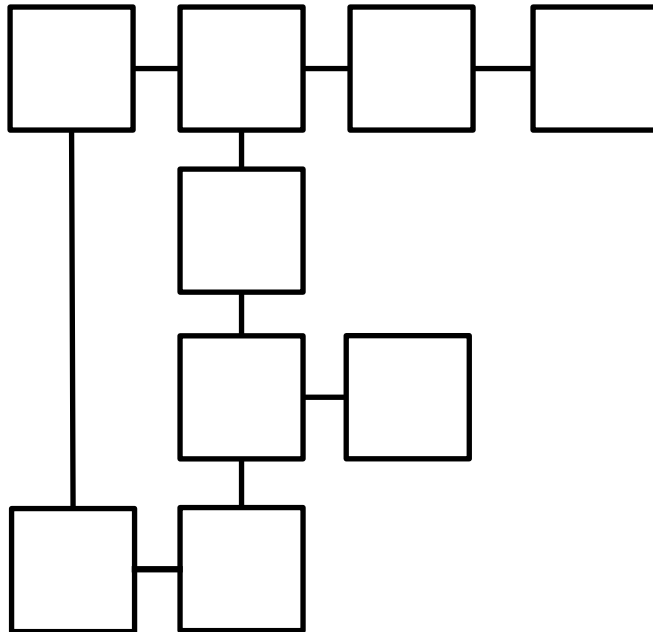


# CONCEPTION D'UN DATAWAREHOUSE

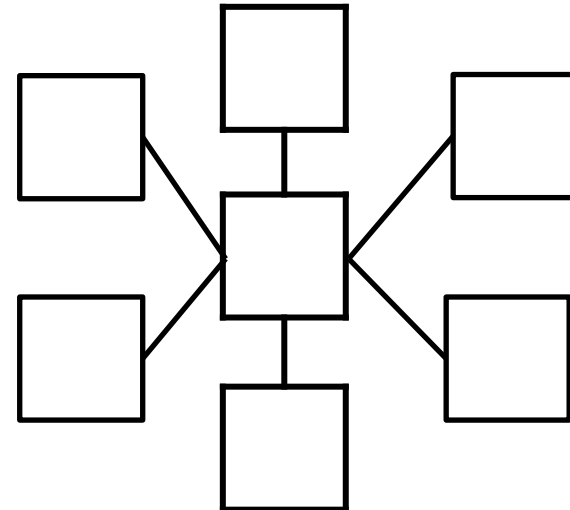


# DIFFÉRENCE DE CONCEPTION

## Schéma Relationnel – Entité/Relation



## Schéma en Etoile



# MODÉLISATION

## **Schéma en étoile**

Une table de faits au milieu du schéma est connectée à un ensemble de tables de dimensions

## **Schéma flocon de neige (snowflake)**

Un raffinement du schéma en étoile où des tables de dimensions sont décomposées

## **Constellation de faits**

Plusieurs tables de faits partagent des tables de dimension (constellation d'étoiles)

# MODÉLISATION MULTIDIMENSIONNEL

## **Table de fait**

La table de faits constitue une table de référence centrale permettant d'accéder aux événements ou activités archivés et inhérents à un processus déterminé.

## **Dimension**

Les tables de dimension sont des compagnons intégrés à une table de fait. Les tables de dimension contiennent le contexte textuel associé à un événement de mesure des processus métiers. Ils décrivent le «qui, quoi, où, quand, comment, et pourquoi" associé à l'événement.

## **Attribut**

Les attributs qualifient les dimensions. Généralement, les attributs sont textuels et discrets (par opposition aux faits).



# TABLE DE FAITS

Les tables de faits sont des collections de **mesures** associées à un processus métier spécifique. Les mesures sont stockés dans les colonnes.

Les tables de faits contiennent généralement un grand nombre de lignes, parfois dans les centaines de millions d'enregistrements lorsqu'elles contiennent une ou plusieurs années d'historique pour un grand organisme.

## Table de faits agrégée

L'agrégation est le processus de calcul des données de synthèse à partir de données plus détaillées. Il permet de réduire la taille des tables de faits par l'agrégation des données dans les comptes rendus analytiques.

# TABLE DE FAITS - GRANULARITÉ

Le grain détermine le niveau de détail de la mesure de table de faits:

## **Par exemple**

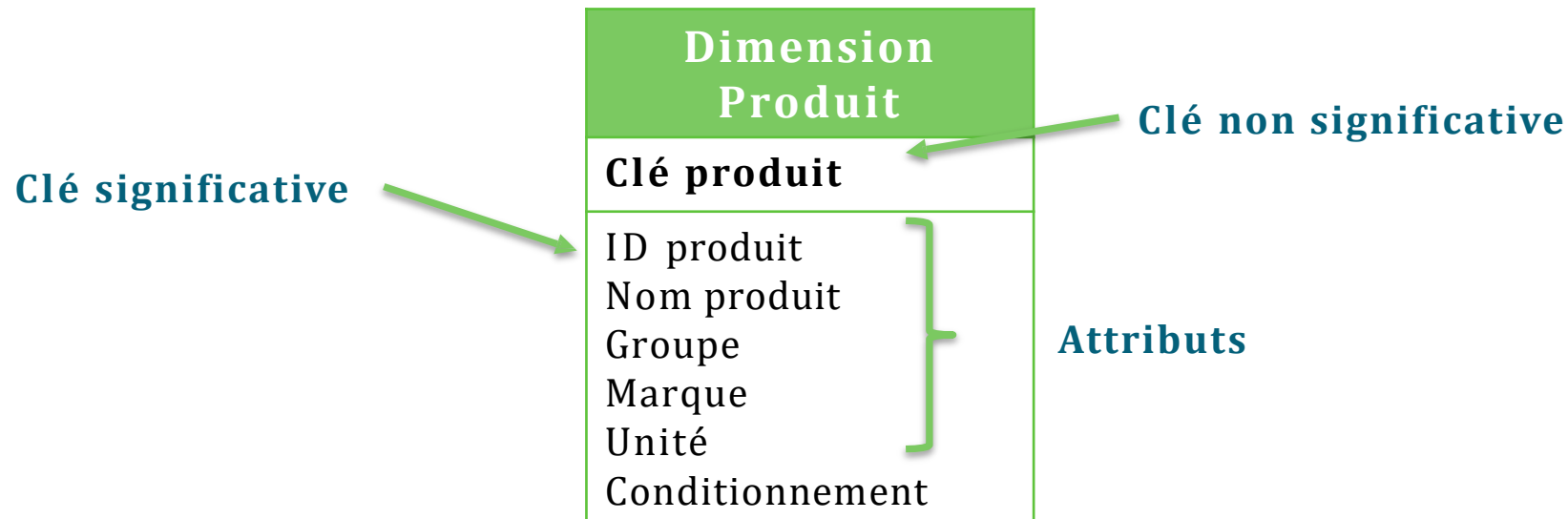
- Transactions individuelles
- Transactions Instantanés (points dans le temps)
- Éléments de ligne sur un document

