

TP 01 Entrepôt de données et Analyse en ligne

Cas d'étude 1 :

On veut construire un entrepôt de données afin de stocker les informations sur les consultations d'un pays. On veut notamment connaître le nombre de consultations, par rapport à différents critères (personnes, médecins, spécialités, etc. Ces informations sont stockées dans les relations suivantes :

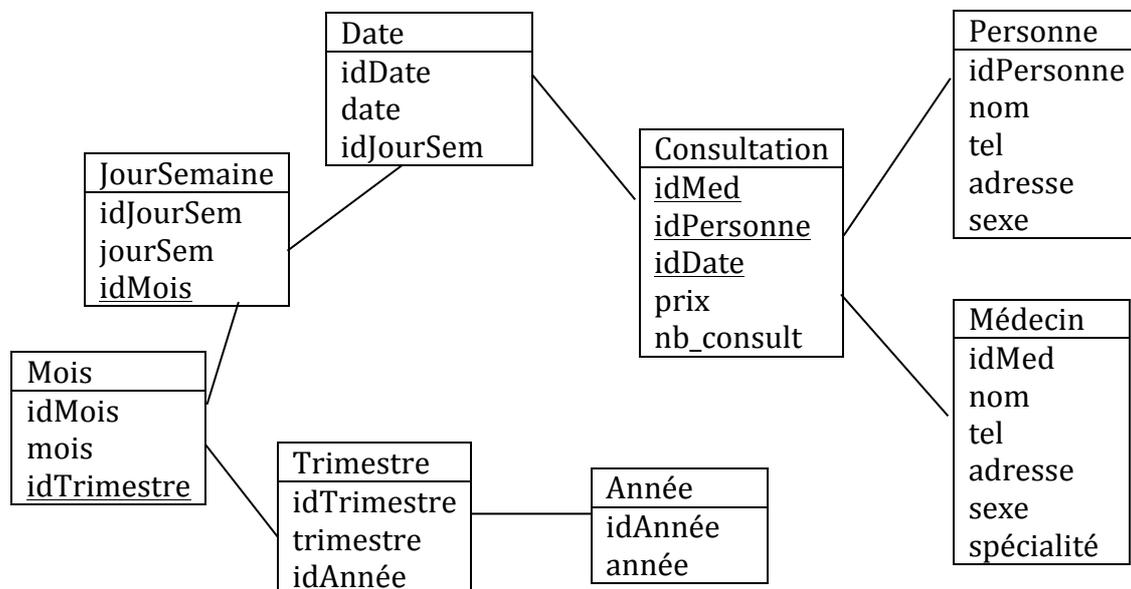
PERSONNE (id, nom, tel, adresse, sexe)

MEDECIN (id, tel, adresse, spécialité)

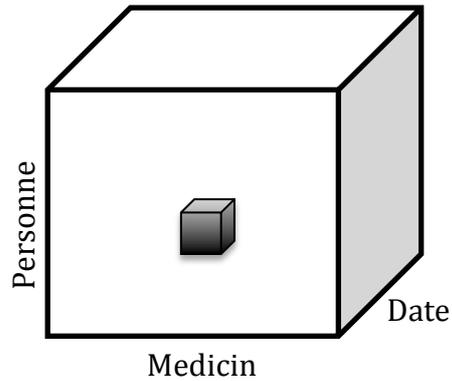
CONSULTATION (id_med, id_pers, date, prix)

Travail à faire :

1. Proposer un schéma relationnel qui tienne compte de la date, du jour de la semaine, du mois, du trimestre et de l'année.



2. Quelle est la table des faits?
Consultation
3. Quels sont les faits?
Le prix et le nombre de consultations (nb_consult)
4. Combien de dimensions ont été retenues? Quelles sont-elles?
Trois dimensions : Médecin, Personne, Temps
5. Quelles sont les hiérarchies des dimensions? Dessinez les.
Date: Année → Trimestre → Mois → JourSemaine → Date
6. Faites une représentation du cube OLAP sans tenir compte des hiérarchies.



