

Systemes d'information décisionnels (Data Warehouse)

Dr Dendani-Hadiby Nadjette
Université Badji Mokhtar Annaba
Département d'Informatique
n_dendani@yahoo.fr

2021/2022

Plan

1. Introduction

- » Problématique- Le Système d'Information - La Suite Décisionnelle

2. L'Entrepôt de Données

- Définition
- Différence avec un SGBD
- Caractéristiques

3. Architecture d'un système décisionnel

4. Modélisation multidimensionnelle

- Niveau conceptuel
- Niveau logique
- Niveau physique

1. Introduction - Problématique

- Objectif

Comment répondre aux besoins de décideurs afin d'améliorer les performances décisionnelles de l'entreprise?

» En répondant aux demandes d'analyse des décideurs

Exemple:

- ✓ clientèle : Qui sont mes clients ? Pourquoi sont-ils mes clients ? Comment les conserver ou les faire revenir ? Ces clients sont-ils intéressants pour moi ?

- ✓ marketing, actions commerciales : Où placer ce produit dans les rayons ? Comment cibler plus précisément le mailing concernant ce produit ?

1. Introduction - Problématique

- En donnant un accès rapide et simple à l'information stratégique
- En donnant du sens aux données
- En donnant une vision transversale des données de l'entreprise (intégration de différentes bases de données)
- En extrayant, groupant, organisant, corrélant et transformant (résumé, agrégation) les données.

1. Introduction - Problématique

- Une grande masse de données :
 - » Distribuée
 - » Hétérogène
 - » Très Détaillée
- A traiter :
 - » Synthétiser / Résumer
 - » Visualiser
 - » Analyser
- Pour une utilisation par :
 - » des experts et des analystes d'un métier
 - » NON informaticiens
 - » NON statisticiens

1. Introduction –Le système d'Information Décisionnel

Moyen d'atteindre ces objectifs :

Le Data Warehouse, un système d'information dédié aux applications décisionnelles

- **En Aval** des bases de production

(ie bases opérationnelles)

- **En Amont** des prises de décision

» basé sur des indicateurs (Key Business Indicators (KBI))

Transformer des données de production en informations stratégiques

données
run the business

informations
manage the business



2. L'Entrepôt de Données (Data Warehouse)

- **Définition de Bill Inmon (1996)**

«Le Data Warehouse est une collection de données orientées sujet, intégrées, non volatiles et historisées, organisées pour le support d'un processus d'aide à la décision.»

- **Principe**

- Base de Données utilisée à des fins d'analyse.

- Caractéristiques :

- orientation sujets («métiers»)
- données intégrées
- données non volatiles
- données datées

Pourquoi pas un SGBD ? (1)

- Fonctions d'un SGBD :
 - Systèmes transactionnels (OLTP)
 - Permettre d'insérer, modifier, interroger rapidement, efficacement et en sécurité les données de la base
 - Sélectionner, ajouter, mettre à jour, supprimer des tuples
 - Répondre à de nombreux utilisateurs simultanément

Pourquoi pas un SGBD ? (2)

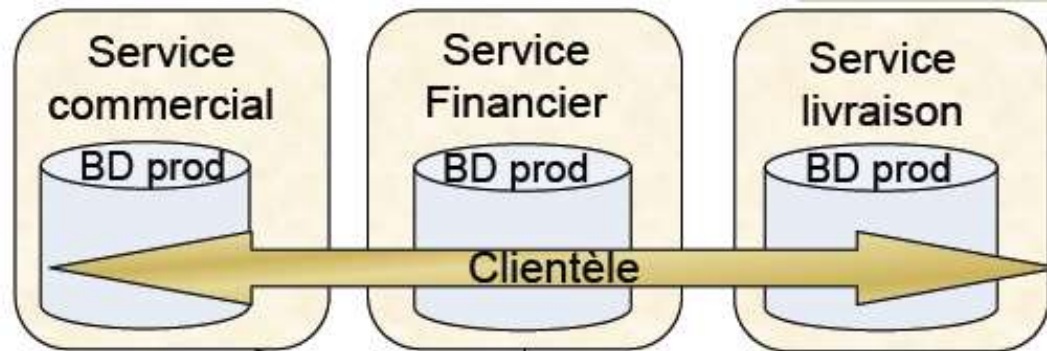
- Fonctions d'un DW :
 - Systèmes pour l'aide à la prise de décision (OLAP)
 - Regrouper, organiser des informations provenant de sources diverses
 - Intégrer et stocker les données pour une vue orientée métier
 - Retrouver et analyser l'information rapidement et facilement

Pourquoi pas un SGBD ? (3)

