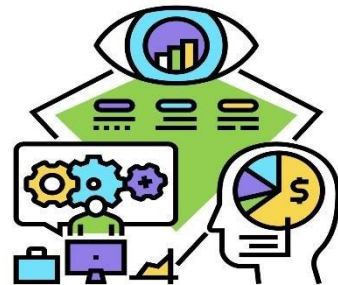


# Partie 2:

## Methodologies de Gestion de projet BI

**BUSINESS  
INTELLIGENCE**



# Méthodes appliquées pour un projet BI

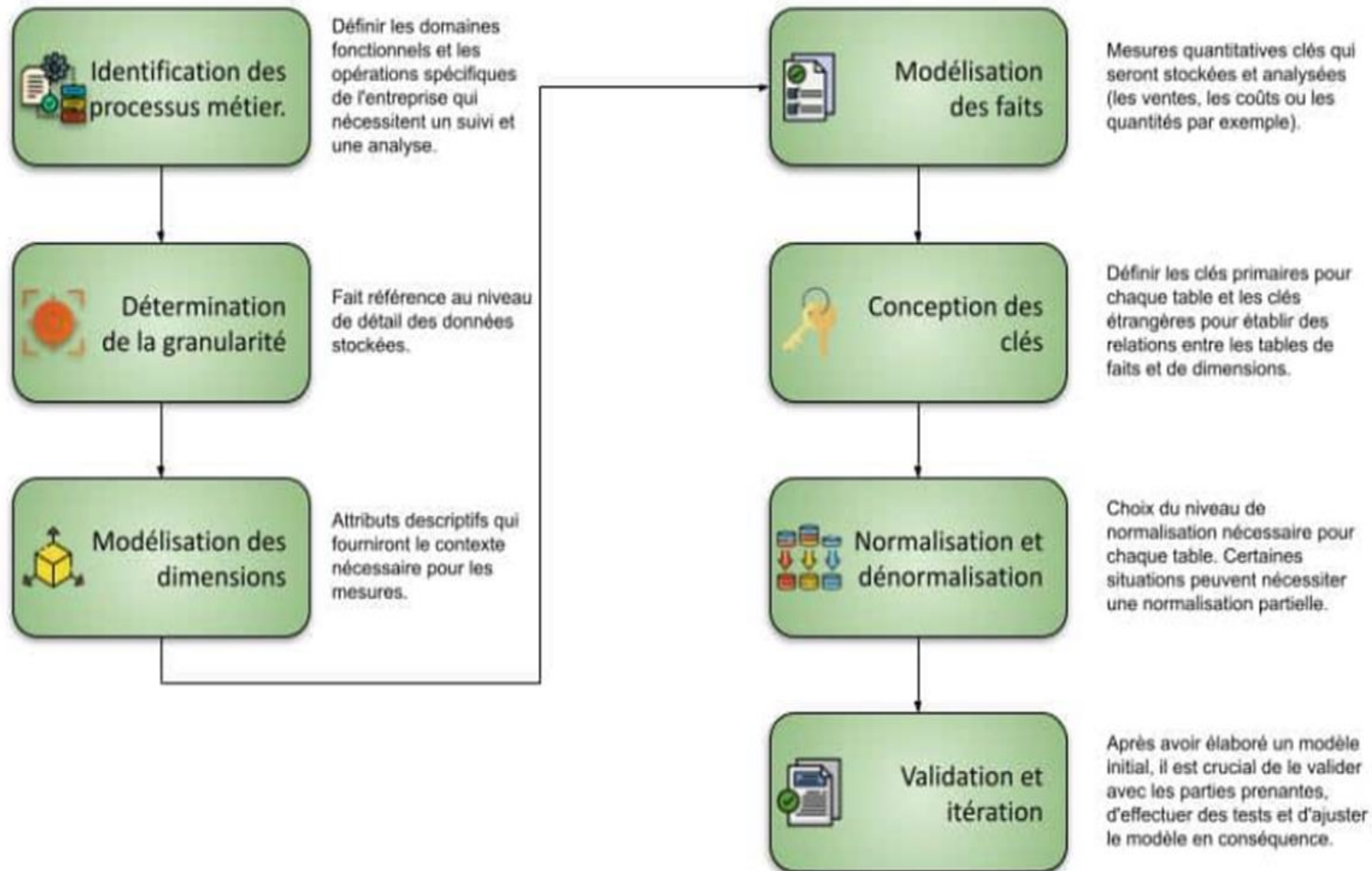
Une **méthodologie de gestion de projet** est un ensemble structuré de pratiques, principes, et processus utilisés pour planifier, exécuter, et finaliser un projet avec succès. Elle fournit un cadre pour organiser les tâches, gérer les ressources, et contrôler les risques tout en respectant les délais et le budget.

Pourquoi l'utiliser ?

- **Clarté et organisation** : Elle structure le projet en étapes définies, facilitant la gestion.
- **Gestion des risques** : Elle aide à anticiper et à résoudre les problèmes potentiels.
- **Efficacité** : Elle optimise l'utilisation des ressources et améliore la communication entre les équipes.
- **Suivi des objectifs** : Elle permet de mesurer les progrès et d'assurer que les objectifs sont atteints.

# Méthodes appliquées pour un projet BI

## Méthode KIMBALL



# Méthodes appliquées pour un projet BI

## Méthode GIMSI



# Méthodes appliquées pour un projet BI

## Méthode GIMSI

### Project management method : GIMSI

#### Identification

- Business Environment
- Company Identification

#### Design

- Objective Definition
- Dashboard Design
- Choosing Key Performance Indicators (KPIs)
- Data Collection
- Dashboard system

#### Implementation

- Software Selection
- Integration and Deployment

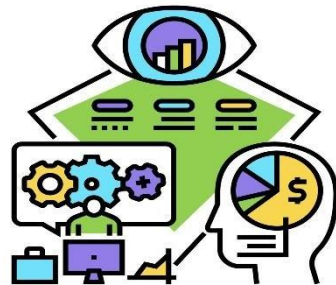
#### Improvement

- System Audit

# Partie 2:

## Entrepôt de données (Datawarehouse)

**BUSINESS  
INTELLIGENCE**



# Data Warehouse ou entrepôt de données

## Définition et objectifs (1/4)

- Les datawarehouse sont des systèmes conçus pour l'aide à la prise de décision. (Mode de travail: **OLAP** On-Line Analytical Processing)

- D'après BILL Inmon (1996) :

**« Un DW est une collection de données orientées sujet, intégrées, non volatiles, historisées, organisées pour la prise de décision. »**

- Les objectifs principaux sont
  - ◆ regrouper, organiser des informations provenant de sources diverses,
  - ◆ les intégrer et les stocker pour donner à l'utilisateur une vue **orientée métier (sujet)**,
  - ◆ retrouver et analyser l'information facilement et rapidement.