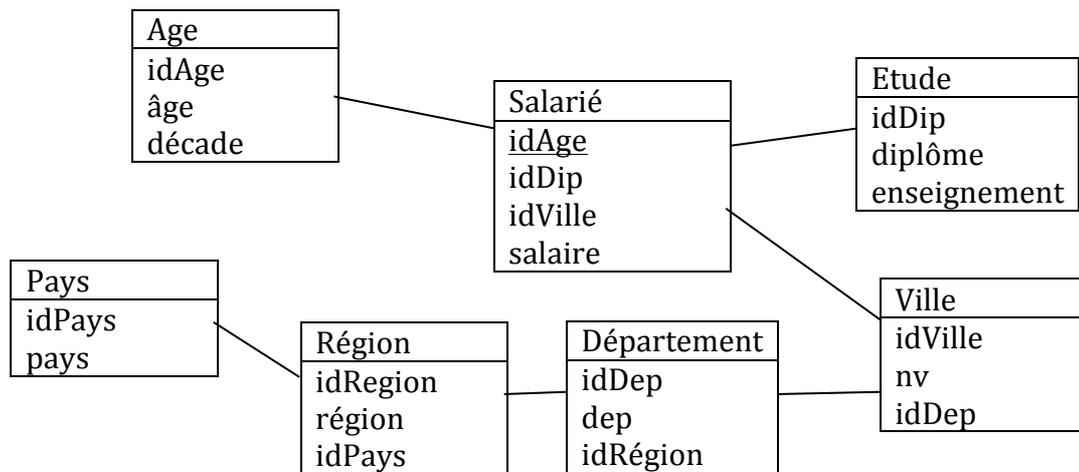
7. A partir de ce cube, indiquez quelles opérations OLAP (roll up, drill down, slice, dice) il faut appliquer pour obtenir les informations suivantes :
- a. le coût total des consultations par médecin en 2012 et 2013
 (On fait deux requêtes, une pour chaque année)
 Roll-up on temps to Année
 Slice Année = 2012
 Roll-up on temps to Année
 Slice Année = 2013
 (Coût « total », donc Slice avec une opération d'agrégation)
 - b. le nombre de consultations par jour de la semaine, par spécialité et par sexe du patient
 Roll up on Date to jourSemaine
 Roll-up on Médecin to spécialité
 Roll-up on Personne to sexe
 - c. le coût des consultations par patiente pour les mois d'octobre
 Roll-up on Date to mois
 Dice mois = octobre
 (Coût des consultations, donc Dice sans une opération d'agrégation)

Cas d'étude 2 :

On dispose d'un outil OLAP pour analyser les salaires selon l'âge et le niveau d'étude des personnes et la situation géographique des entreprises. L'analyse selon l'âge peut se faire par année ou par décade (tranches de 10 années à partir de 14 ans et jusqu'à 73 ans). L'analyse du niveau d'étude peut se faire par le niveau d'enseignement atteint en fin d'études (primaire, secondaire, supérieur) ou par le dernier diplôme obtenu (certificat de fin d'étude primaire, BEPC, Bac, Licence, Master). L'analyse de la situation géographique peut se faire par ville, département, région ou pays.

Travail à faire :

1. Quel est le schéma relationnel en étoile le plus approprié pour cette analyse.



2. Quelle est la table des faits?
Salarié
3. Quels sont les faits?
Un seul fait : salaire
4. Combien de dimensions ont été retenues? Quelles sont-elles?
Trois dimensions : Ville, Etude et Age
5. Quelles sont les hiérarchies des dimensions? Dessinez les.
Ville → Département → Région → Pays
6. Faites une représentation du cube OLAP en tenant en compte les hiérarchies.
S'inspirer des slides 45 et 46 du cours, disponible sur
http://hugo.alatristasalas.free.fr/Cours/CM-DW_OLAP.pdf