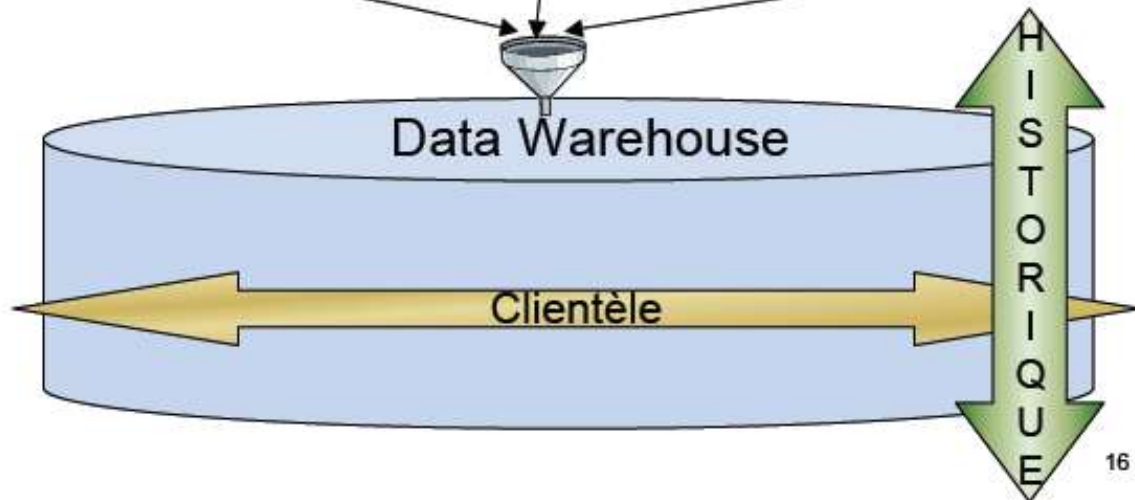
	OLTP	DW
Utilisateurs	Nombreux Employés	Peu Analystes
Données	Alphanumériques Détaillées / atomiques Orientées application Dynamiques	Numériques Résumées / agrégées Orientées sujet Statiques
Requêtes	Prédéfinies	« one-use »
Accès	Peu de données (courantes)	Beaucoup d'informations (historisées)
But	Dépend de l'application	Prise de décision
Temps d'exécution	Court	Long
Mises à jour	Très souvent	Périodiquement

Pourquoi pas un SGBD ? (4)

OLTP: On-Line
Transactional
Processing



OLAP: On-Line
Analitical
Processing



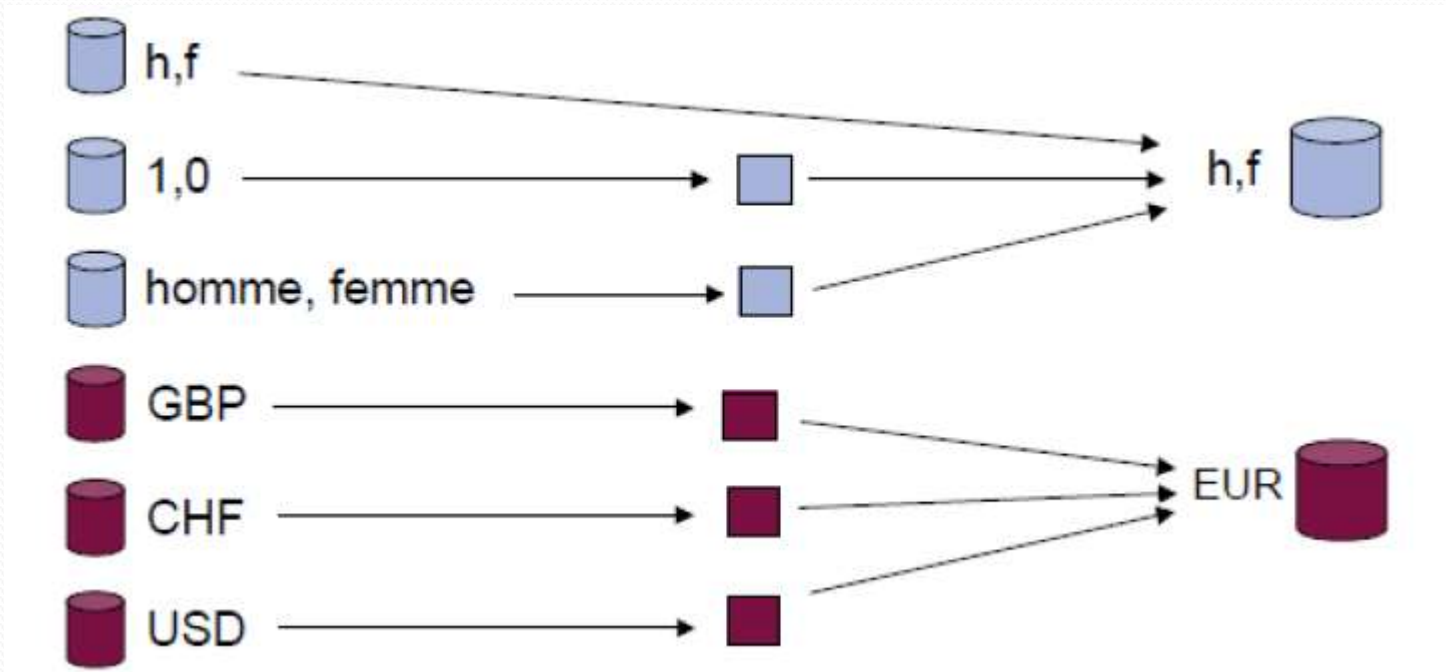
Caractéristiques d'un DW (1)

- **Données orientées sujet**
 - Regroupe les informations des différents métiers
 - Ne tiens pas compte de l'organisation
 - fonctionnelle des données