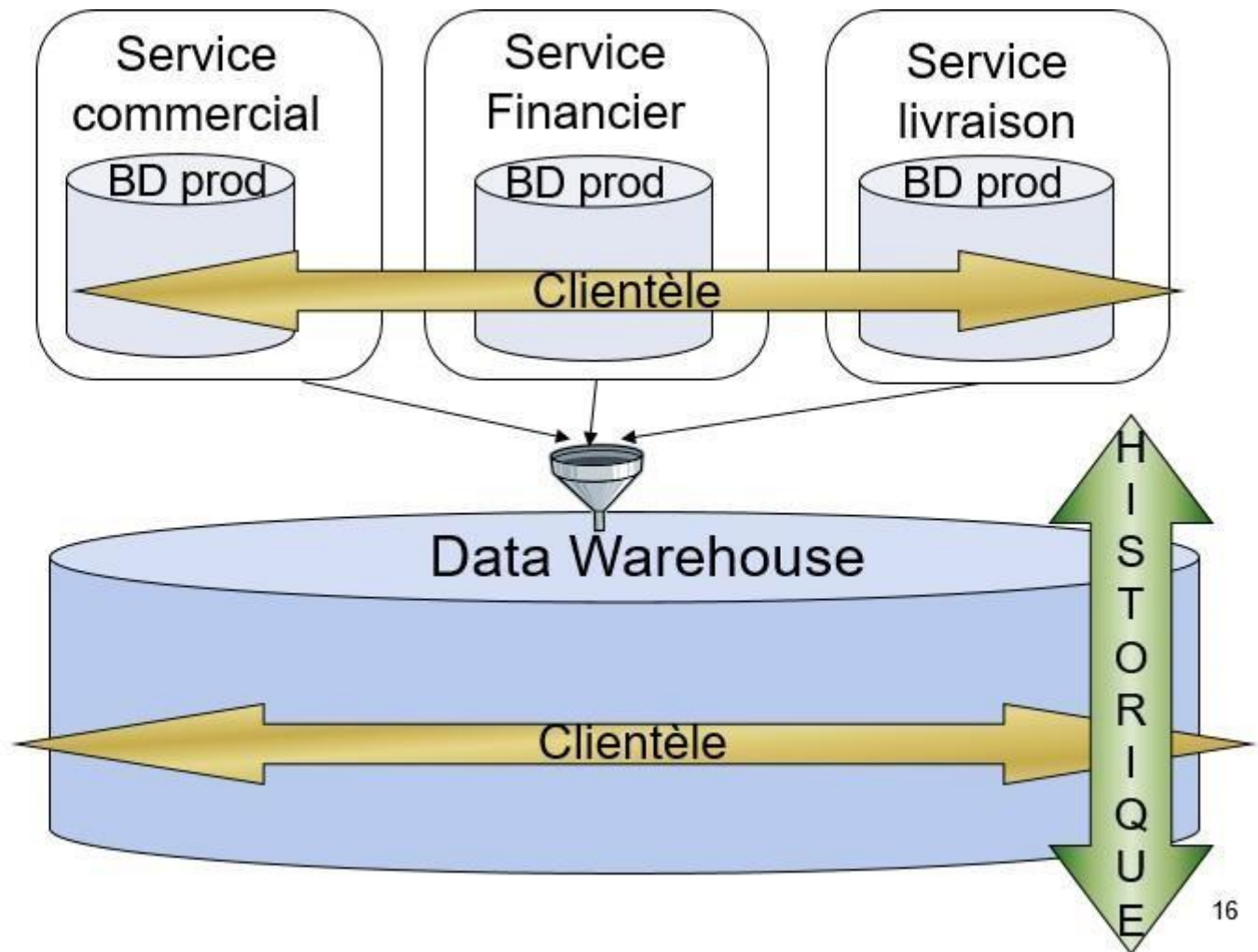


# Data Warehouse ou entrepôt de données

## Définition et objectifs (2/4)

OLTP: On-Line  
Transactional  
Processing

OLAP: On-Line  
Analytical  
Processing





# Data Warehouse ou entrepôt de données

## Définition et objectifs (3/4)

Dans un entrepôt de données les données sont:

**Orientées sujet:** thèmes par activités majeures ; Le datawarehouse est lui organisé autour des sujets majeurs de l'entreprise.

**Intégrées:** Les données proviennent de plusieurs sources différentes. Avant d'être intégrées au sein du datawarehouse elles doivent être mise en forme et unifiées afin d'en assurer la cohérence. Cela nécessite une forte normalisation.