Le plus petit niveau de granularité est meilleur pour l'analyse (mais pas pour le stockage)

# DIMENSION

Une table de dimension se compose de:

- Une clé non significative établissant un lien avec les lignes de la table de faits
- Une clé significative reprise d'une source de données opérationnelle ou externe
- Un nombre d'attributs permettant de caractériser la dimension



## CLÉ NON SIGNIFICATIVE

Aussi appelé clé de substitution (Surrogate Key)

Les clés de table de dimension sont non intelligentes et non liées à l'entreprise car:

- Les clés métier (clé significative) peuvent changer avec le temps
- Les clés de substitution gardent une trace des valeurs des attributs non clés pour une clé significative donnée
- Les clés de substitution sont plus simples et plus courtes
- Les clés de substitution peuvent avoir la même longueur et le même format pour toutes les clés



# SLOWLY CHANGING DIMENSION (SCD)

## Dimension à variation lente

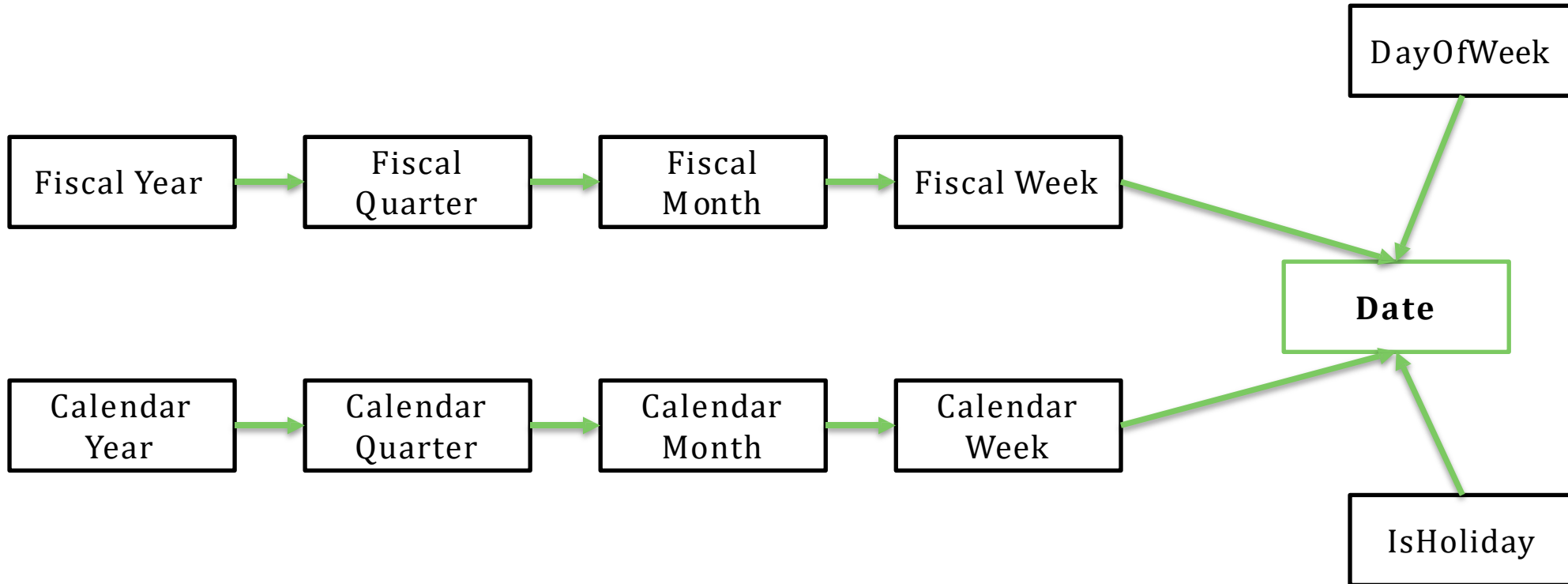
- **Type 1:** remplacer les anciennes données par les nouvelles (perte des données historiques)
- **Type 2:** créer une nouvelle ligne de table de dimension chaque fois que l'objet de dimension change, avec toutes les caractéristiques de la dimension au moment du changement.
  - **Approche la plus courante**
- **Type 3:** pour chaque attribut changeant, créer un champ de valeur courant et plusieurs champs d'ancienne valeur (plusieurs valeurs)

# DIMENSION TEMPORELLE

Dimension Date
Clé date
date complète
jour
mois
année
semestre
Trimestre
... ..

- Type particulier de dimension
- Dimension cruciale pour l'analyse
- Dupliquer les attributs même s'ils peuvent être déduits

# DIMENSION TEMPORELLE



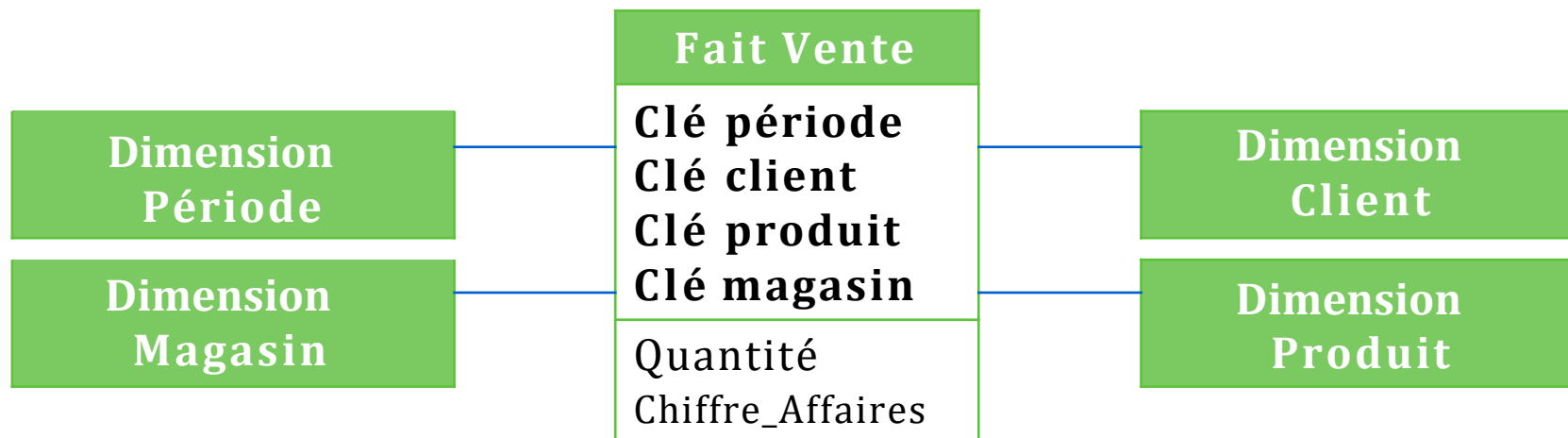
# FAIT VS DIMENSION

Les observations numériques continuellement estimées

→ **Faits**

Les observations numériques discrètes tirées d'une petite liste

→ **Dimensions**



# SCHÉMA EN ÉTOILE

- Une (ou plusieurs) table(s) de faits.
- Plusieurs tables de dimension **dénormalisées**
- Les tables de dimension ne sont pas reliées entre elles