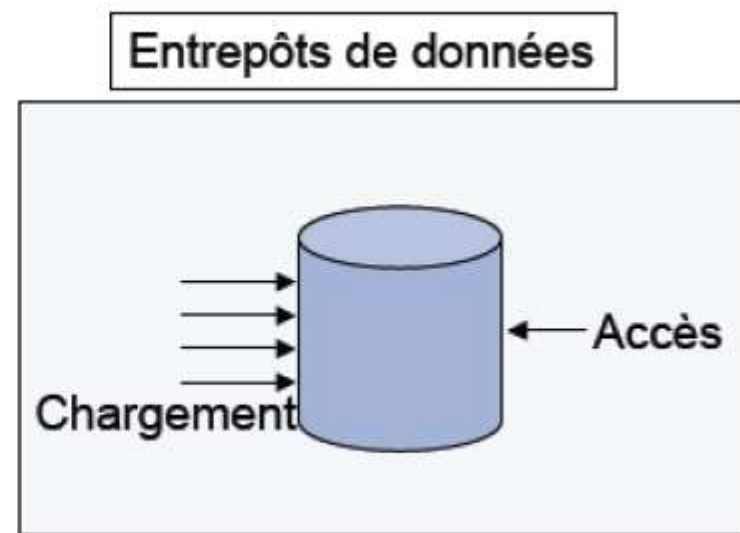
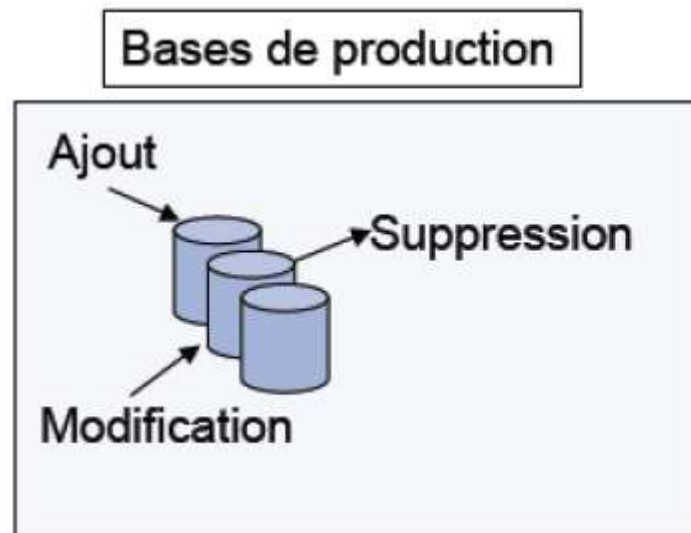
Caractéristiques d'un DW (2)

- **Données intégrées**
 - Normalisation des données
 - Définition d'un référentiel unique



Caractéristiques d'un DW (3)

- **Données non volatiles**
 - Traçabilité des informations et des décisions prises
 - Copie des données de production



Caractéristiques d'un DW (4)

- **Données historisées / datées**
 - Les données persistent dans le temps
 - Mise en place d'un référentiel temps

Base de production

Image de la base en Mai 2005

Répertoire

Nom	Ville
Dupont	Paris
Durand	Lyon

Image de la base en Juillet 2006

Répertoire

Nom	Ville
Dupont	Marseille
Durand	Lyon

Entrepôt de données

Calendrier

Code	Année	Mois
1	2005	Mai
2	2006	Juillet

Répertoire

Code	Année	Mois
1	Dupont	Paris
1	Durand	Lyon
2	Dupont	Marseille

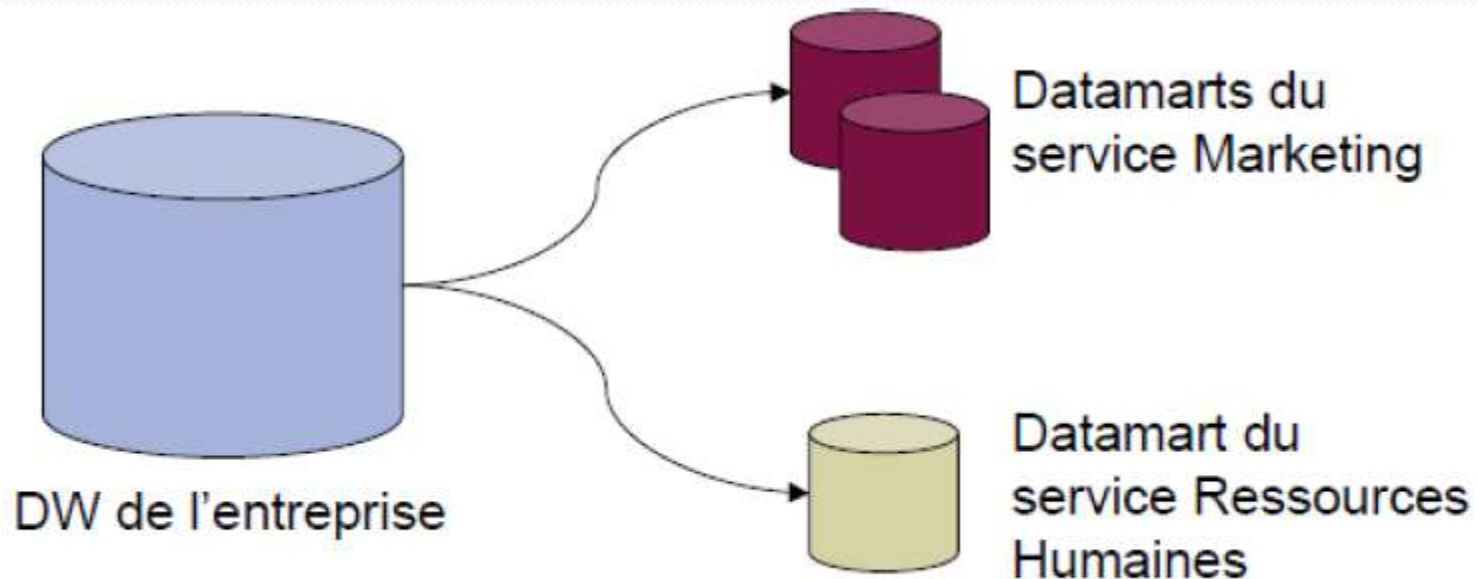
Caractéristiques d'un DW (5)