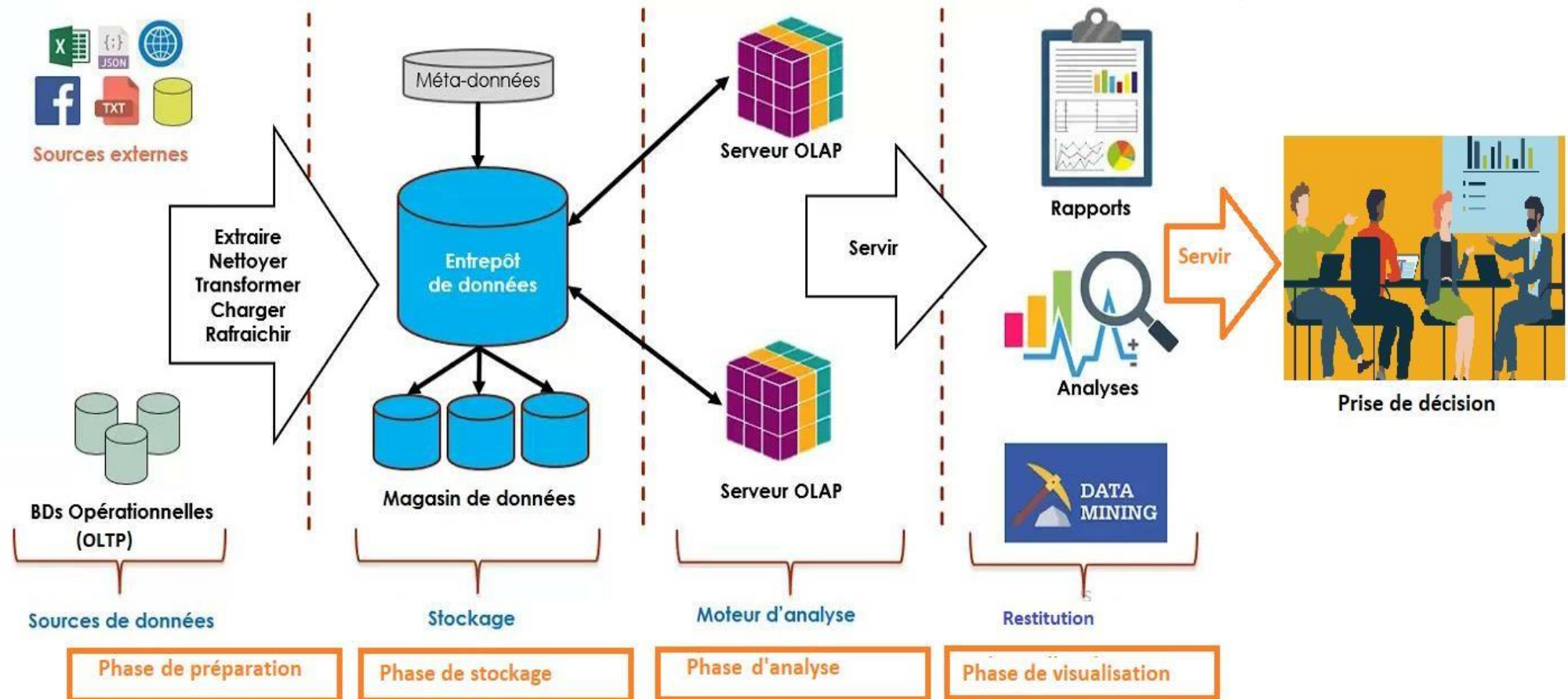
# Datawarehouse ou entrepôt de données

## Définition et objectifs (4/4)

**Non volatiles:** Un datawarehouse veut conserver la traçabilité des informations et des décisions prises. Les données ne sont ni modifiées ni supprimées. Une requête émise sur les mêmes données à plusieurs mois d'intervalles doit donner le même résultat.

**Historisées:** Contrairement au système de production les données ne sont jamais mises à jour. Chaque nouvelle données est insérées. Un référentiel de temps doit être mis en place afin de pouvoir identifier chaque donnée dans le temps.

# Architecture Globale d'un système décisionnel



# Architecture Globale d'un système décisionnel

## Phase de préparation



Cette phase fait intervenir des processus ETL qui se chargeront de récupérer les données issues de différentes sources de stockage, de les formater, nettoyer et consolider.

## Phase de stockage



Cette phase permet de stocker les données sous une forme adaptée (l'entrepôt de données ou datawarehouse).

Le datawarehouse se charge de stocker et de centraliser les données en vue de la constitution du système d'information décisionnel.

## Phase d'analyse



Cette phase permet de structurer les données sous formes de cubes à l'aide des serveurs OLAP afin de faciliter l'analyse multidimensionnelle.

## Phase de visualisation

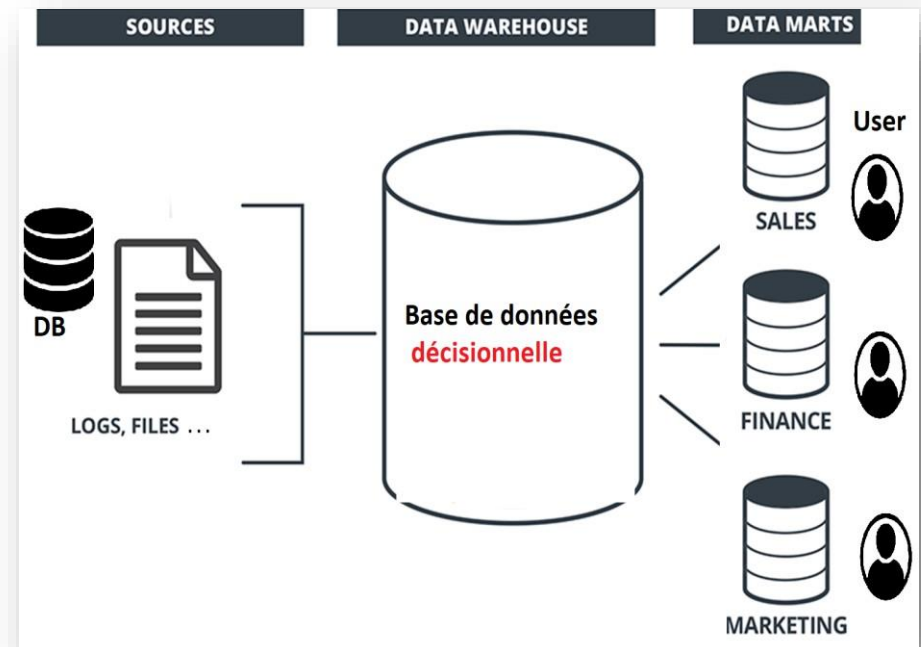


Les utilisateurs finaux interviennent dans cette dernière étape pour exploiter et analyser les données qui leur sont fournies.