Facilité de navigation  
Nombre de jointures limité



Redondance dans les dimensions  
Toutes les dimensions ne concernent pas les mesures  
Alimentation complexe

# RAPPEL - NORMALISATION

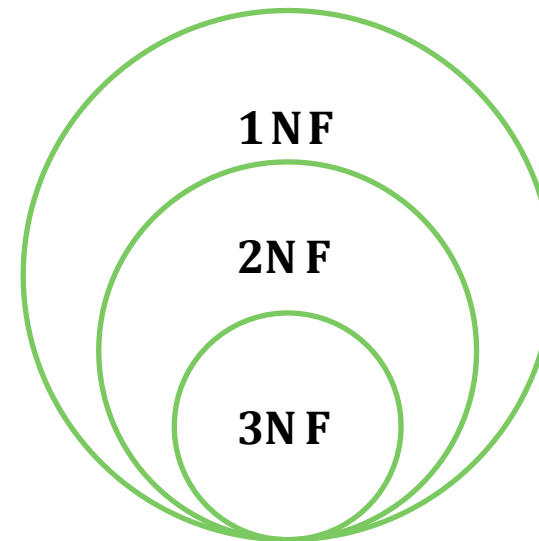
## Définition

- La normalisation est le processus qui permet d'éviter les données redondantes dans les bases de données.
- Cela implique de restructurer les tables pour atteindre successivement des formes plus élevées de normalisation.
- Une base de données correctement normalisée respecte les caractéristiques suivantes:
  - Les valeurs scalaires (atomiques) dans chaque champs
  - Absence de redondance.
  - Utilisation minimale des valeurs nulles.
  - Perte minimale d'informations.

## RAPPEL - NORMALISATION

Il existe une séquence aux formes normales:

- 1NF est considéré comme le plus faible,
- 2NF est plus fort que 1NF
- 3NF est plus fort que 2NF



**Chaque niveau supérieur est un sous-ensemble du niveau inférieur**

## RAPPEL - FORMES NORMALES

### Première forme normale

Une relation est en première forme normale si et seulement si tout attribut contient une valeur atomique.

*Livre(ID\_livre, titre, auteurs)*

→ *Livre(ID\_livre, titre, auteur1, auteur2, auteur3)*

ou mieux encore

*Livre(ID\_livre, titre, auteurs)*

→ *Livre(ID\_livre, ID\_auteur, titre)*

→ *Auteur(ID\_auteur, nomAuteur)*



## RAPPEL - FORMES NORMALES

### Deuxième forme normale

Les attributs non clé dépendent de toute la clé et non d'une partie de la clé

*Commande(ID\_fournisseur, refArticle, raisonSocialeFournisseur, adresseFournisseur, quantité, prix)*

→ *Fournisseur(ID\_fournisseur, raisonSociale, adresse)*

*Commande(ID\_commande, ID\_fournisseur, refArticle, quantité, prix)*

**Pour qu'une table soit en 2NF, il faut que:**

- La table soit en première forme normale
- Tous les attributs non clés de la table doivent être fonctionnellement dépendants de la clé primaire entière

## RAPPEL - FORMES NORMALES

### Troisième forme normale

Chaque attribut de la relation ne dépend que de la clé et pas d'un autre attribut de la relation

*Employe(ID\_Employe, nomEmploye, posteEmploye, ID\_service, nomService)*

→ *Employe(ID\_Employe, nomEmploye, posteEmploye, ID\_service)*

→ *Service (ID\_service, nomService)*

**Pour qu'une table soit en 3NF, il faut que:**

- **La table soit en deuxième forme normale**
- **Aucun attribut ne dépend de manière transitoire de la clé primaire**

# SCHÉMA EN ÉTOILE

Dim
Clé 1 (PK)
Attribut Attribut ... ..

Les tables de faits contiennent  
des données factuelles ou  
quantitatives

Clé 1 (PK)(FK)
Clé 2 (PK)(FK)
Clé 3 (PK)(FK)
Clé 4 (PK)(FK)
Data column Data column ... ..

Dim
Clé 3 (PK)
Attribut Attribut ... ..

Les tables de dimensions  
sont **dénormalisées** pour  
optimiser la performance

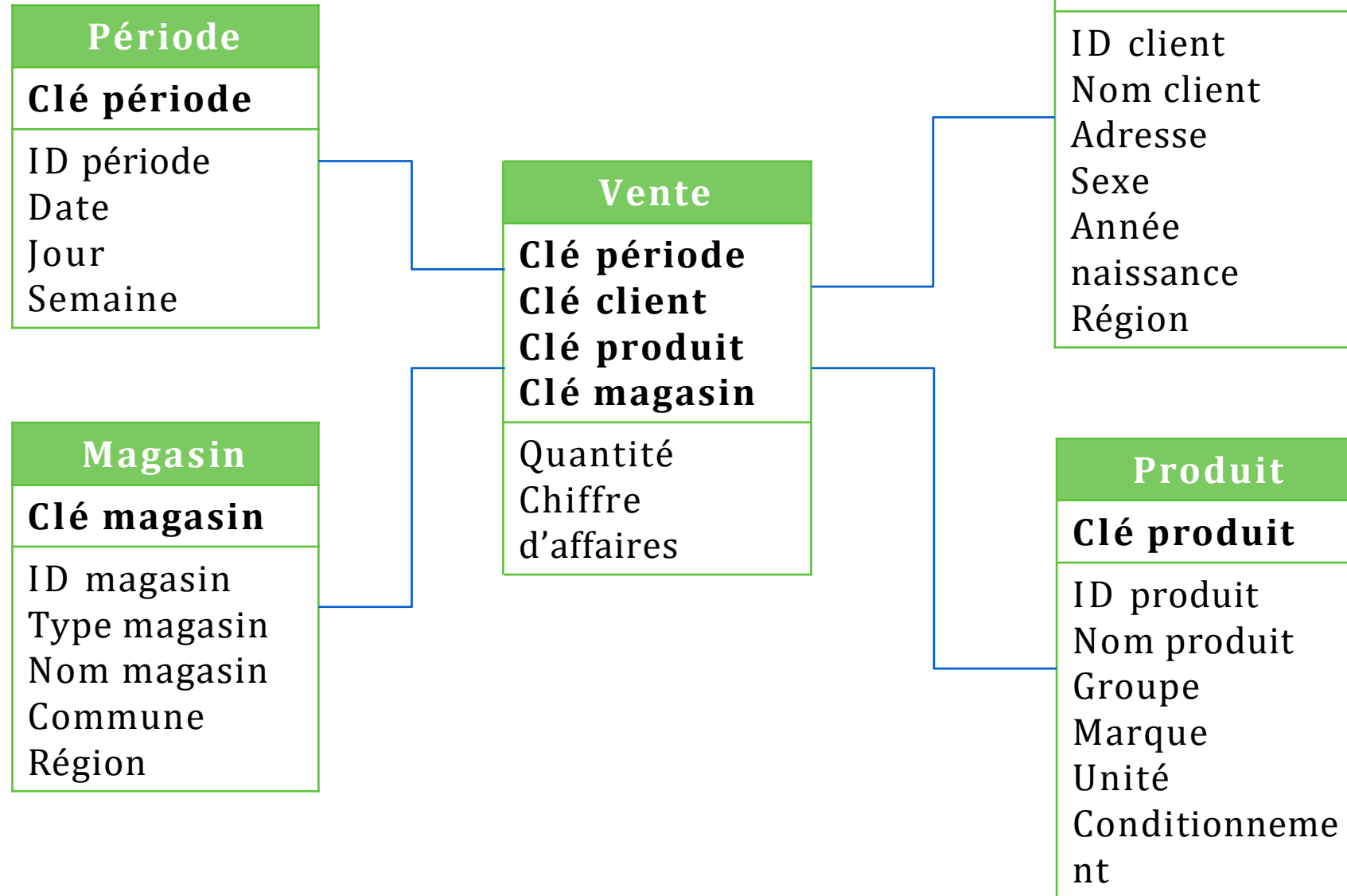
Dim
Clé 2 (PK)
Attribut Attribut ... ..

