

# TD Data Mining

## 1 Entrepôt de données - OLAP

### Exo1

On veut construire un entrepôt de données afin de stocker les informations sur les consultations médicales au sein de l'Union Européenne. Après analyse, il s'avère que l'on peut centraliser les informations sous la forme du schéma relationnel suivant : `personne(id, nom, tel, adresse, sexe)`

`médecin(id, nom, tel, adresse, sexe, spécialité)`

`consultation(id_med, id_pers, date, prix)`

1. Proposer un schéma en étoile qui tienne compte pour la date, du jour de la semaine (lundi, ...), du mois, du trimestre et de l'année (on a 2 mesures : `prix_consult` et `nombre`).
2. A partir du cuboïde de base, quelles opérations OLAP (roll up, driwn down, slice, ...) faut-il appliquer pour :
  - (a) avoir le coût total des consultations par docteur en 2003
  - (b) le nombre de consultations par jour de la semaine, par spécialité et par sexe du patient
  - (c) le coût des consultations par patiente pour les mois d'octobre.

### Exo2

Supposons qu'un cuboïde de base a 3 dimensions (A, B, C), avec le nombre suivant de cellules :  $|A| = 1\,000\,000$ ,  $|B| = 100$ , et  $|C| = 1\,000$ .

1. Si chaque dimension n'a qu'un niveau, dessiner le treillis des cuboïdes
2. Si chaque cellule est stockée sur 4 octets, quelle est la taille totale du cuboïde?

## 2 Application

Vous trouverez sur la page web du cours , le fichier contenant un module python rudimentaire pour la gestion d'un cube molap défini à partir d'un fichier csv.

Le but de l'exercice est de rajouter les fonctionnalités `'slice'`, `'dice'` et `'view2d'` à la classe `MolapCube` de ce module.