

## CHAPITRE 3

# IMPLANTATION D'UN DWH

## Fondement data warehouse

2BC  
2020-2021

## **Utilisation d'un SGBD Relationnel (systèmes ROLAP)**

- *SGBDR : Nécessité des adaptations pour répondre aux besoins des ED*
- *Stockage des données dans un SGBDR*
- *Utilisation d'un middle-ware pour implémenter les opérations spécifiques de l'OLAP*

## **Utilisation d'un SGBD Multidimensionnel (systèmes MOLAP)**

- *SGBD capable de stocker et traiter des données multidimensionnelles*
- *Basé sur un stockage par tableau (technique des matrices creuses)*
- *Indexation rapide des données calculées*

## **Utilisation d'un SGBD Hybride (systèmes HOLAP)**

*Tirer profit des avantages des technologies ROLAP et MOLAP :*

- *un ROLAP pour stocker et gérer les données détaillées*
- *un MOLAP pour stocker et gérer les données agrégées*

# PRINCIPE

## **Analyse des besoins clients**

- a. Exprimer les besoins sous la forme de requêtes décisionnelles*
- b. Réaliser les vues hiérarchiques pour chaque requête*

## **Analyse des données**

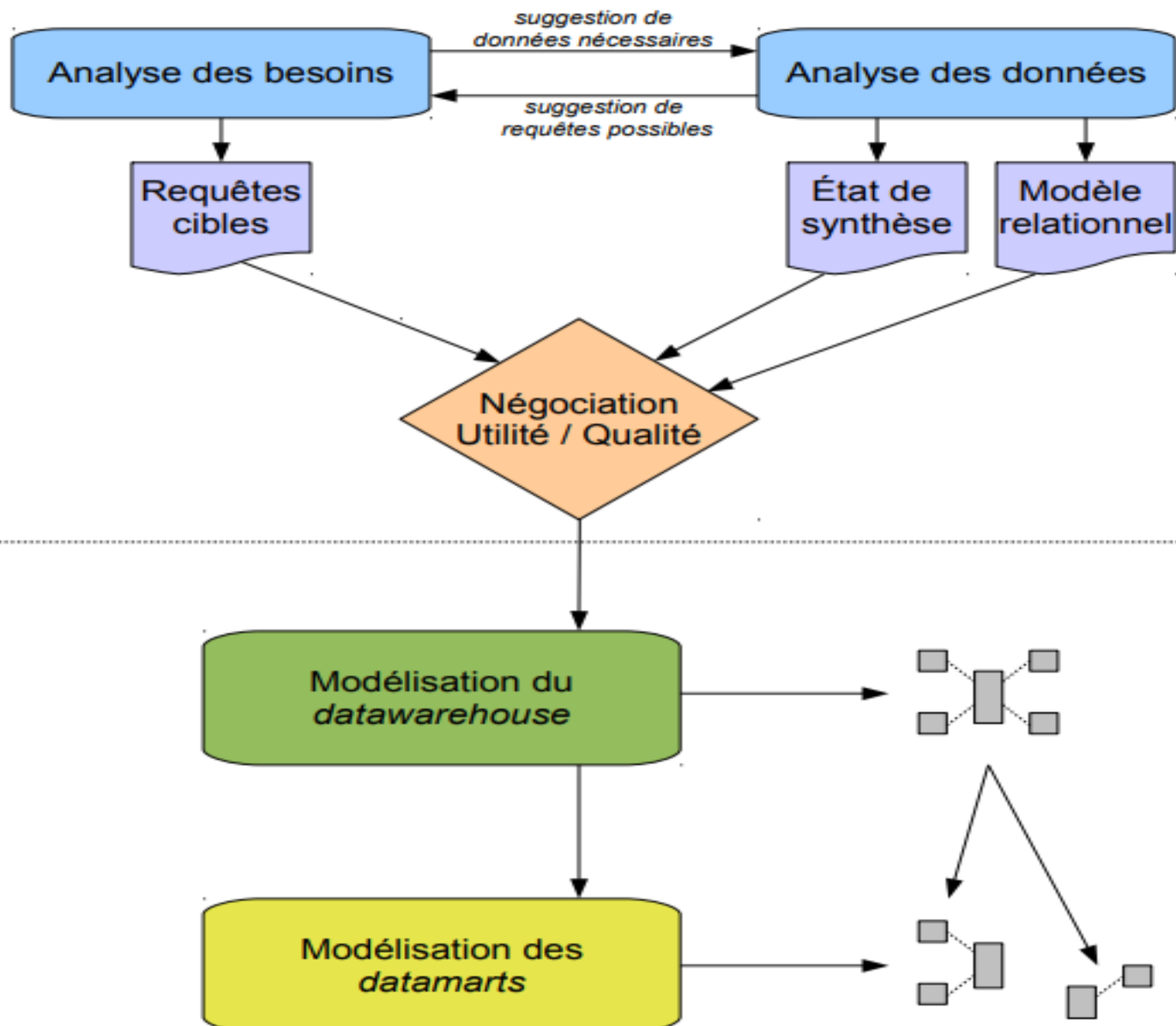
- a. Étude des sources de données (quantification, analyses générales)*
- b. Qualification des données (qualité et intérêt)*
- c. Intégration logique des données (simulation d'un schéma relationnel virtuel)*
- d. Normalisation du schéma virtuel en 3NF pour en avoir une vue cohérente*