- **Inconvénient :**

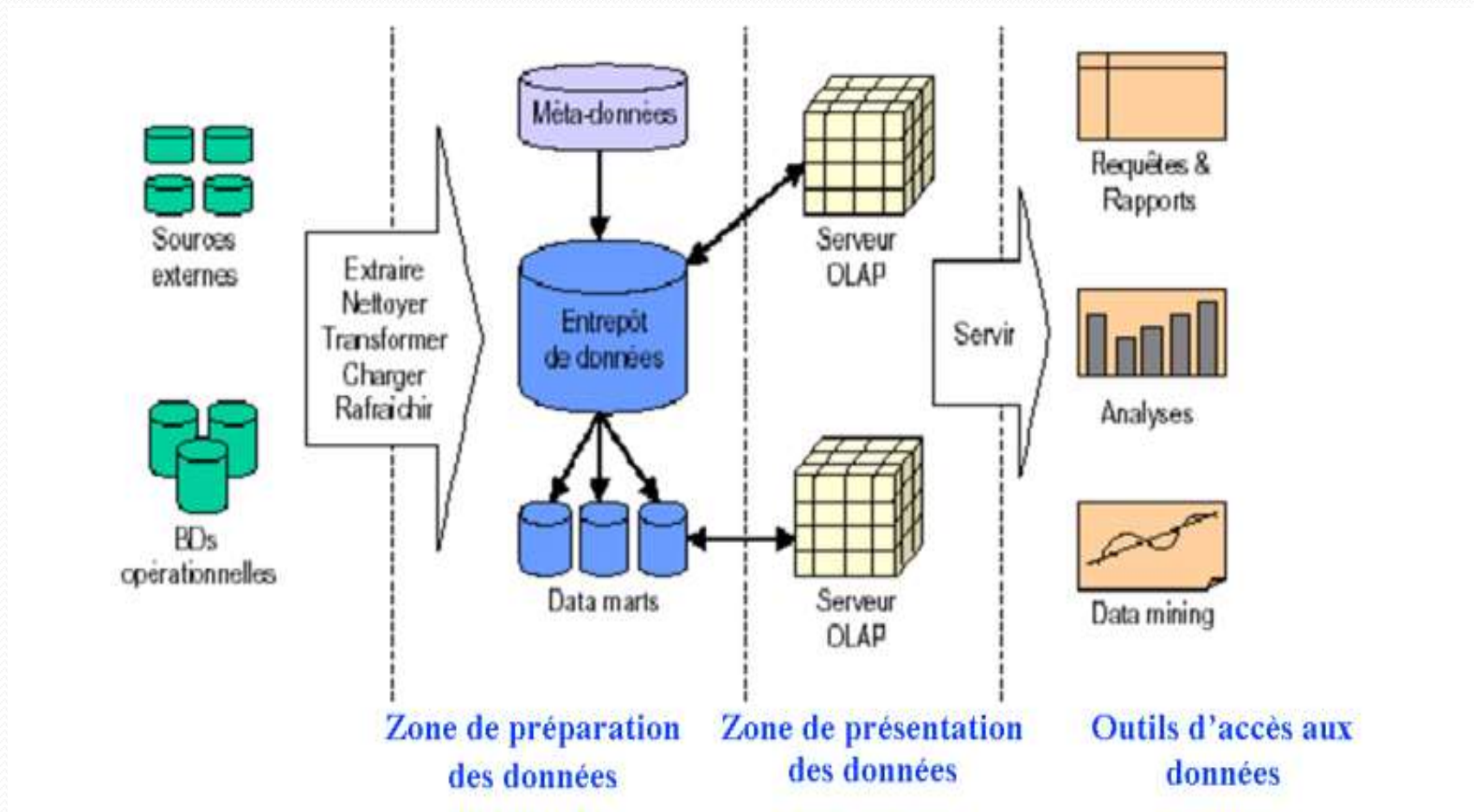
De par sa taille, le DW est rarement utilisé directement par les décideurs car il contient plus que nécessaire pour une classe de décideurs

Le datamart

- Sous-ensemble d'un entrepôt de données
- Destiné à répondre aux besoins d'un secteur ou d'une fonction particulière de l'entreprise
- Point de vue spécifique selon des critères métiers



3. Architecture d'un système décisionnel



4. Modélisation multidimensionnelle

Niveau conceptuel, Niveau logique, Niveau physique

- Niveau conceptuel

Description de la base multidimensionnelle
indépendamment des choix d'implantation

- Les concepts:
- Dimensions et hiérarchies
- Faits et mesures

Modélisation multidimensionnelle

Dimension (1)

- Axes d'analyse avec lesquels on veut faire l'analyse
Géographique, temporel, produits, etc.
- Chaque dimension comporte un ou plusieurs attributs/membres
- **Une dimension est tout ce qu'on utilisera pour faire nos analyses.**
- Chaque membre de la dimension a des caractéristiques propres et est en général textuel
- Remarque importante :
tables de dimension << Table de fait

Modélisation multidimensionnelle

Dimension (2)

Clé de substitution

Attributs de la dimension

Dimension produit

Clé produit (CP)