Les tables de dimensions  
contiennent des descriptions sur  
les sujets d'analyse

Clé 4 (PK)
Attribut Attribut ... ..

Relation 1: N entre  
les tables de  
dimensions et les tables de faits

# SCHÉMA EN ÉTOILE

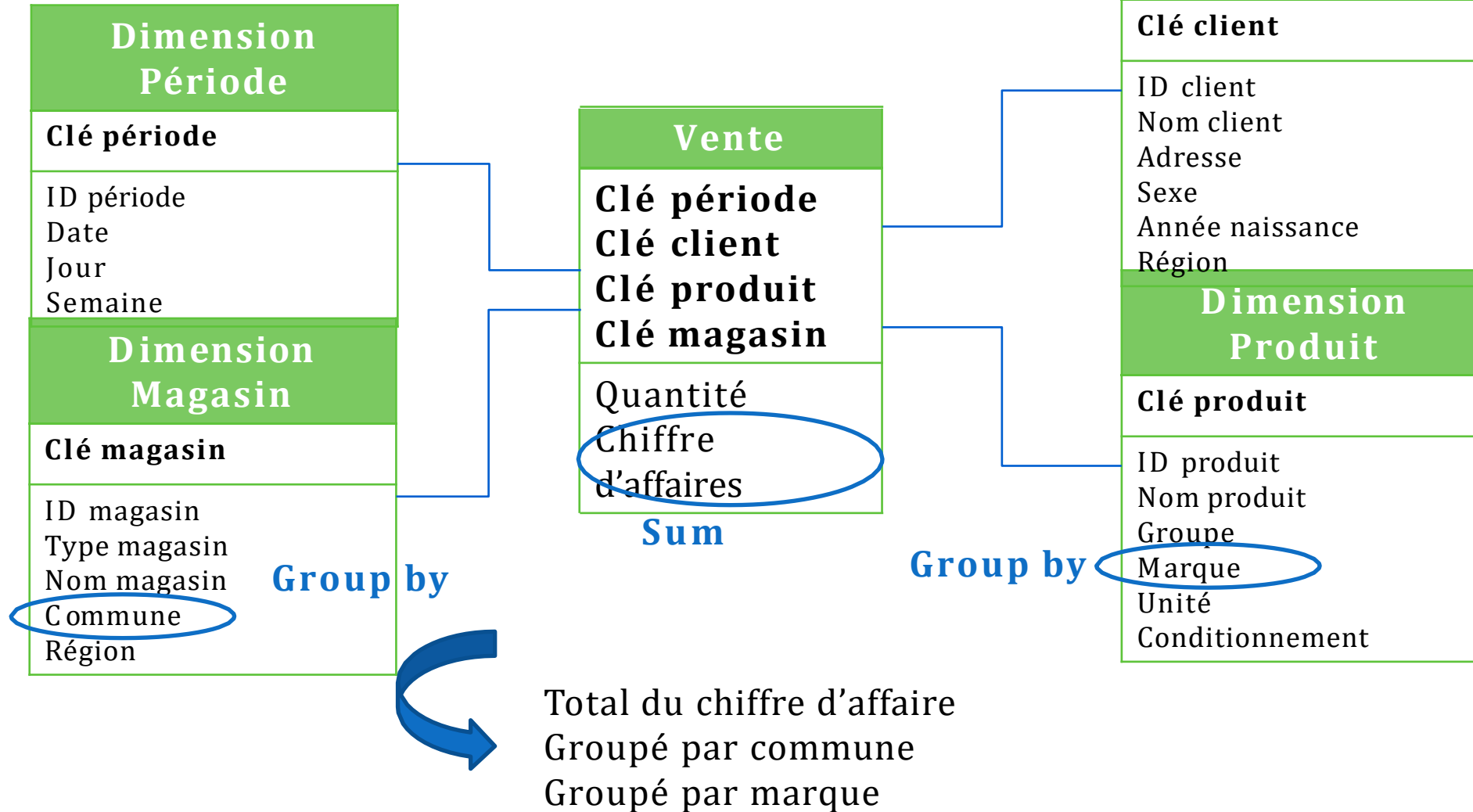


## SCHÉMA EN ÉTOILE

Un schéma en étoile contient une seule table centrale, appelée une **table de faits**, entouré de plusieurs tables appelées **dimensions**.

Une schéma en étoile couvre un secteur d'activité. Dans ce cas, le schéma couvre les ventes d'une entreprise. Un datawarehouse couvre plusieurs domaines d'activité et se compose de plusieurs schémas étoiles et/ou flocon de neige.

