . .	2
Question 3 . . . . .	2
Question 4 . . . . .	2
Question 5 . . . . .	2
<b>2. Un entrepôt de données pour Amazon</b>	<b>3</b>
Question 1 . . . . .	3
Question 2 . . . . .	3
<b>3. Requêtes Analytiques</b>	<b>4</b>
<b>4. Classification des faits</b>	<b>5</b>
4.1 - 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. . . . .	5
4.2 - 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. . . . .	5
4.3 - 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 $j$ . . . . .	5
4.4 - 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
4.5 - Un fait $(c,e,j)$ existe lorsqu'un appel du client $c$ le jour $j$ est traité par l'employée. Aucune mesure n'existe. . . . .	6
4.6 - 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. . . . .	6
4.7 - Un fait $(c,e,j,x)$ existe lorsqu'un appel du client $c$ le jour $j$ est traité par l'employée. La mesure $x$ est la durée de l'appel en secondes. . . . .	6
4.8 - 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. . . . .	6
4.9 - 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 $j$ . . . . .	6

# 1. Les interrogations : requêtes transactionnelles vs analytiques.

## Question 1

C'est une requête *transactionnelle* car le but ici n'est pas de faire de l'analyse sur le résultat de cette requête, au contraire on affiche une information très précise.

## Question 2

C'est une requête *analytique* car ici on veut analyser les résultats de cette requête afin de pouvoir potentiellement tirer des bénéfices en offrant plus de places et plus de séances.

## Question 3

C'est une requête *analytique* car ici on affiche le nombres de places totales par le titre du film, nom du cinéma et le mois ceci pourrait permettre, par la suite, une analyse plus complexe de ces données.

## Question 4

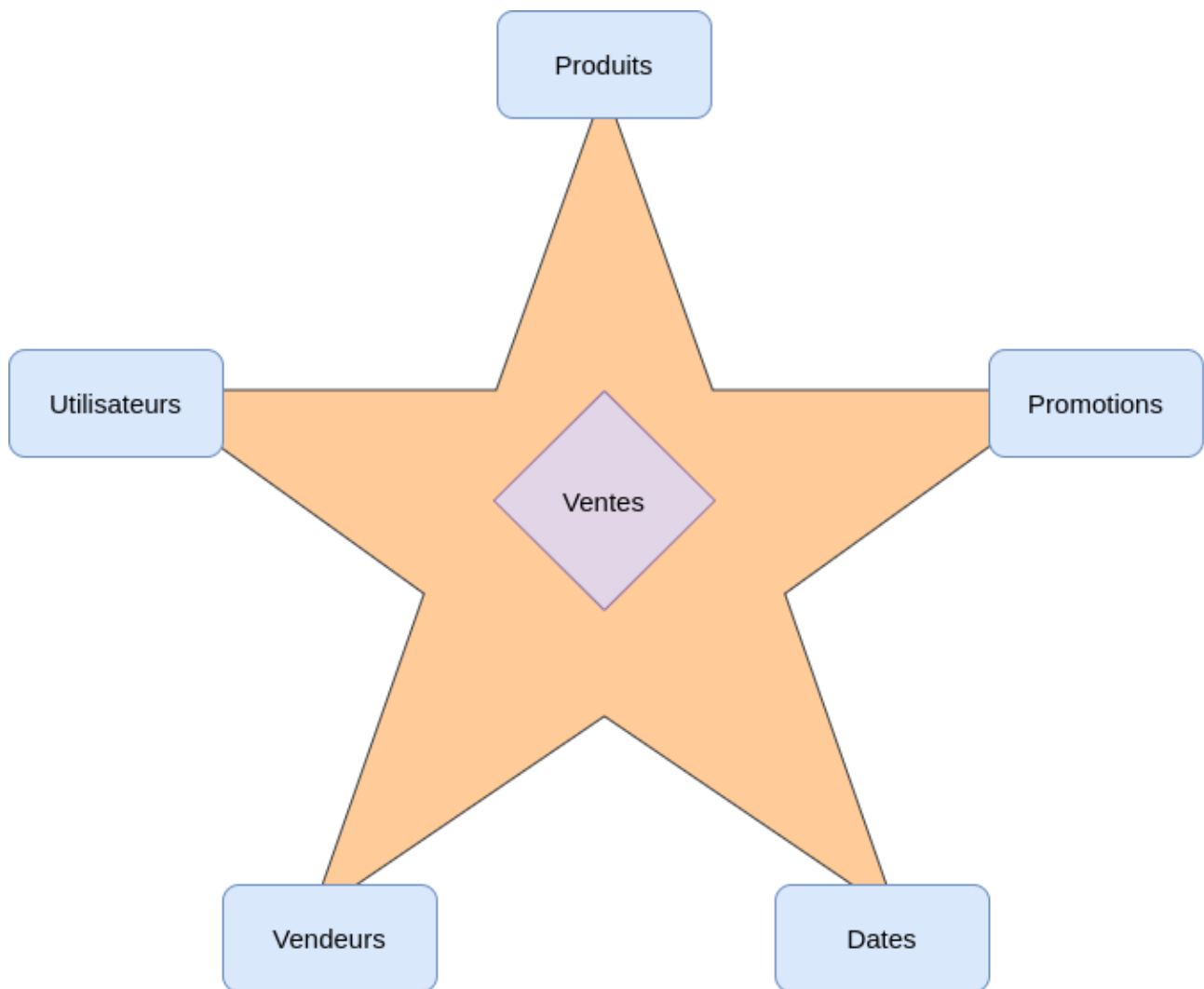
C'est une requête *analytique* car on calcul le nombre de ventes pour chaque créneau.