Code produit

Description du produit

Famille du produits

Marque

Emballage

Poids

Modélisation multidimensionnelle

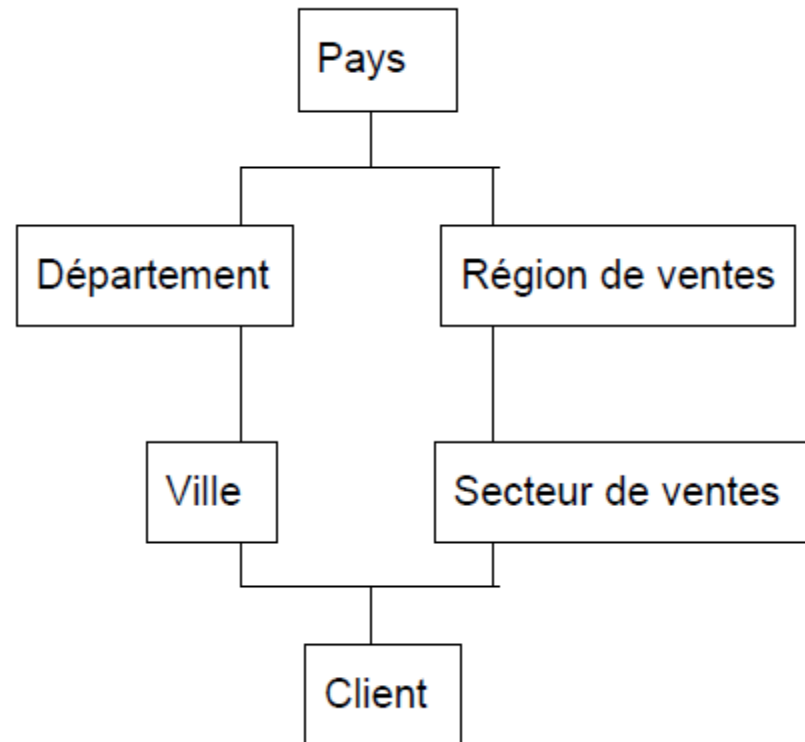
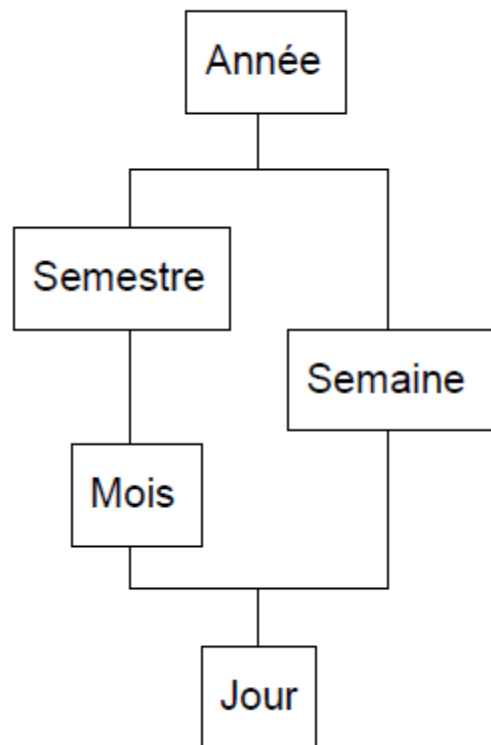
Hiérarchie (1)

- Les attributs/membres d'une dimension sont organisés suivant des hiérarchies
 - Chaque membre appartient à un niveau hiérarchique (ou niveau de granularité) particulier
 - Exemples :
 - Dimension temporelle : jour, mois, année
 - Dimension géographique : magasin, ville, région, pays
 - Dimension produit : produit, catégorie, marque, etc.
- Attributs définissant les niveaux de granularité sont appelés paramètres
- Attributs informationnels liés à un paramètre sont dits attributs faibles

Modélisation multidimensionnelle

Hiérarchie (2)

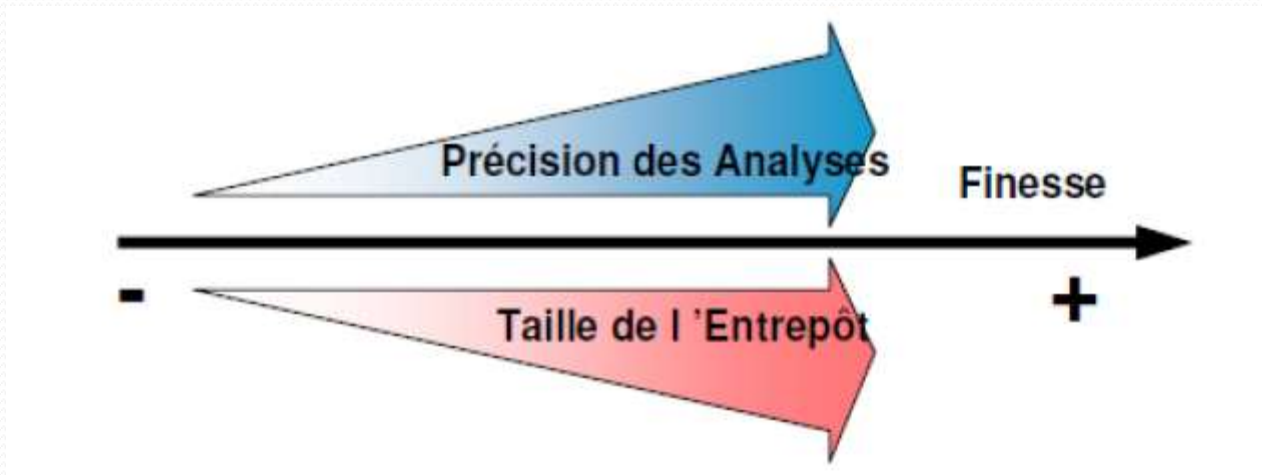
- Hiérarchies multiples dans une dimension