# SCHÉMA EN ÉTOILE



# SCHÉMA SNOWFLAKE

- Le schéma snowflake (**en flocon**) est dérivé du schéma en étoile où les tables de dimensions sont **normalisées**
- La table des faits reste inchangée
- Les dimensions sont **décomposées** selon sa (ou ses) hiérarchie(s)

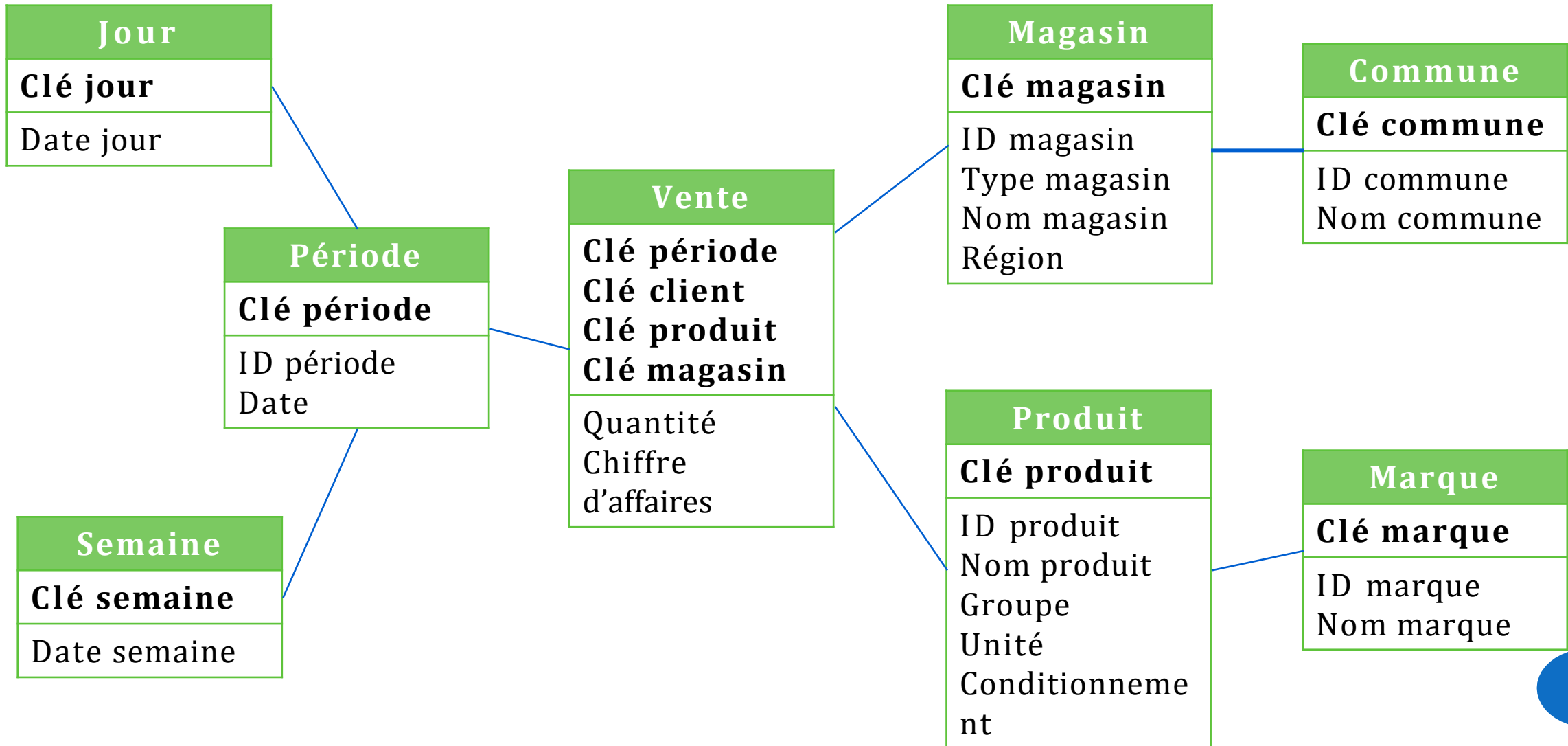


l'amélioration des performances de requête en raison de la réduction du stockage sur disque et de l'assemblage des tables



des efforts de maintenance supplémentaires nécessaires en raison de l'augmentation du nombre de tables de recherche.