## Question 5

C'est une requête *transactionnelle*, cette requête n'a pas vocation à être analysée

## 2. Un entrepôt de données pour Amazon

### Question 1



### Question 2

Pour intégrer les commentaires client dans ce modèle on pourrai l'ajouter en attribut de la table *Ventes*.

### 3. Requêtes Analytiques

Les options ROLLUP et CUBE permettrait d'enlever la plupart des requêtes ci-dessous puisqu'elles effectueraient les différentes combinaisons.

Les requêtes utilisées :

---

```
0 SELECT id_produit, SUM(montant_journalier)
1 FROM ventes_monoprix
2 GROUP BY id_produit;
3
4 SELECT id_produit, id_ville, SUM(montant_journalier)
5 FROM ventes_monoprix
6 GROUP BY id_produit, id_ville;
7
8 SELECT id_produit, id_date, SUM(montant_journalier)
9 FROM ventes_monoprix
10 GROUP BY id_produit, id_date;
11
12 SELECT id_magasin, id_date, AVG(montant_journalier)
13 FROM ventes_monoprix
14 GROUP BY id_magasin, id_date;
15
16 SELECT id_ville, id_date, SUM(montant_journalier)
17 FROM ventes_monoprix
18 GROUP BY id_ville, id_date;
19
20 SELECT id_produit, id_ville, id_date, SUM(montant_journalier)
21 FROM ventes_monoprix
22 GROUP BY id_produit, id_ville, id_date;
```

---

## 4. Classification des faits

**4.1 - 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.**

a. Il s'agit d'un fait transactionnel, si c'était un snapshot il y aurait énormément de ligne qui valent zéro (les jours où il n'achète pas).

b. La mesure est additive car l'on peut ajouter les différentes dépenses.

**4.2 - 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.**

a. Il s'agit d'un snapshot car avec un fait transactionnel il faudrait mettre à jour la base de donnée à chaque transaction.

b. La mesure est additive car on somme les gains du produit.

**4.3 - 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  $j$ .**

a. Il s'agit d'un snapshot car c'est la valeur du produit  $p$  au jour  $j$  au moment de la requête qui nous intéresse.

b. La mesure est semi-additive car il n'y a pas d'intérêt à additionner le stock de chaque jour mais l'on peut en étudier la variation.

**4.4 - 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  $j$ .**

a. Il s'agit d'un snapshot car c'est la valeur du produit  $p$  des magasins jusqu'au jour  $j$ .

b. La mesure est non-additive car on somme les ventes du début de l'année jusqu'au jour j.

**4.5 - Un fait  $(c,e,j)$  existe lorsqu'un appel du client c le jour j est traité par l'employée. Aucune mesure n'existe.**

a. Il s'agit d'un fait transactionnel car si c'était un snapshot il y aurait énormément de ligne qui valent zéro (les jours où il n'appelle pas).

b. Il n'y a pas de mesure.

**4.6 - 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.**

a. Il s'agit d'un fait transactionnel car si c'était un snapshot il y aurait énormément de ligne qui valent zéro (les produits non noté) de plus le client ne va pas renoté chaque produit chaque jour.

b. La mesure est semi-additive car l'on pourrait faire la moyenne des notes d'un produit par exemple.

**4.7 - Un fait  $(c,e,j,x)$  existe lorsqu'un appel du client c le jour j est traité par l'employée. La mesure x est la durée de l'appel en secondes.**

a. Il s'agit d'un fait transactionnel car si c'était un snapshot il y aurait énormément de ligne qui valent zéro (les jours où il n'y a pas d'appel) de plus même chose qu'au dessus.

b. La mesure est semi-additive car l'on pourrait calculer la moyenne du temps d'appel.

**4.8 - 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.**

a. Il s'agit d'un snapshot car on part du principe qu'une banque est tellement massive qu'il y aura forcément des changes de chaque monnaie existante chaque jour.

b. La mesure est additive car l'on pourrait faire la moyenne entre chaque jour.

**4.9 - 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  $j$ .**

- a. Il s'agit d'un snapshot car nous obtenons le cours de change moyen toutes les 24heures.
- b. La mesure est non-additive car on obtient la moyenne des changes de la journée.