# Data Warehouse et Datamart

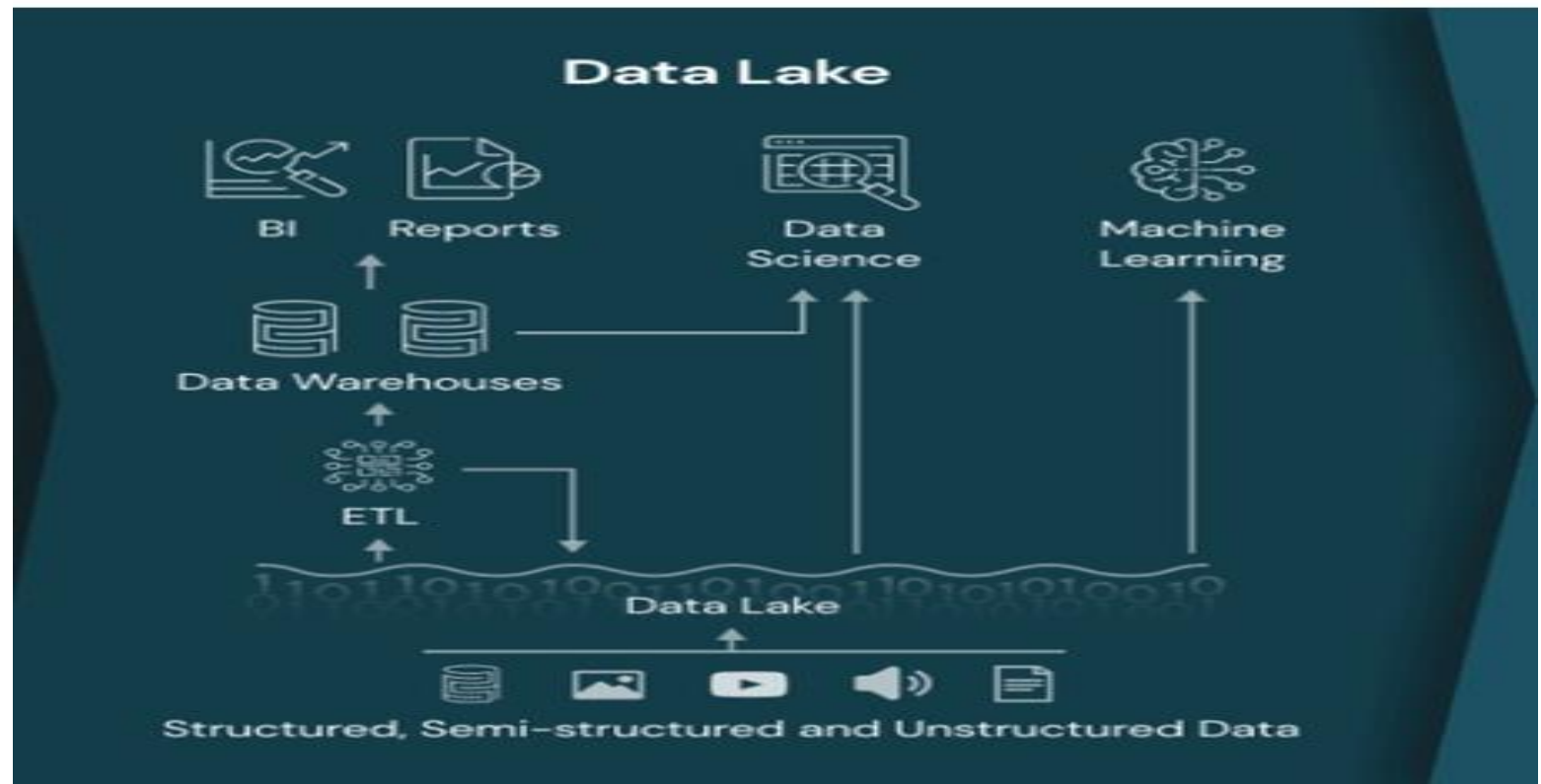
- Un Data Warehouse et un Data Mart se distinguent par le spectre qu'il recouvre :
  - Le Data Warehouse recouvre l'ensemble des données et problématiques d'analyse visées par l'entreprise.
  - Le Datamart recouvre une partie des données et problématiques liées à un métier ou un sujet d'analyse en particulier.
  - Un Datamart est un sous-

ensemble du Data Warehouse de l'entreprise, obtenu par extraction et agrégation des données de celui-ci.



# Data lake

- Un data lake est un emplacement de stockage centralisé.
- Il peut stocker des données structurées, semi-structurées ou non structurées, ce qui signifie que les données peuvent être conservées sous des formats plus souples pour une utilisation ultérieure.
- Lorsqu'il importe les données, le data lake les associe à des identificateurs et des balises de métadonnées pour une récupération plus rapide.





# Data Warehouse vs Data lake

**Le schéma d'un datawarehouse est défini et structuré avant le stockage.**

**Un data lake n'applique pas de schéma prédéfini, ce qui lui permet de stocker les données dans leur format natif.**

**Un data warehouse est facilement accessible aux utilisateurs techniques et non techniques grâce à son schéma clairement défini et documenté.**

**Accessibilité complexe les données ne sont pas organisées sous forme simplifiée avant leur stockage, un data lake a souvent besoin d'un expert ayant une compréhension approfondie des différents types de données existants.**

**Avec un data warehouse, vous devez prévoir non seulement du temps pour définir le schéma initial, mais aussi des ressources considérables pour modifier ce schéma à l'avenir chaque fois que les besoins de l'entreprise évoluent.**

**Les data lakes s'adaptent très facilement aux changements. Enfin, lorsque les besoins en capacité de stockage augmentent, il est plus facile de faire évoluer les serveurs d'un cluster de data lake.**



# Datawarehouse ou entrepôt de données

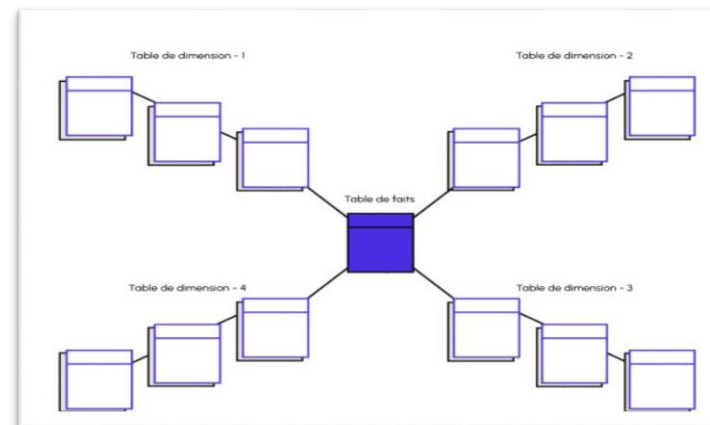
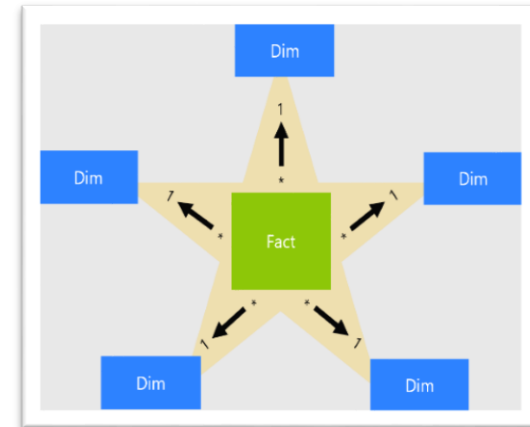
## Modélisation (1/2)

➤ Il existe trois schémas possibles pour la modélisation de l'entrepôt :

- Le schéma en étoile
- le schéma en flocon.
- Le schéma en constellation

➤ Quel que soit le modèle considéré, on distinguera :

- **la table des faits** qui contient l'information à analyser (par exemple les ventes)
- **des tables de dimensions** qui contiennent les informations sur les dimensions d'analyse (par exemple le lieu, le temps, la description du produit).