Modélisation multidimensionnelle

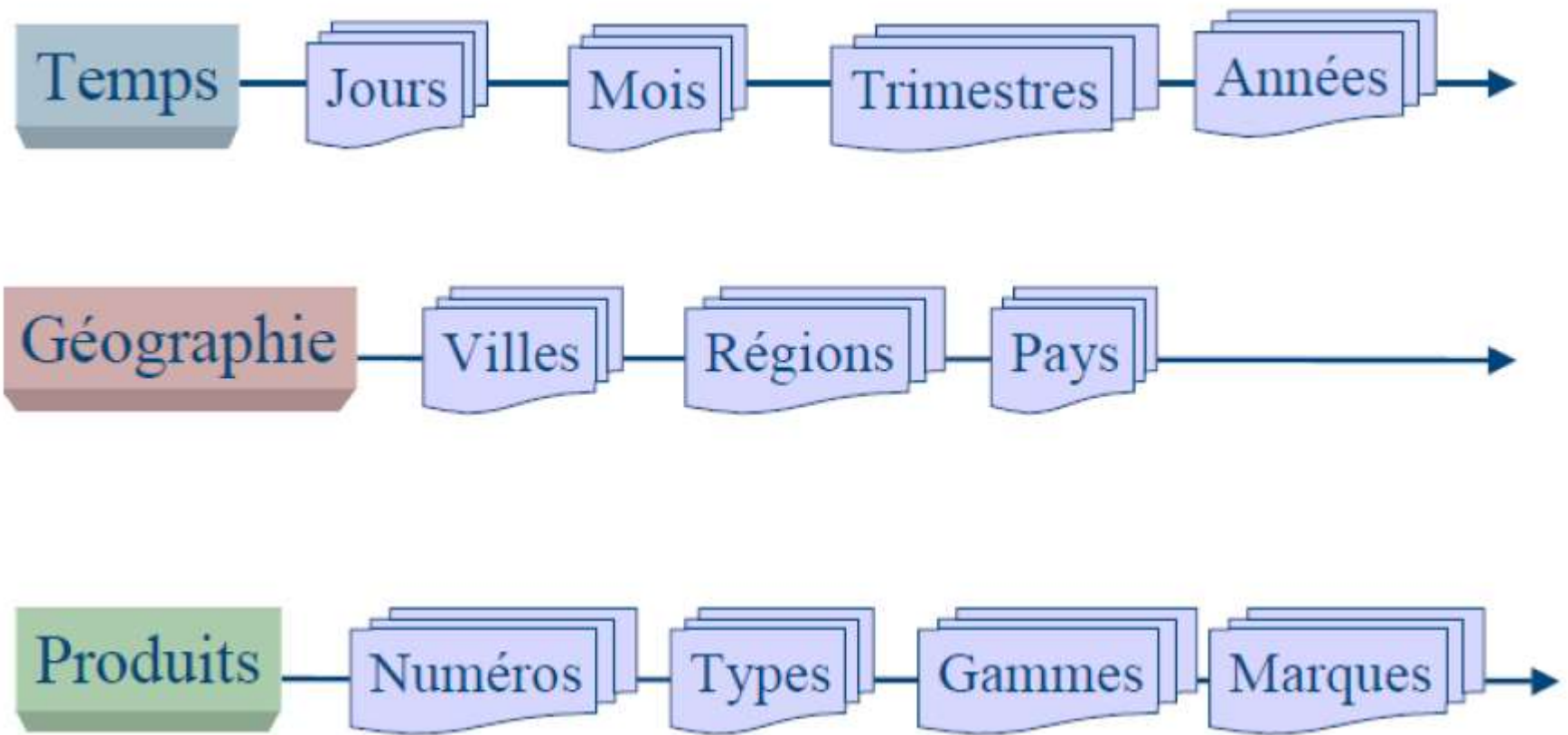
Granularité (1)

- Niveau de détail de représentation
 - Journée > heure du jour
 - Magasin > rayonnage
- Choix de la granularité



Modélisation multidimensionnelle

Granularité (2)



Modélisation multidimensionnelle

Fait

- Sujet analysé
- un ensemble d'attributs appelés mesures (informations opérationnelles)
 - les ventes (chiffre d'affaire, quantités et montants commandés, volumes des ventes, ...)
 - les stocks (nombre d'exemplaires d'un produit en stock, ...),
 - les ressources humaines (nombre de demandes de congés, nombre de démissions, ...).
- Un fait représente la valeur d'une mesure, calculée ou mesurée, selon un membre de chacune des dimensions
- **Un fait est tout ce qu'on voudra analyser.**

Exemple : 250 000 euros est un **fait** qui exprime la valeur de **la mesure Coût des travaux** pour le membre **2002** du **niveau Année** de la dimension **Temps** et le membre **Versailles** du **niveau Ville** de la dimension **Découpage administratif**.
- La table de fait contient les valeurs **des mesures** et **les clés** vers les tables de dimensions

Modélisation multidimensionnelle

Mesure

- Élément de donnée sur lequel portent les analyses, en fonction des différentes dimensions.
- Ces valeurs sont le résultat d'opérations d'agrégation sur les données

Exemple :

- Coût des travaux
- Nombre d'accidents
- Ventés
- ...

Modélisation multidimensionnelle

Clés

- Tables de dimension
 - Clé primaire
- Tables de fait
 - Clé composée
 - Clés étrangères des tables de dimension

