

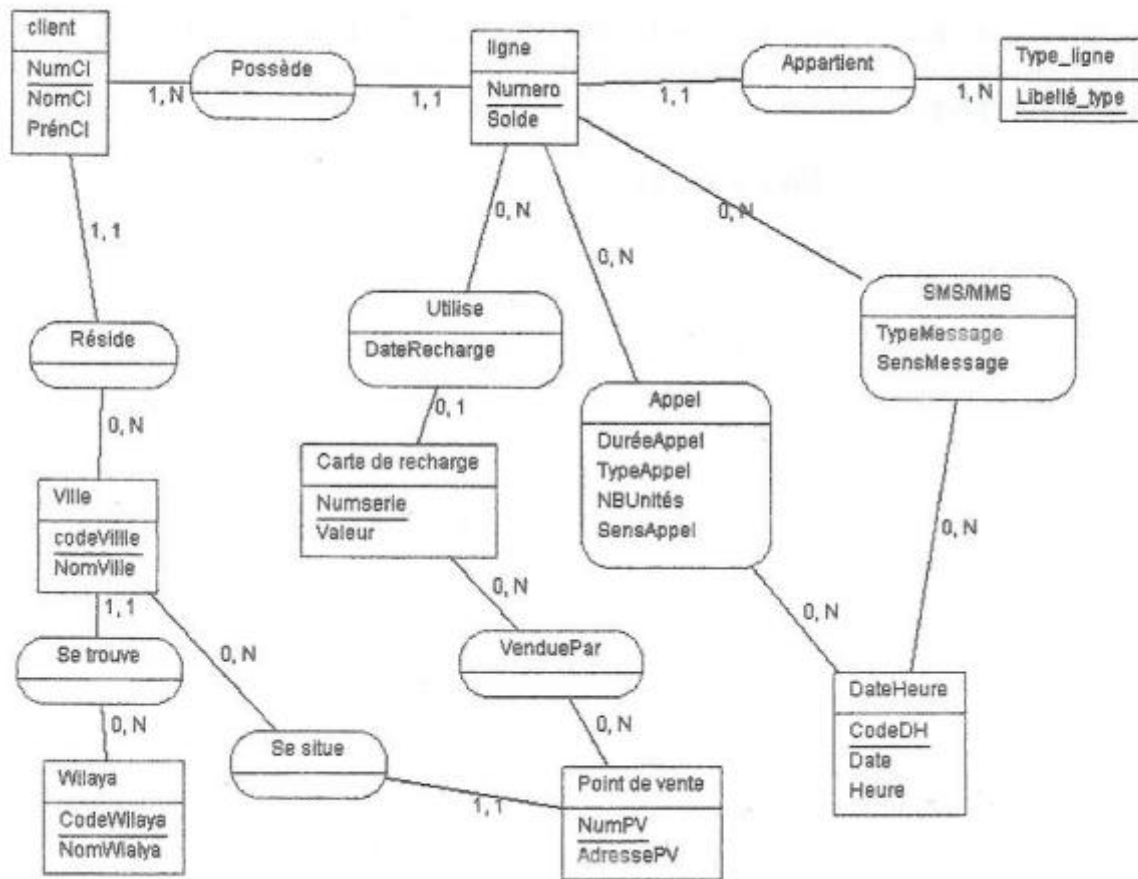
Méthodologie de Kimball

« Se base sur l'architecture des DM indépendants »

80% d'analyse 20% de dev

- **Choisir le sujet (Identifier l'activité à analyser)**
- **Choisir les faits**
 - **Le fait** : action définie par des mesures
 - fait -> 1 ou plusieurs indicateurs.
 - **Exemple de mesure** : *gain, taux*, nombre de..
 - **Formules de calculs des mesures** : COUNT, SUM, AVG etc.
 - **Type des faits** :
 - *Additif* : additionnable suivant toutes les dimensions (ex. Qté, CA)
 - *Semi additif* : additionnable suivant certaines dimensions (ex. niveau de stocke additionnel suivant le temps)
 - *Non additif* : non additionnable quel que soit la dimension (comptage de faits où affichage 1 par 1)
 - Fait instantané périodique (photos à différents instant)
 - fait transactionnel (événement ponctuel)
 - fait transactionnel intervallaire (hausse de salaire)
 - fait instantané cumulatif / récapitulatif
 - Emmagasinier les calculs préliminaires dans la table des faits
 - Stocker les pré-calculs
- **Les Identifier les dimensions**
 - Les dimensions
 - dimensions changeantes 4 types nomenclature Kimball (dim 0 invariantes, dim 1 non historisées, dim 2 historisées par création de nouvelles occurrences, dim 3 par ajout de colonne) (timestamp, fait valide)
 - Définir les hiérarchies
 - Séparer date et heure, important de garder heure
 - Snowflake -> normalization
 - **Niveau logique**
 - ROLAP → Relational OLAP
 - MOLAP → Multidimensional OLAP
 - HOLAP → Hybrid OLAP
 - OOLAP → Object OLAP
 - **Cas du ROLAP**
 - Ajouter les clés primaires au niveau des dimensions, ensuite rajouter les clés étrangères dans la table des faits
- **Choisir la durée de la base**
- **Suivre les dimensions lentement évolutives**
- **Décider des requêtes prioritaires, des modes de requêtes**

Exercice N°1 : Soit le MCD ci-dessous décrivant une partie de la base de données d'un opérateur téléphonique.