# Datawarehouse ou entrepôt de données

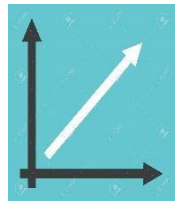
## Modélisation (2/2)

- Pour la Modélisation conceptuelle BD : entité et relation
- Pour la Modélisation de **DW** : dimension et mesure
- Les mesures sont les valeurs numériques que l'on compare



(ex : montant\_ventes, qte\_vendue)

- ◆ Ces valeurs sont le résultat d'une opération d'agrégation des données
- Les dimensions sont les points de vues depuis lesquels les mesures peuvent être observées :
  - ◆ Ex : date, localisation, produit, etc.
  - ◆ Elles sont stockées dans les tables de dimensions



# Modélisation du DW

## Dimension

- Une dimension peut être définie comme :
  - ◆ un thème, ou un axe (attributs), selon lequel les données seront analysées
  - ◆ Ex : Temps, Région, Produits...
- Une dimension contient des membres organisés en hiérarchie:
  - ◆ Chacun des membres appartient à un niveau hiérarchique (ou niveau de granularité) particulier
  - ◆ Ex : pour la dimension Temps:
    - ◆ année – semestre – mois – jour

# Modélisation du DW

## Mesure – fait – table de faits

- Une mesure est un élément de donnée sur lequel portent les analyses, en fonction des différentes dimensions
  - ◆ **Exemple :** coût des travaux, nombre d'accidents, ventes
- Un fait représente la valeur d'une mesure, mesurée ou calculée, selon un membre de chacune des dimensions
  - ◆ **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 »
- Les mesures sont stockées dans les tables de faits
  - ◆ Table de fait contient les valeurs des mesures et les clés vers les tables de dimensions

# Le Modèle en étoile

## Avantages et inconvénients

- ▮ Une table de faits comprenant une (ou plusieurs) mesures.
- Plusieurs tables de dimension dénormalisées
- ▮ Les tables de dimension n'ont pas de lien entre elles

### Avantages :

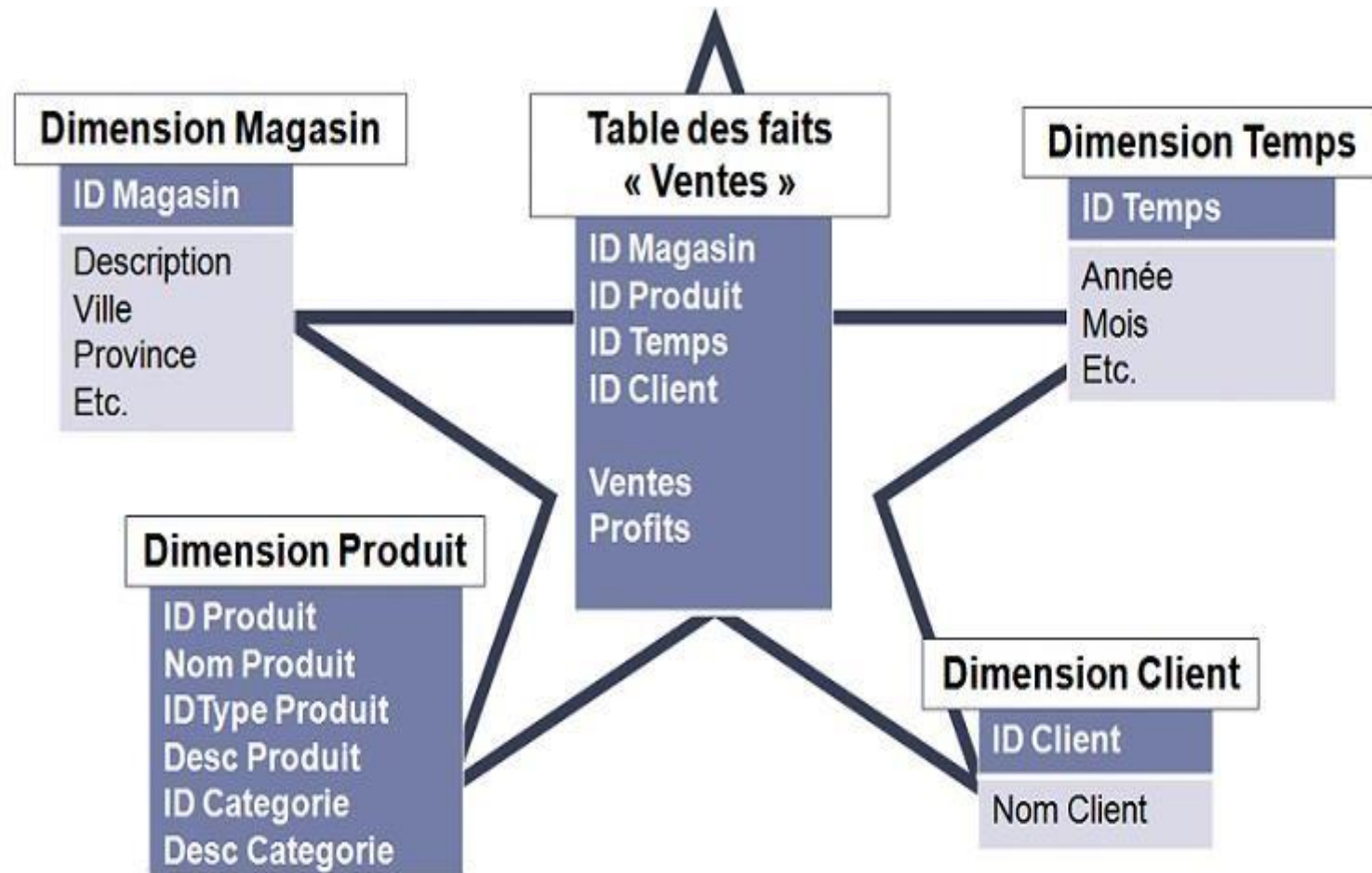
- Facilité de navigation
- ▮ Alimentation facile.
- ▮ Performances : nombre de jointures limité ; gestion des données creuses.
- ▮ Gestion des agrégats

### Inconvénients :

- Toutes les dimensions ne concernent pas les mesures
- ▮ Redondances dans les dimensions
- ▮