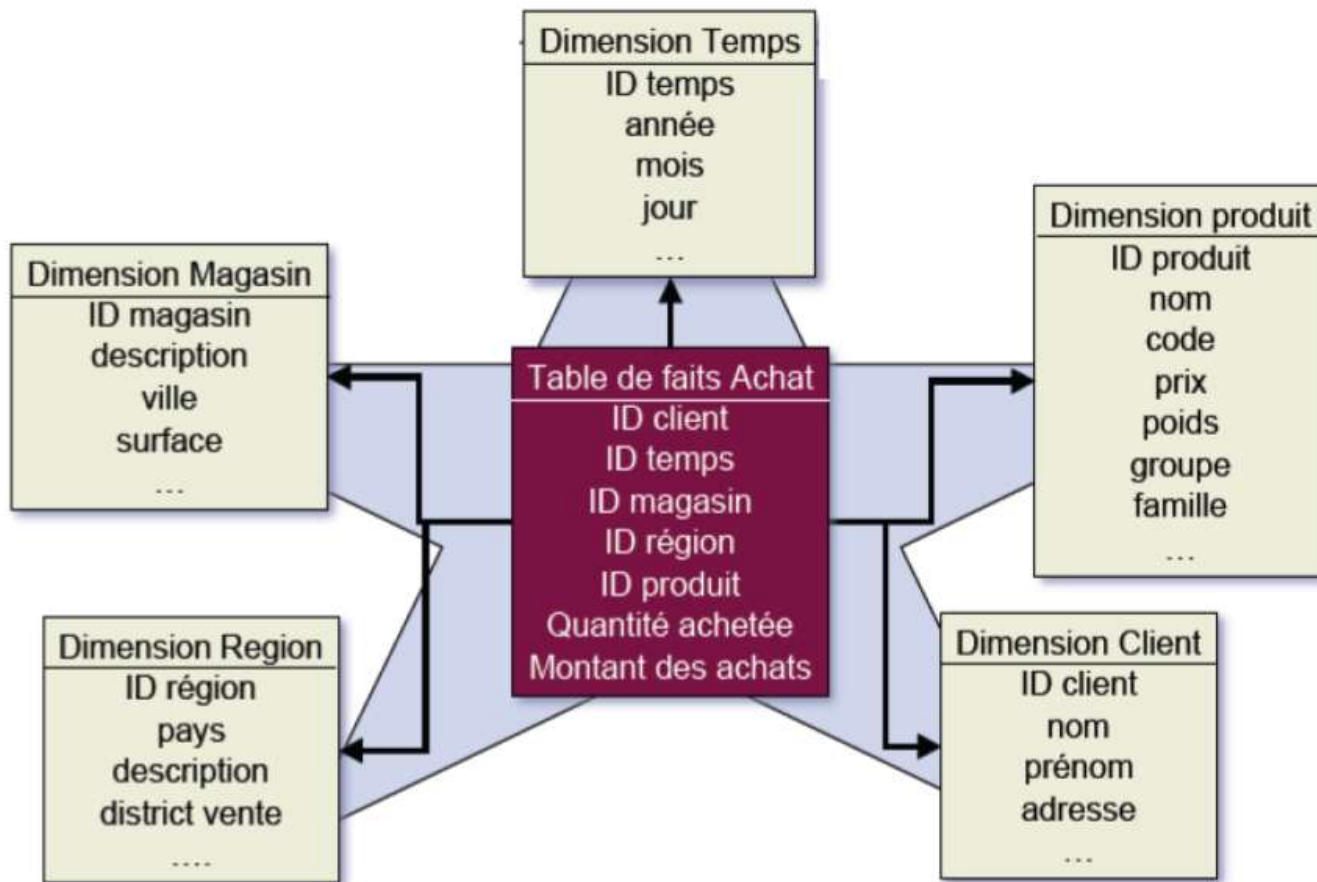
Modélisation

- Au niveau conceptuel, il existe 2 modèles :
 - en étoile (star schema)
 - en constellation (fact constellation schem)
- Au niveau logique, il existe 1 modèle :
 - en flocon (snowflake schema)

Modèle en étoile (1)

- Une table de fait centrale et des dimensions
- Les dimensions n'ont pas de liaison entre elles
- Avantages :
 - Facilité de navigation
 - Nombre de jointures limité
- Inconvénients :
 - Redondance dans les dimensions
 - Toutes les dimensions ne concernent pas les

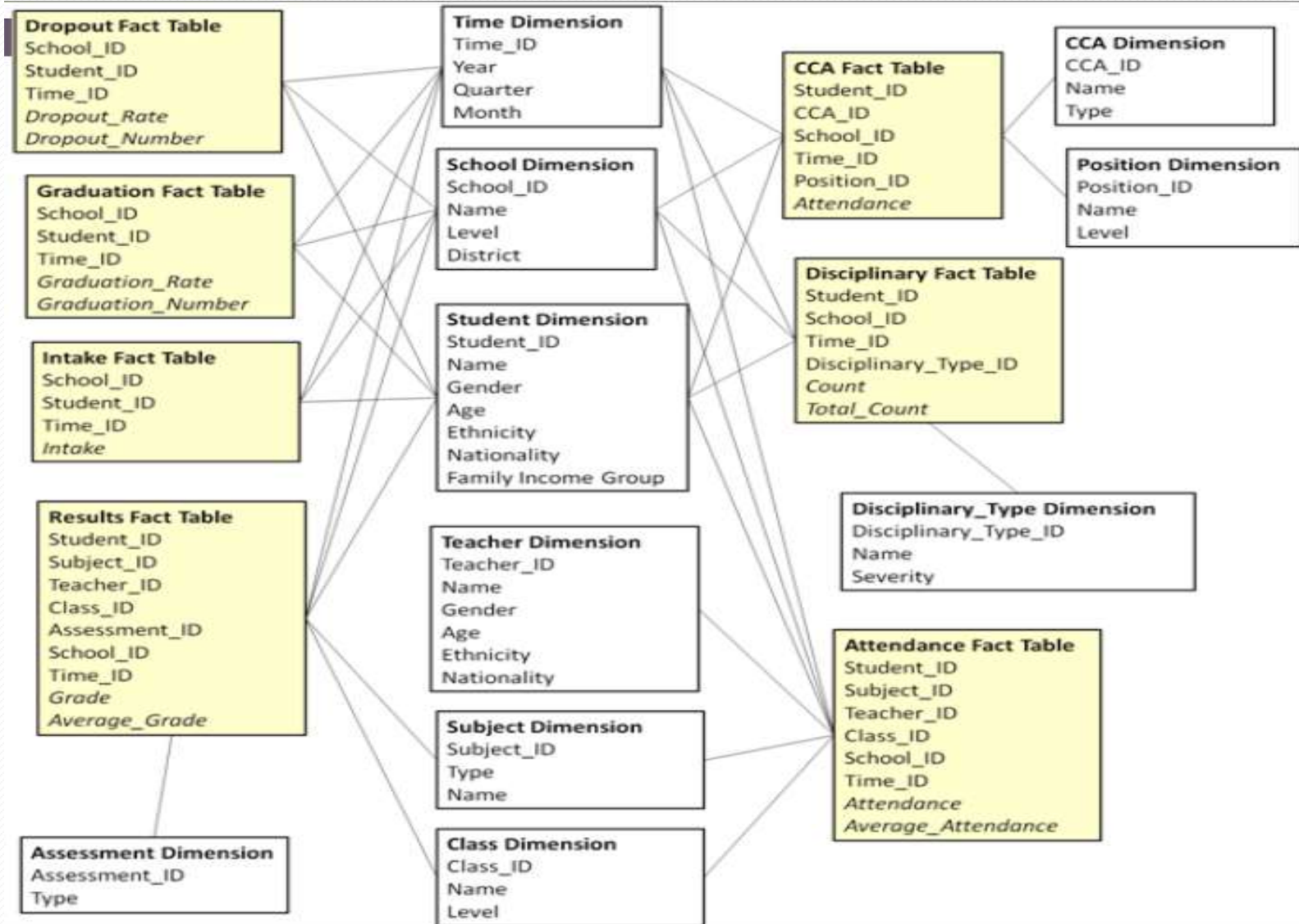
Modèle en étoile (2)