# SCHÉMA SNOWFLAKE





# SCHÉMA SNOWFLAKE

**Modèle en flocons de neige**

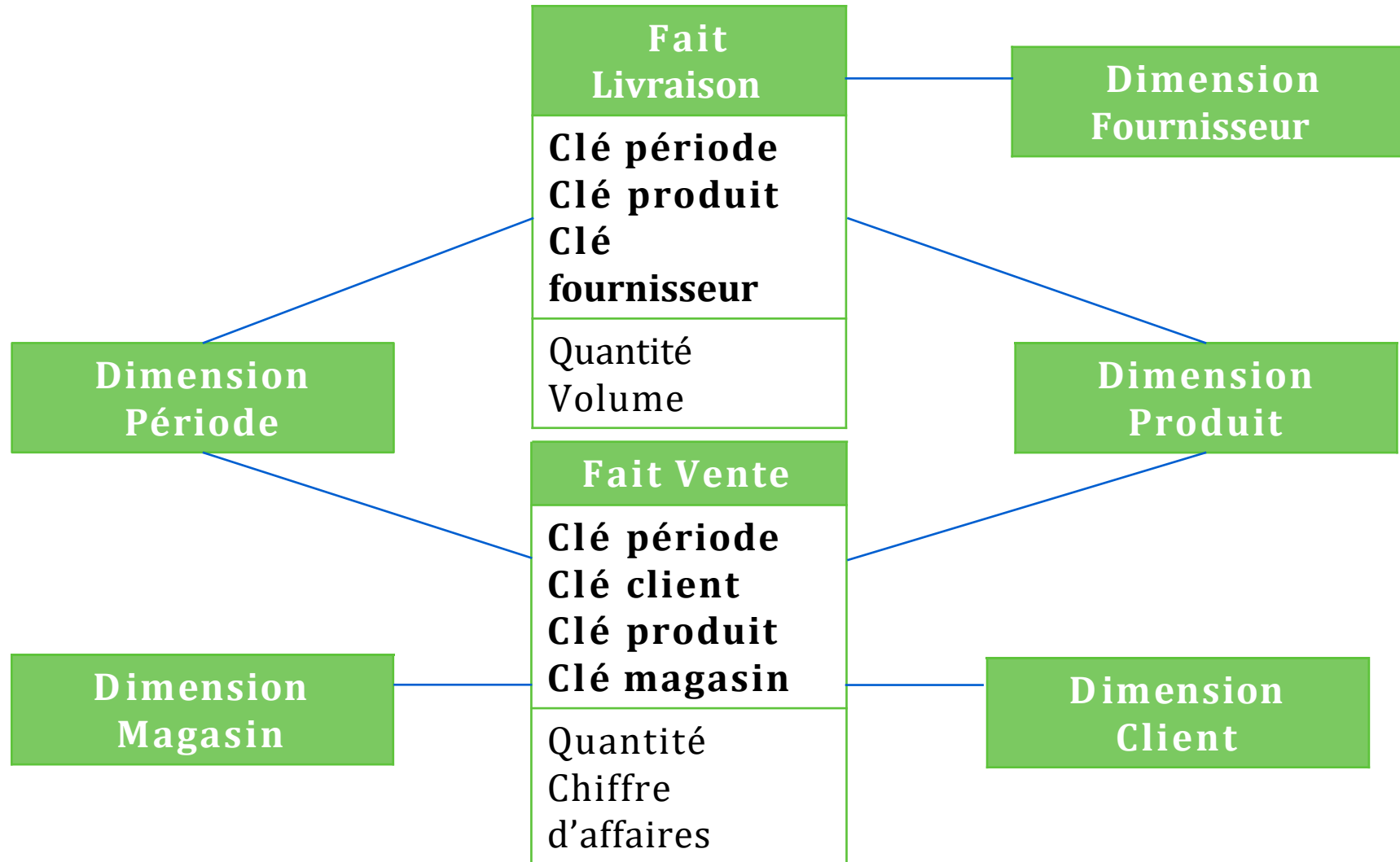
**=**

**Modèle en étoile + normalisation des  
dimensions**

## SCHÉMA EN CONSTELLATION

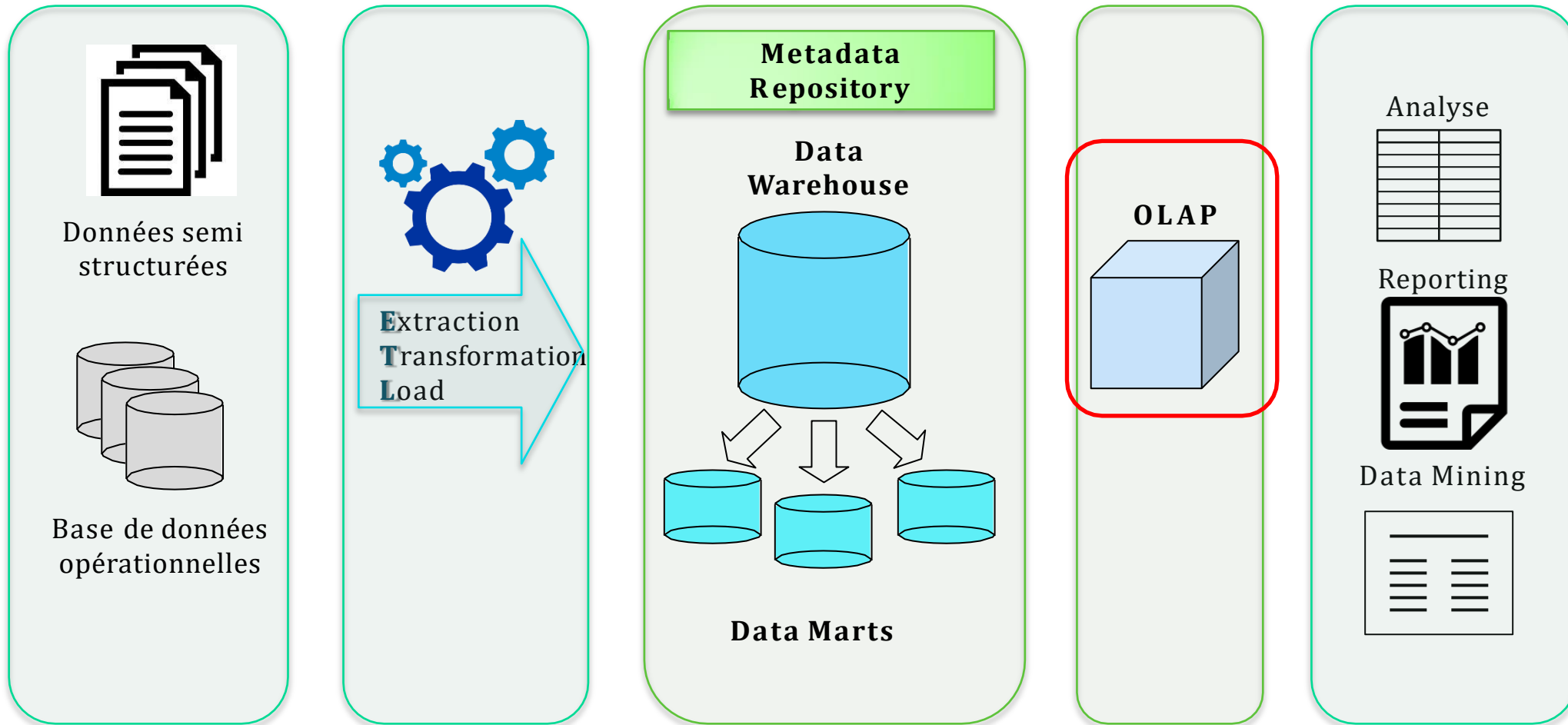
- fusionne plusieurs **modèles en étoile** qui utilisent des dimensions communes
- contient plusieurs **table de faits** et des dimensions communes ou pas

# SCHÉMA EN CONSTELLATION



# **ANALYSE MULTIDIMENSIONNELLE DES DONNÉES OLAP**

# ARCHITECTURE



# OLAP

## OLAP



OnLine Analytical Processing

OLAP effectue une analyse **multidimensionnelle** des données métier et fournit la possibilité pour les calculs complexes, l'analyse des tendances et la modélisation des données.

# OLAP

Les systèmes OLAP doivent :

- Supporter les exigences complexes des décideurs en termes d'analyse,
- Analyser les données à partir de différentes perspectives (dimensions métiers)
- Supporter les analyses complexes impliquant des ensembles de données volumineux.

# OLAP vs OLTP

## OLTP - On-Line Transaction Processing

- Technologie utilisée pour effectuer les mises à jour sur les systèmes opérationnels ou transactionnels

## OLAP - On-Line Analytical Processing

- Technologie utilisée pour effectuer des analyses complexes des données dans un entrepôt de données



# IMPLÉMENTATION OLAP

## **MOLAP (Multidimensional OLAP)**

MOLAP est la façon la plus courante de l'analyse OLAP. Dans MOLAP, les données sont stockées dans un cube multidimensionnel. MOLAP est souvent utilisé dans le cadre d'une application datawarehouse.

## **ROLAP (Relational OLAP)**

Les données multidimensionnelles sont stockées dans une base de données relationnelle standard.