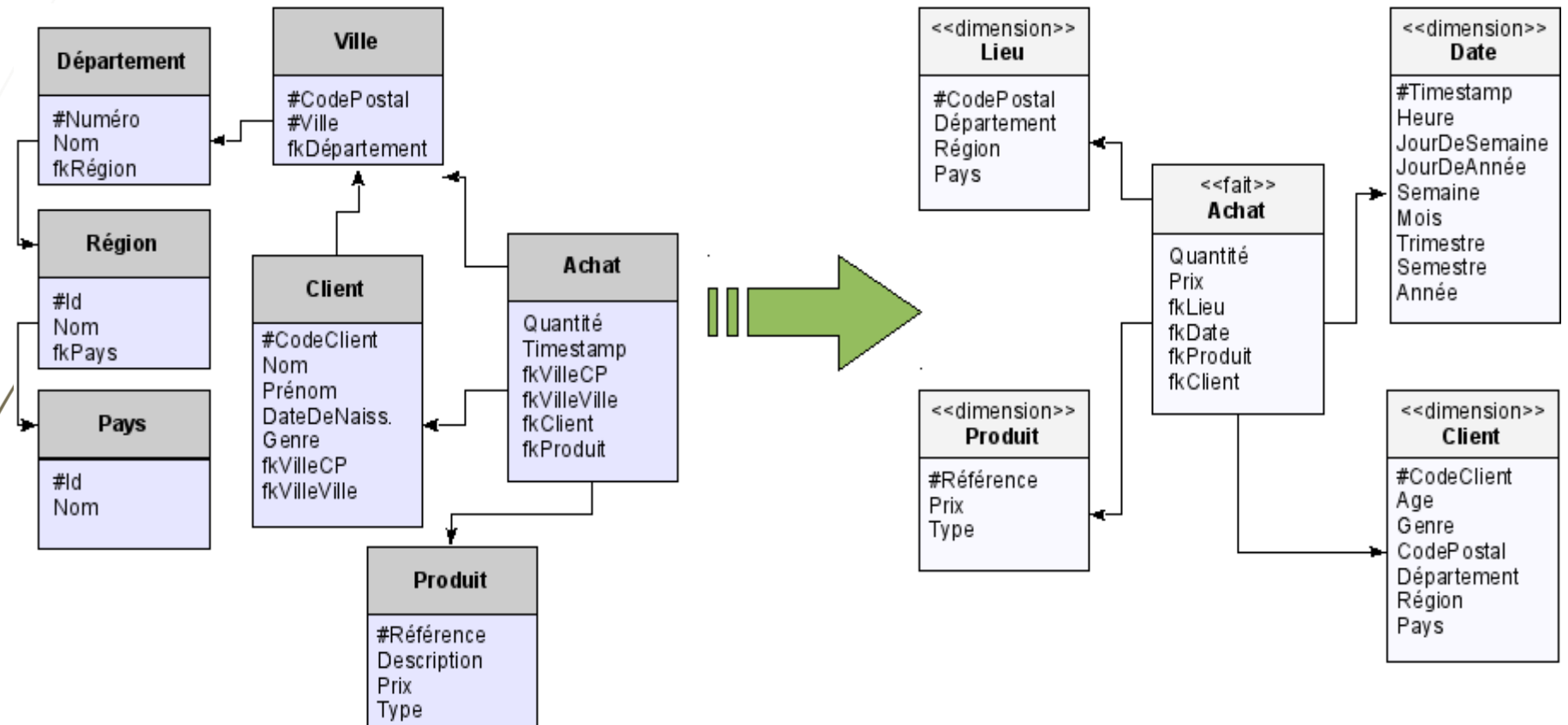
# Le Modèle en étoile

## Exemple1



# Le Modèle en étoile

## Exemple2





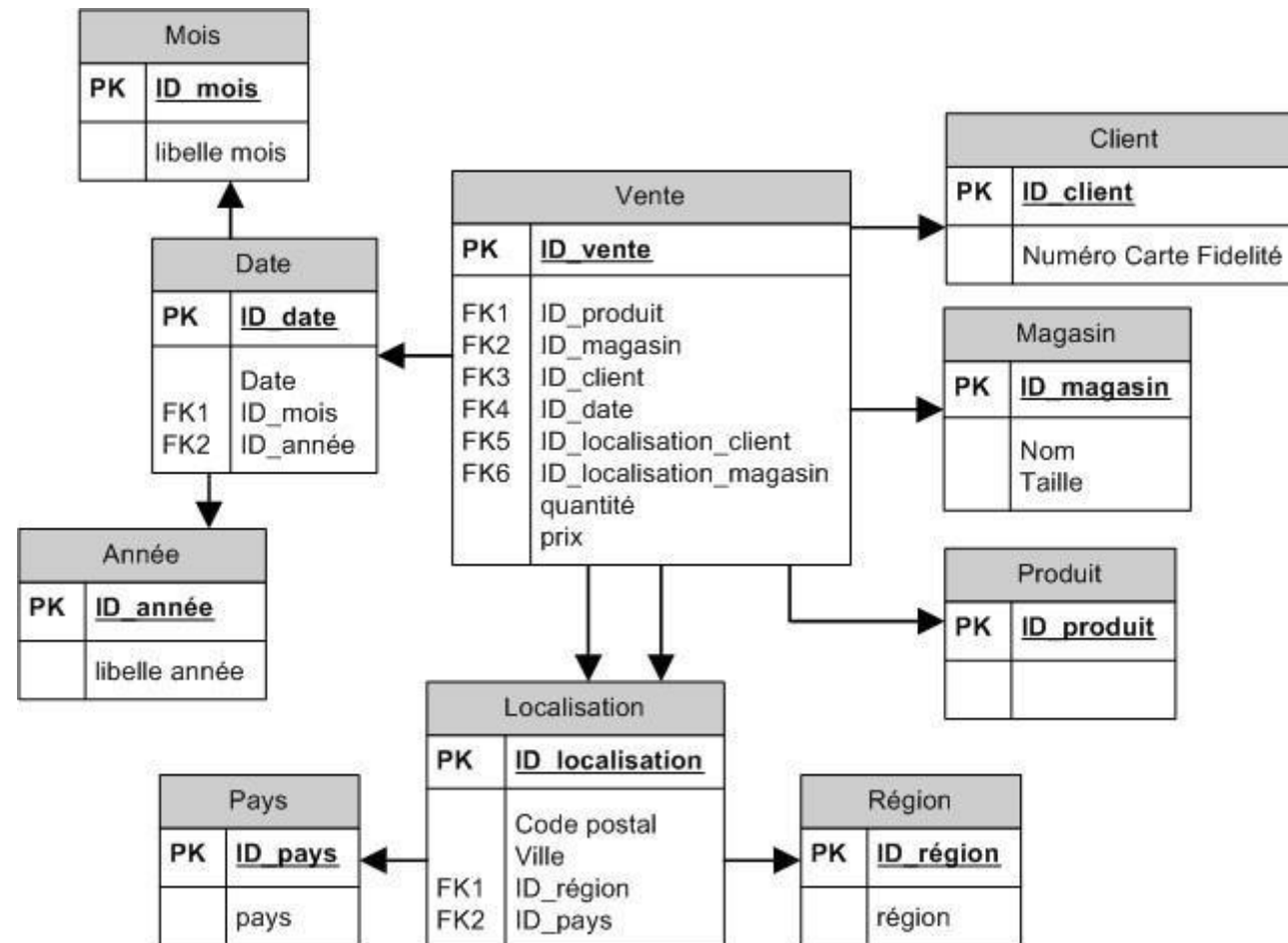
# Le Modèle en flocons

## Avantages et inconvénients

- Le schéma en flocon est dérivé du schéma en étoile où les tables de dimensions sont normalisées (la table des faits reste inchangée).
- Avec ce schéma, chacune des dimensions est décomposée selon sa (ou ses) hiérarchie(s).
  - Exemple : Commune, Département, Région, Pays, Continent