**Sélectionner les requêtes qui seront effectivement réalisables en fonction des données disponibles**

## **4. Conception du data warehouse et des data marts**

- a. Séparer les requêtes en fonction de la granularité de la table des faits (grain fin des ventes, grain plus grossier du ticket de caisse, etc.)*
- b. Créer un data warehouse intégrant toutes les requêtes de grain fin*
- c. Extraire un data mart par niveau de grain supérieur et/ou pour des thématiques particulières nécessitant par exemple une pré-agrégation*

# CONCEPTION LOGIQUE D'UN DWH



# CONCEPTION LOGIQUE D'UN DWH

- Définition des objets; Définition des relations entre objets

**=> Choix d'un modèle de conception (schéma)**

Utilisation, par exemple, d'Oracle Designer ou Oracle WareHouse Builder



**Analyse des besoins :**  
ce que je souhaite étudier.

**Analyse des données disponibles :**  
ce que je peux étudier.

**ROLAP** : schéma de BD relationnelle reflétant la vue de l'analyste

- *Multidimensionnelle*
- *Hiérarchisée*
- *schéma en étoile (star schema)*
- *schéma en flocon (snowflake schema)*
- *constellation de faits (fact constellation)*

NB: le schéma en **étoile** est souvent utilisé pour l'implantation **physique**

Structure simple utilisant le modèle entité-relation

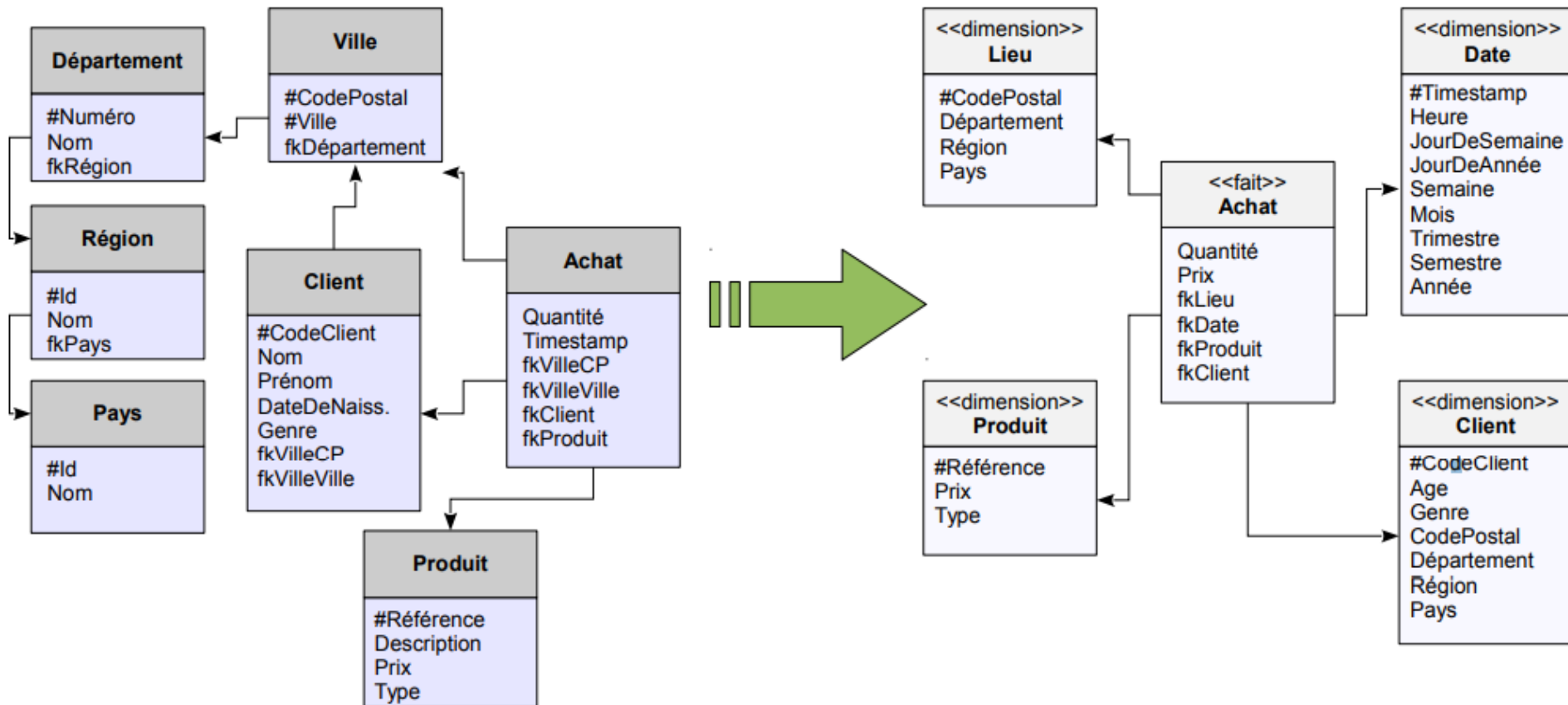
**une entité/table centrale (table des faits)**

- objets de l'analyse
- taille très importante
- beaucoup de champs

**des entités/tables périphériques (tables de dimensions)**

- critères/dimension de l'analyse
- taille peu importante
- peu de champs

# SCHÉMA EN ÉTOILE





## Evolution du schéma en étoile

- Décomposition des dimensions du modèle en étoile en sous-hiérarchies
- Conservation du fait
- Eclatement des dimensions suivant leur hiérarchie des paramètres
- Normalisation des tables de dimensions
  - Structure hiérarchique des dimensions
  - Un niveau inférieur identifie un niveau supérieur