Remarque :

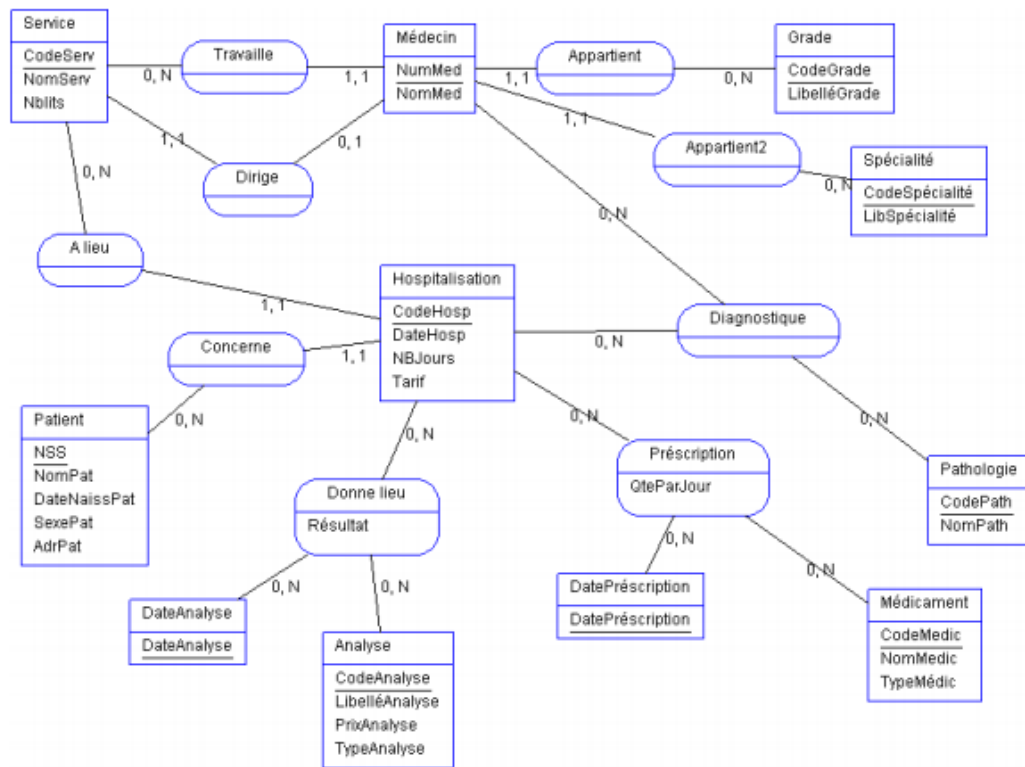
- Sens appel (rentrant/sortant).
- Type Appel : Externe (autre opérateur)/ interne.
- Libellé_type : Carte/ abonnement.
- Dans le cas d'une ligne de type abonnement, le solde correspond au montant facturé.

Questions :

1. Nous identifions deux magasins de données "suivi des appels", "suivi des SMS/MMS", proposez d'autres magasins possibles en fonction des données disponibles.
2. Proposez les modèles conceptuels des deux magasins
3. Donnez le MLD ROLAP normalisé du premier magasin
4. Donnez les scripts de réalisation de la dimension temps (vue matérialisée+ dimension)
5. Donnez le script de création de la vue matérialisée représentant le fait.