- Utilisé lorsque les tables sont trop volumineuses
  - ◆ Avantages :
    - réduction du volume,
  - ◆ Inconvénients :
    - navigation difficile,
    - nombreuses jointures.
    - Alimentation complexe.

# Le Modèle en flocons

## Exemple

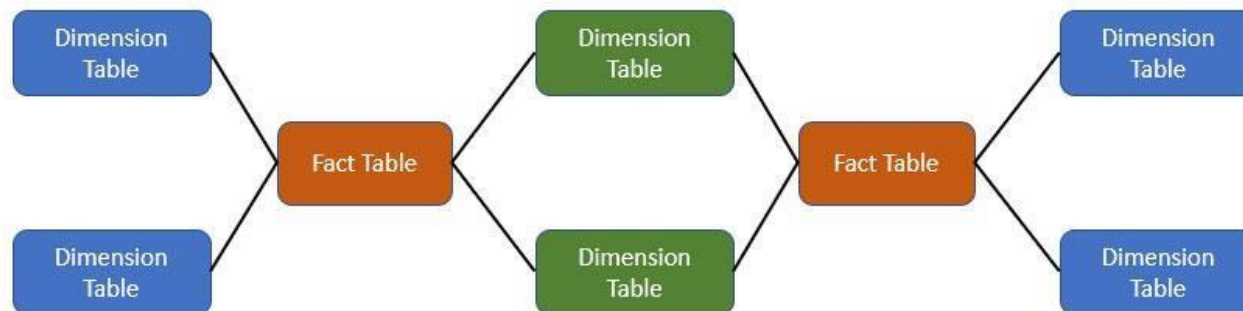


# Modèle en constellation

## Présentation

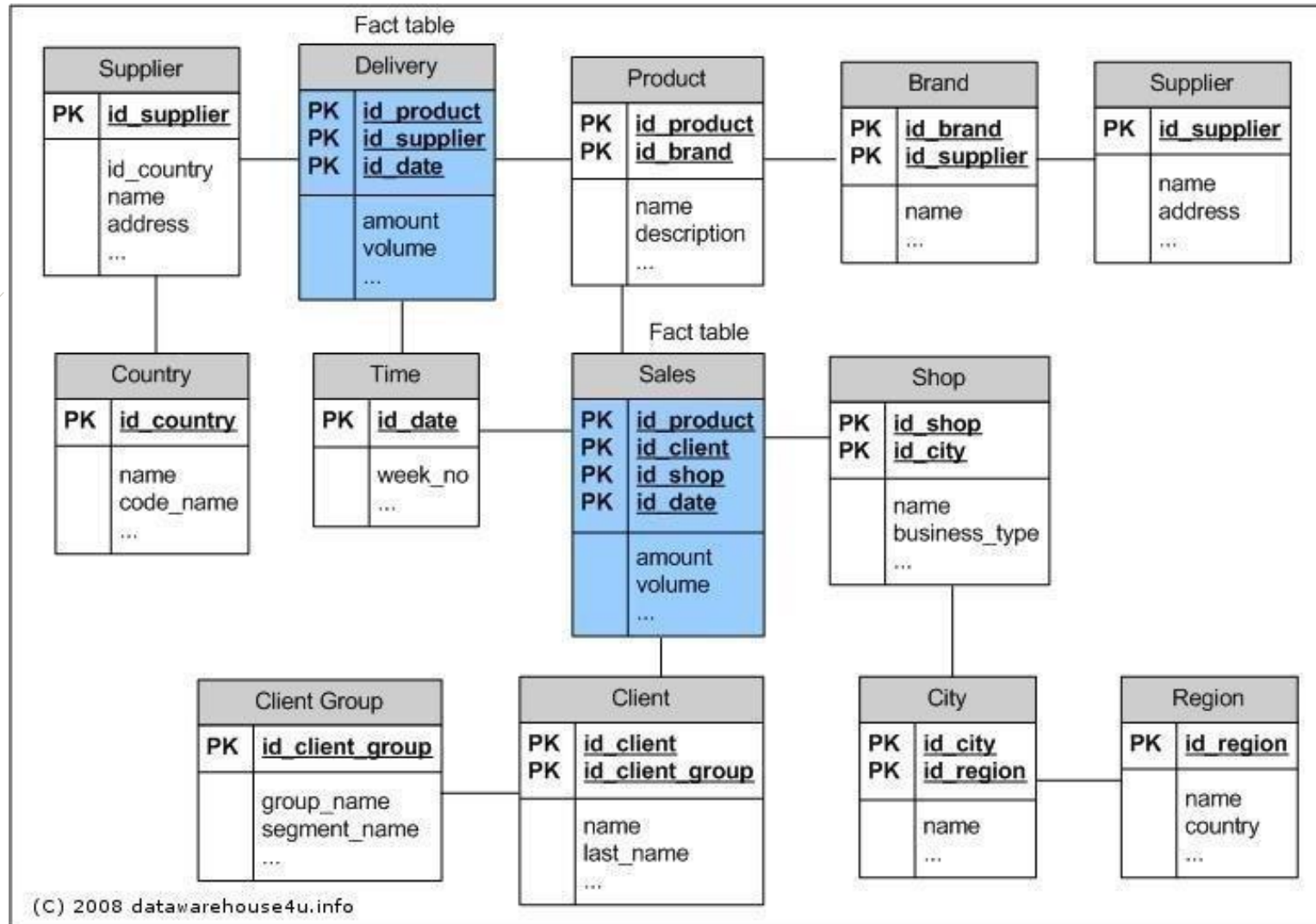
La modélisation en constellation consiste à fusionner plusieurs modèles en étoile qui utilisent des dimensions communes.