Elles sont organisées en schémas en forme d'étoiles ou de flocon

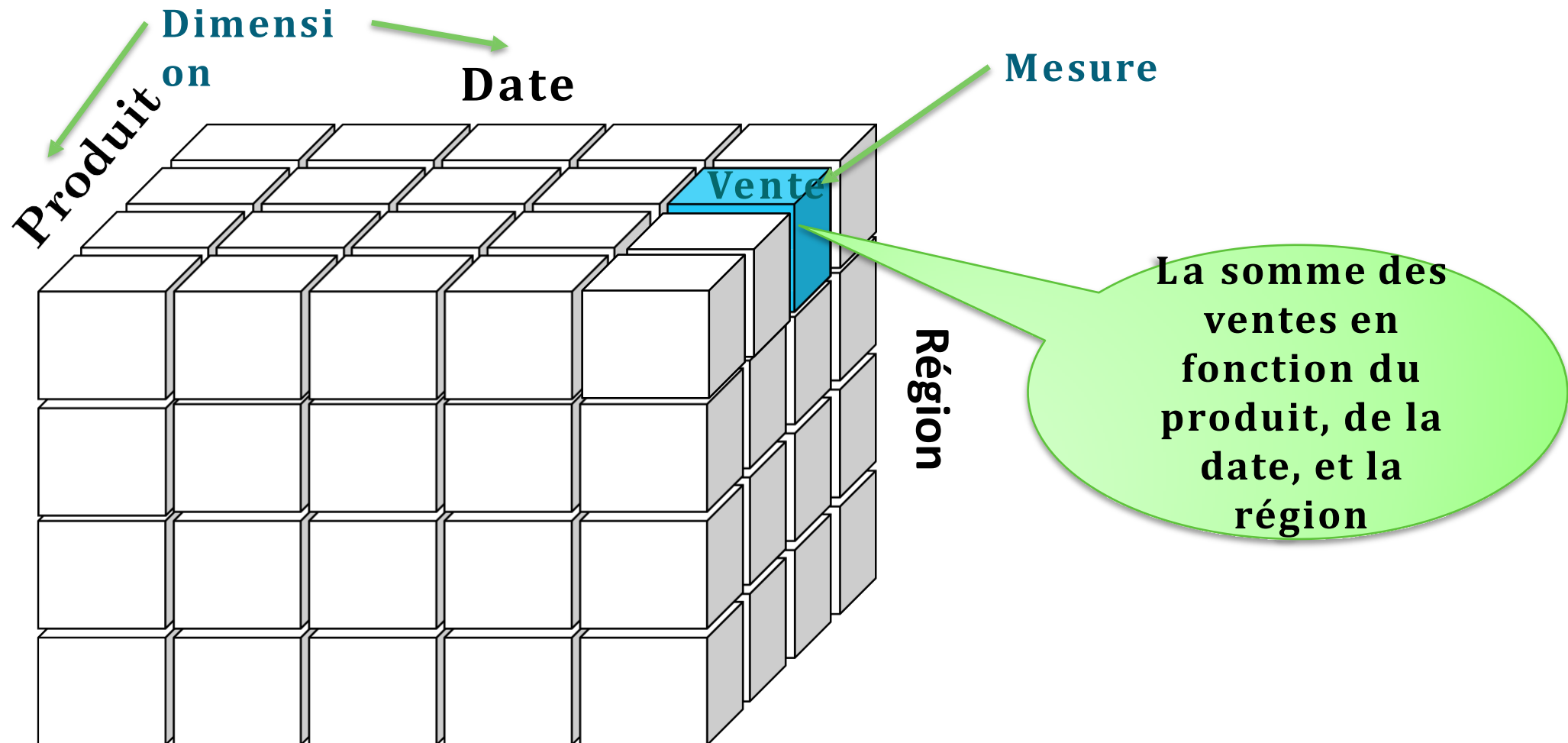
# IMPLÉMENTATION OLAP

## **HOLAP (Hybrid OLAP)**

HOLAP combine les meilleures caractéristiques de MOLAP et ROLAP dans une architecture unique.

HOLAP tente de combler le fossé technologique des deux produits en permettant l'accès ou l'utilisation à la fois des bases de données multidimensionnelles et des bases de données de datawarehouse relationnelles.

# DONNÉE MULTIDIMENSIONNELLE



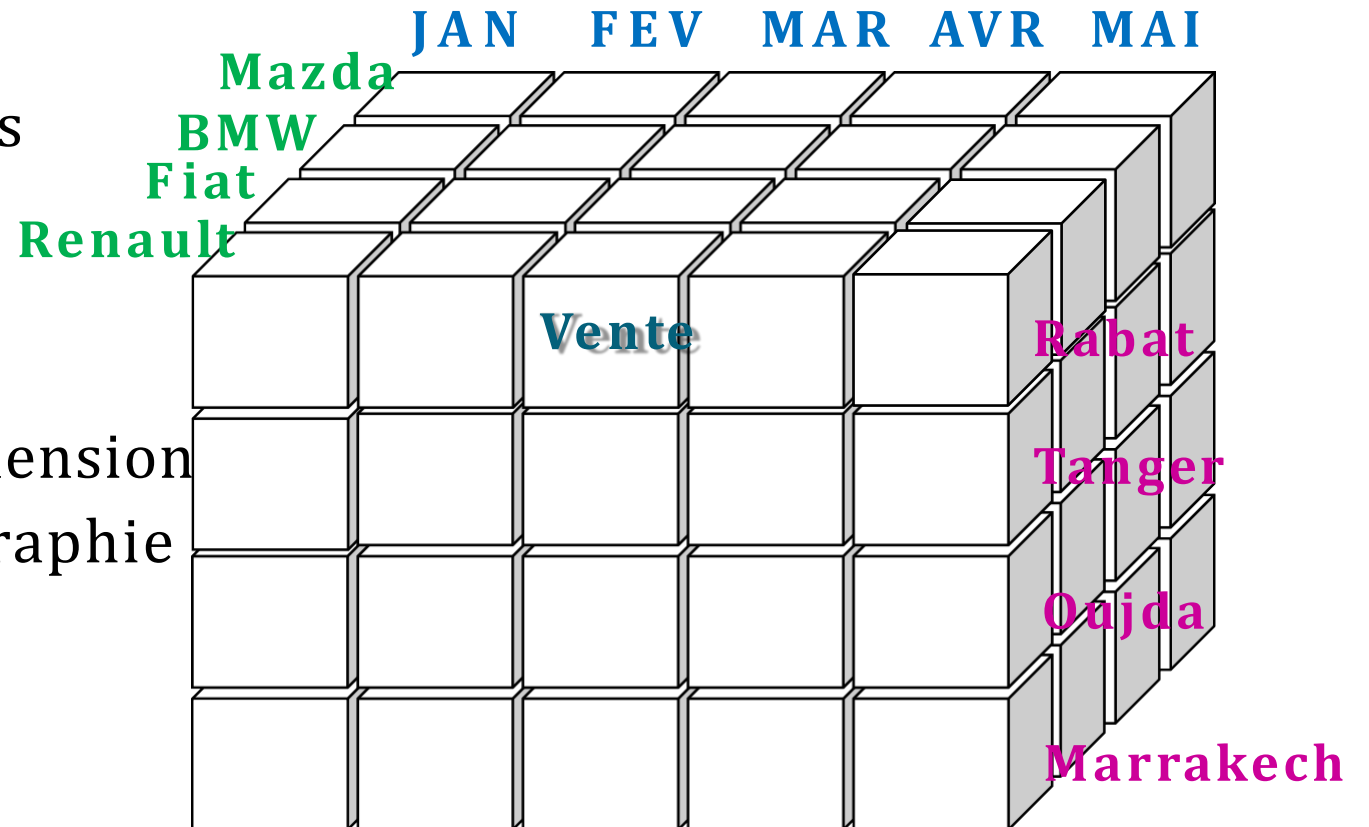
# MESURES ET DIMENSIONS

## Mesure

Stockée dans les tables de faits  
Exemple: Montant des ventes

## Dimension

Stockée dans les tables de dimension  
Exemple: Produit, Date, Géographie



# DIMENSIONS

Une dimension contient des attributs, des hiérarchies, et des membres

## **Attribut**

Une colonne de la table

## **Membre**

La valeur réelle des données dans un attribut.