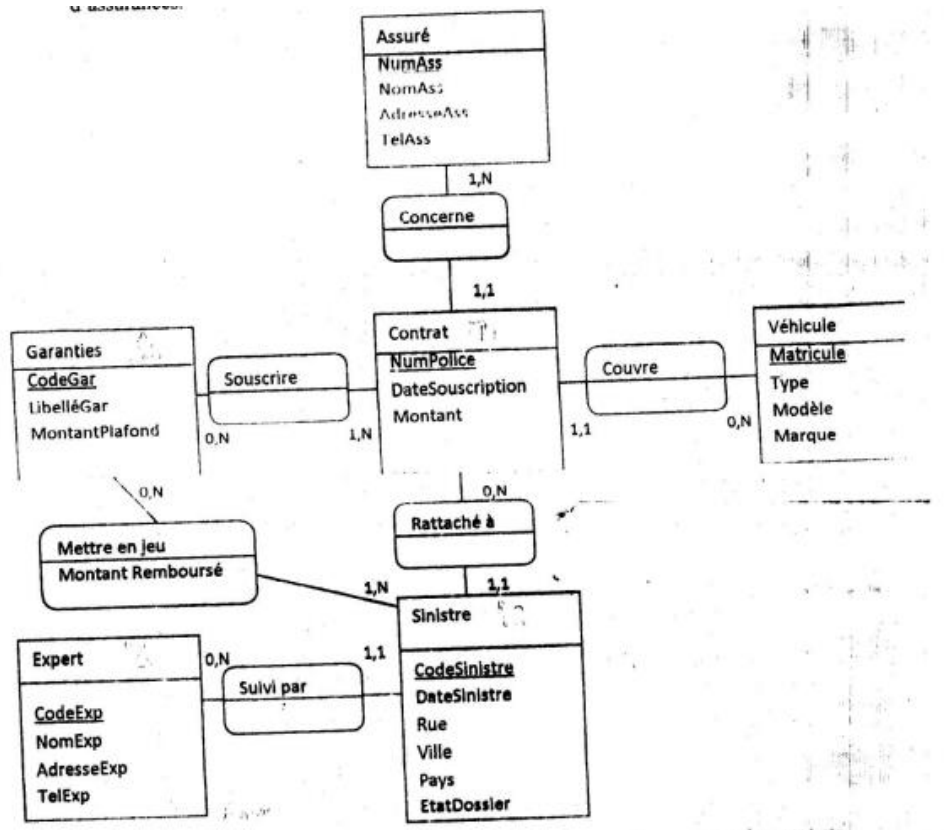
Exercice N°2 :

Considérons le MCD ci - dessous représentant un système d'information d'un hôpital :



- 1- On veut concevoir un premier magasin de données concernant l'aspect financier :
 - a. Proposez les mesures intéressantes pour analyser les résultats financiers obtenus par l'hôpital.
 - b. Quelles sont les dimensions d'analyse de chaque mesure (donner le nom et les attributs de chaque dimension) ?
 - c. Donner les hiérarchies possibles au sein de chaque dimension
 - d. Représentez le modèle conceptuel du magasin (en étoile).
 - e. Représentez le modèle logique du magasin (en étoile).
 - f. Au niveau physique, considérons que la table de fait est créée sous forme de vue matérialisée. Donnez le script de création de cette vue.
 - 2- Un autre magasin est dédié au suivi des patients par rapport aux pathologies (maladies) diagnostiquées, ainsi que les prestations données par l'hôpital (hospitalisation dans un service, médicaments administrés, analyses réalisées...etc.).
- Reprendre les questions : a, b, c, d

Exercice 03 : Soit le MCD suivant représentant une partie du système d'information d'une société d'assurances



Nous voulons concevoir deux magasins de données, le premier concerne les contrats d'assurance et le second les sinistres.

1. Proposer un modèle conceptuel pour chacun des deux magasins, en identifiant les mesures utiles pour les analyses, ainsi que les dimensions associées.
2. Donnez le modèle logique normalisé ROLAP du 1^{er} magasin.
3. Donnez le script de création d'une vue matérialisée représentant une dimension (au choix) du 1^{er} magasin.

Exercice 04 : Modélisation logique d'un entrepôt de données

Mr Leclerc directeur d'une entreprise de la grande distribution souhaite analyser et suivre les ventes de produits dans son réseau de magasins. Mr Leclerc souhaite obtenir une réponse aux questions suivantes :

- Quels produits dégagent la plus forte rentabilité dans le temps ?
- Existe-t-il des disparités régionales de consommation des produits ?
- Quelle est la répartition des ventes entre les produits de marque des fabricants et ceux de la marque du distributeur ?
- Quel est le chiffre d'affaire réalisé avec les plus gros fournisseurs ?

L'ensemble des informations seront issues des tickets de caisse. Nous identifions un certain nombre d'axes d'analyse :

- l'axe produit ;
- l'axe magasin ;
- l'axe temps ;
- l'axe département ;
- l'axe fournisseur.

Il faut ensuite décrire la hiérarchie de chacun de ces axes:

- *pour l'axe produit* : un produit appartient à une sous-famille de produits, cette sous-famille appartient à une famille de produits, cette famille appartient à une gamme de produit
- *Pour l'axe magasin*: un magasin est rattaché à une enseigne.
- *Pour l'axe fournisseur*: un fournisseur appartient à un groupe de fournisseurs.
- *Pour l'axe département*: un département est rattaché à une région, cette région est rattachée à un pays.
- *Pour l'axe temps*: un mois est rattaché à un trimestre qui est rattaché à un semestre qui est rattaché à une année.

On cherche alors à décrire les indicateurs suivants : les ventes par produit, par magasin, par fournisseur, par région et dans le temps.

Question 1 : Quelles sont les différentes hiérarchies

Question 2 : Construire un schéma en étoile et en flocon de neige pour ce cahier des charges.