Un modèle en constellation comprend donc plusieurs tables de faits et des tables de dimensions communes ou non à ces tables de faits.



# Modèle en constellation

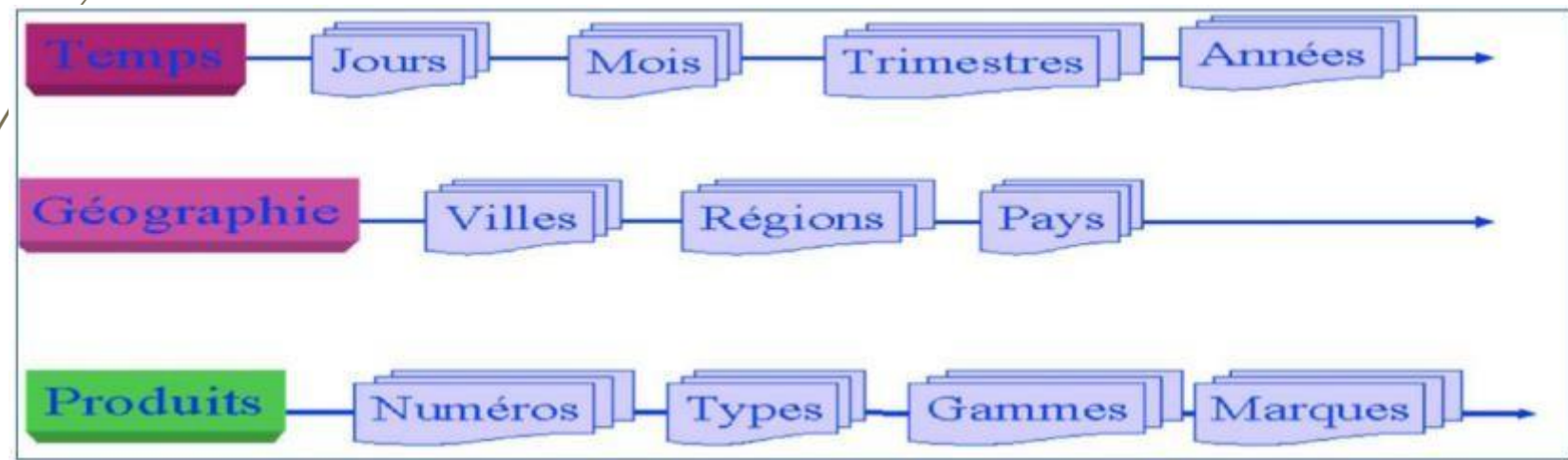
## Exemple



## Hiérarchie des paramètres d'une dimension

- Une mesure est généralement analysée selon les dimensions qui le caractérisent.
- Nécessaire de définir pour chaque dimension ses différents niveaux de détail définissant ainsi une (ou plusieurs) hiérarchie(s) de paramètres

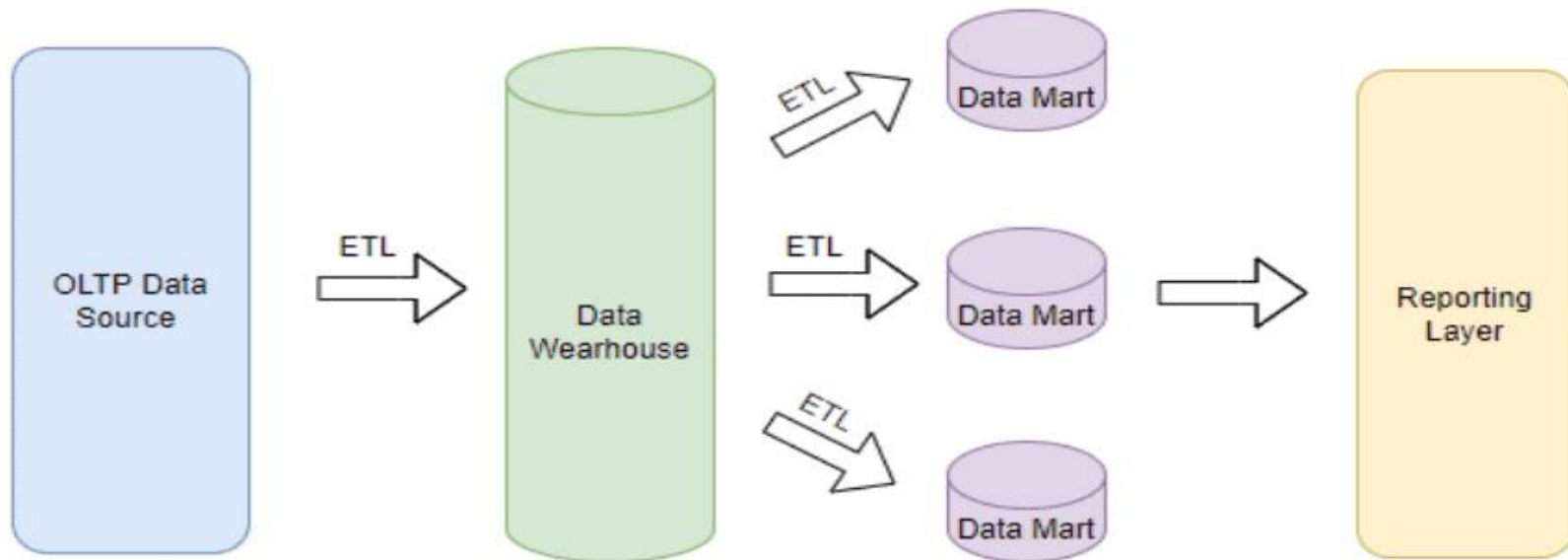
Exemple :



# Approches Modélisation DataWarehouse

- **L'approche d'Inmon (ou approche Top-Down):**

le Data Warehouse est un référentiel centralisé d'entreprise stockant l'information au niveau le plus détaillé. Des Datamarts modélisés sous forme de schémas en étoile sont ensuite créés à partir de ce Data Warehouse.



# Approches Modélisation DataWarehouse

- **L'approche de Kimball (ou approche Bottom-Up)**

Le Data Warehouse peut être vu comme l'union des datamarts cohérents entre eux grâce aux dimensions conformes (data warehouse bus). La modélisation dimensionnelle permet un requetage facile et performant pour les utilisateurs.