**Chaque dimension du schéma en étoile précédent est dénormalisée**

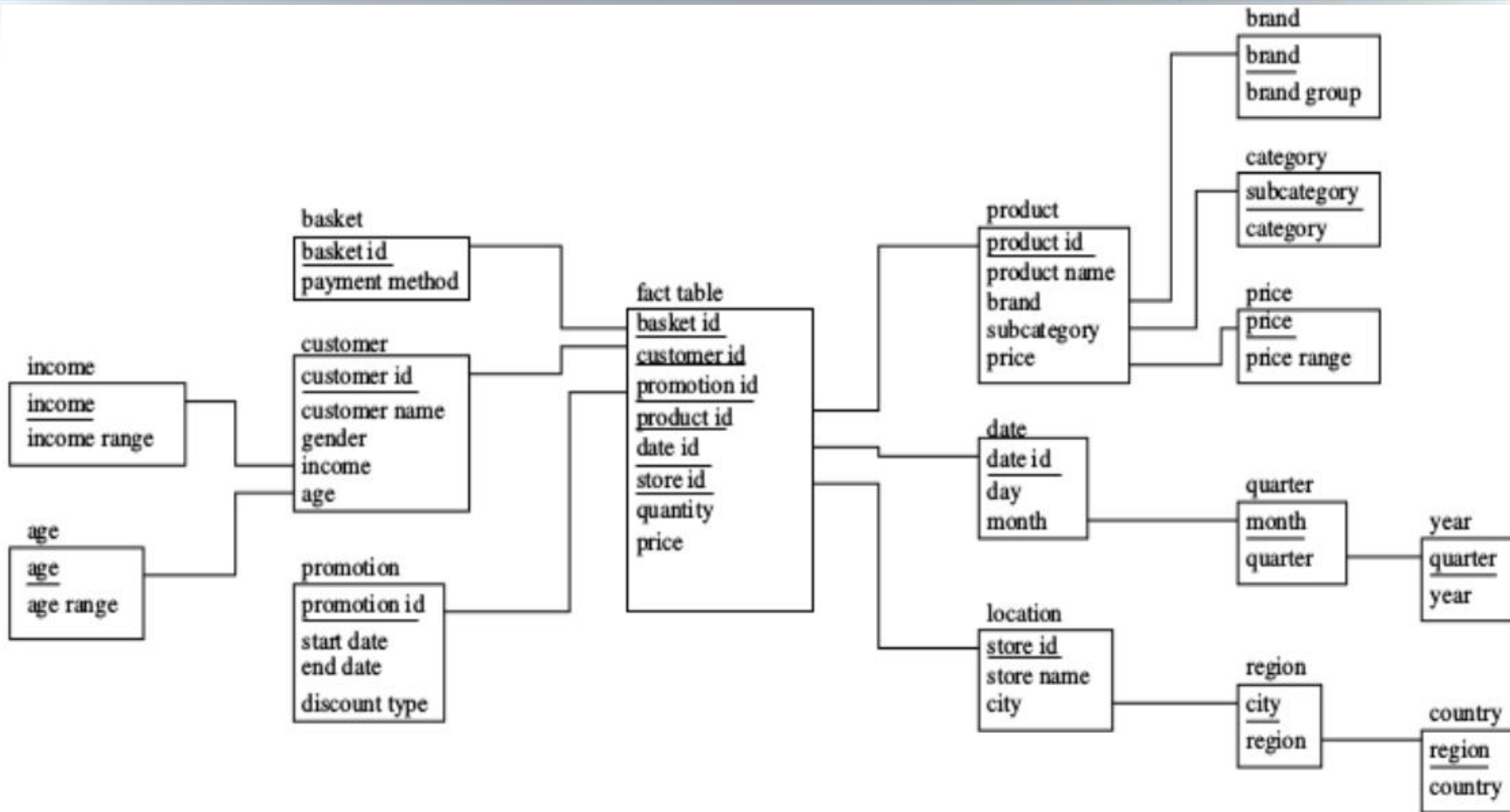
## Avantages

- Formalisation d'une hiérarchie au sein d'une dimension
- Maintenance des tables de dimensions simplifiée
- Réduction de la redondance

## Inconvénients

- complexité en termes de lisibilité et de gestion
- Navigation coûteuse

# LES CUBES



Chaque dimension du schéma en étoile est dénormalisée

## Modèle en constellation :

- Fusion de plusieurs modèles en étoile qui utilisent des dimensions communes
- Enregistrement de plusieurs faits avec des dimensions communes ou non

## Généralisation du schéma en étoile

- Plusieurs tables des faits
- Partage de tables de dimensions

=> En général, on a un schéma de constellation de faits pour l'entrepôt  
une étoile de la constellation pour un magasin de données (*Data Mart*)

## **Agrégation des faits selon une ou plusieurs dimensions**

2 moyens de les représenter :

**Cas 1:** une table des faits séparée/dédiée avec les tables pour les dimensions correspondantes

**Cas 2:** dans la même table des faits, en codant les niveaux hiérarchiques dans les tables de dimensions

## Cas 1:

- `faits1(idProduit, idVille, idJour, 5)`
- `faits2(idProduit, idVille, idMois, 60)`

avec une table **jour** et une table **mois**

## Cas 2:

- `faits(idProduit, idVille, idDate1, 5)`
- `faits(idProduit, idVille, idDate2, 5)`

avec une table date contenant

- `date(idDate1, 22, 01, 2010)`
- `date(idDate2, ALL, 01, 2010)`