Constellation (1)

- Série d'étoiles
- Fusion de plusieurs modèles en étoile qui utilisent des dimensions communes
- Plusieurs tables de fait et tables de dimensions, éventuellement communes

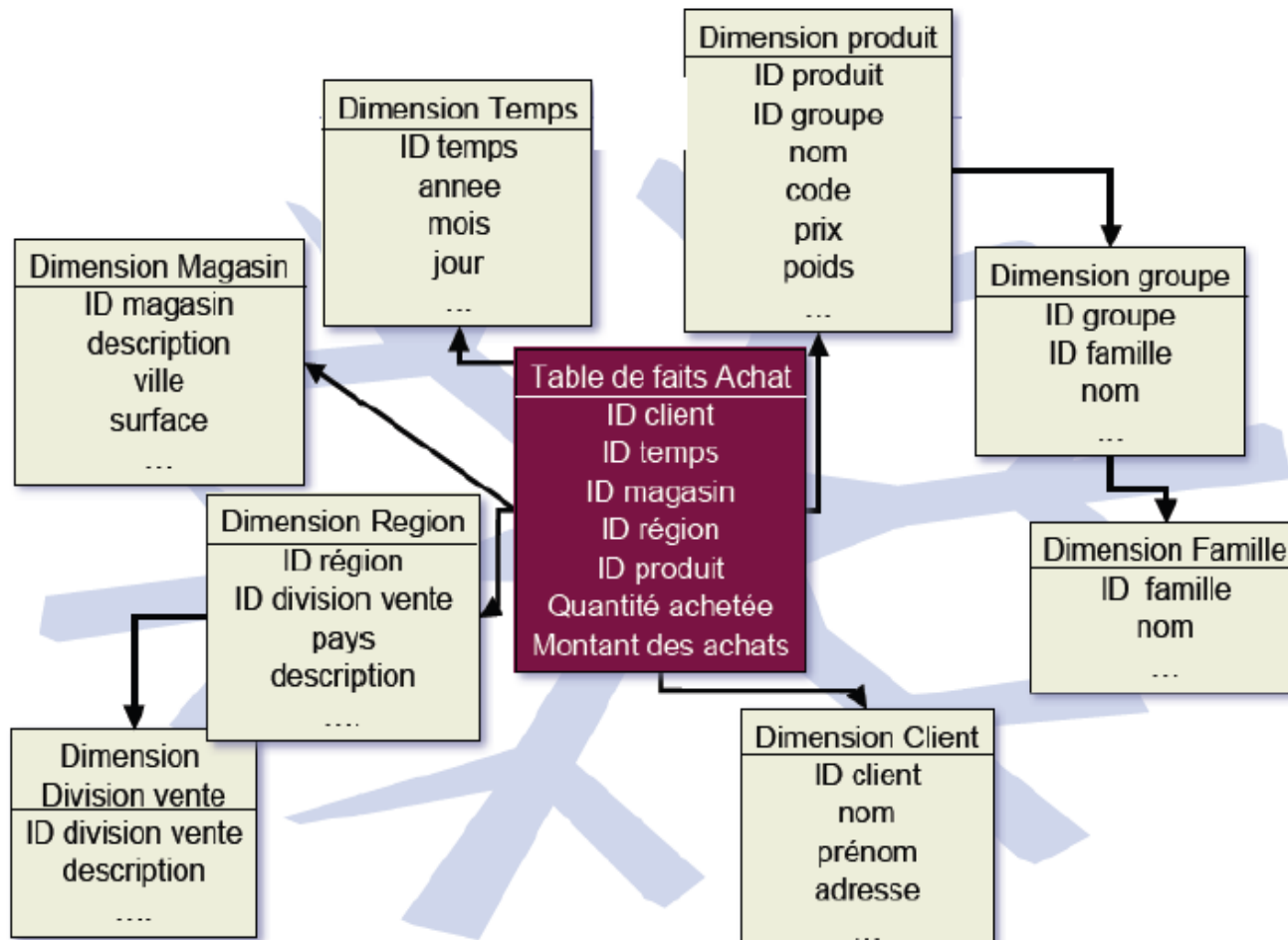
Constellation (2)




Modèle en flocon (1)

- Modèle en étoile + normalisation des dimensions
 - Une table de fait et des dimensions en sous-hiérarchies
 - Un seul niveau hiérarchique par table de dimension
 - La table de dimension de niveau hiérarchique le plus bas est reliée à la table de fait (elle a la granularité la plus fine)
- Avantages :
 - Normalisation des dimensions
 - Economie d'espace disque (réduction du volume)
- Inconvénients :
 - Modèle plus complexe (nombreuses jointures)
 - Requêtes moins performantes

Modèle en flocon (2)





Fin