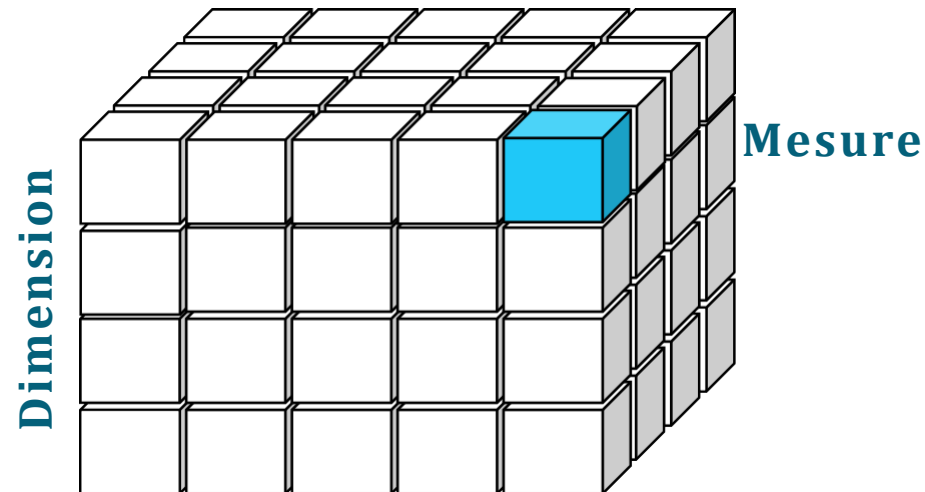
## **Hiérarchie**

Les hiérarchies sont représentées par des relations de groupement

# CUBE OLAP

Un cube OLAP multidimensionnel est un hyper-cube où:

- Les axes ou **dimensions** sont définies par l'utilisateur
- Les points dans l'espace (cellules) contiennent des **mesures** calculées à partir de formules plus ou moins complexes.
- Les opérateurs sur le cube sont **algébriques** (retournent un cube) et peuvent ainsi être combinés



# OPÉRATIONS OLAP - ROTATE

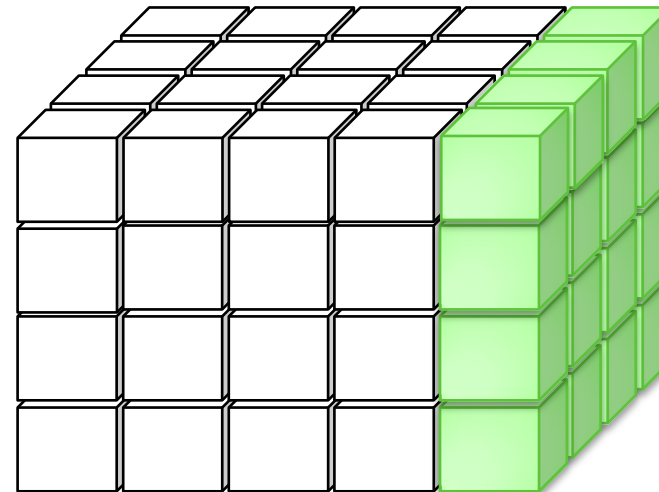
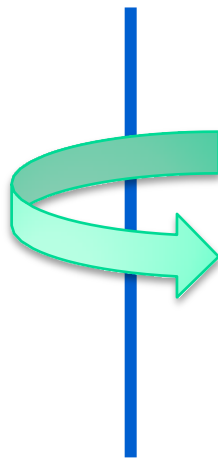
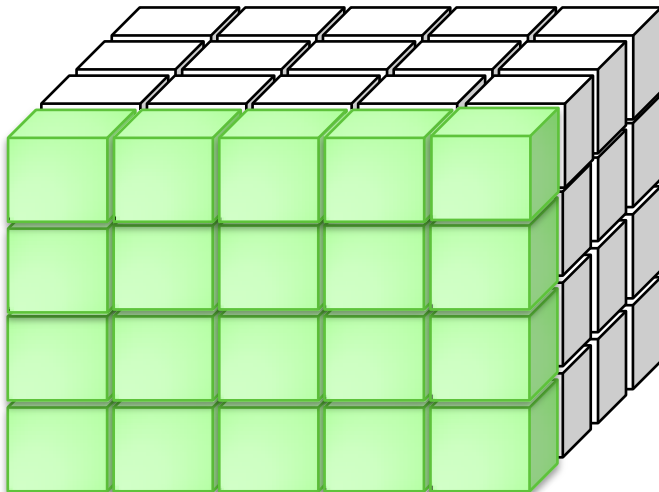
## Rotate

Rotation des axes du cube pour fournir une vue alternative des données

**(Produit , Date) → (Région , Produit)**

	MAZDA	BMW
JAN	200	86
FEV	178	90

	RABAT	TANGER
MAZDA	102	21
BMW	35	14



# OPÉRATIONS OLAP - ROLL UP /DRILL DOWN

## **Roll-up**

- Remonter dans la hiérarchie de la dimension **Semaine → Mois**

## **Drill-down**

- Descendre dans la hiérarchie de la dimension **Mois → Semaine**



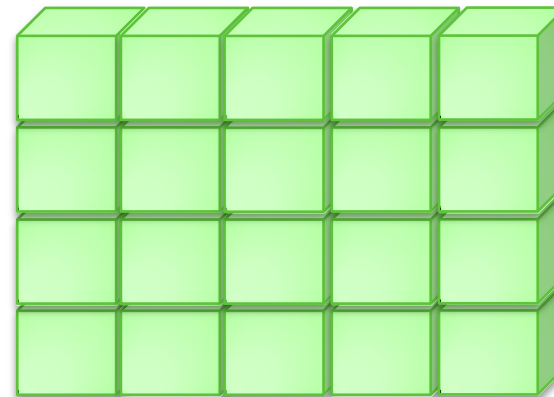
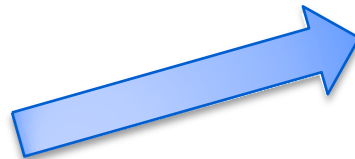
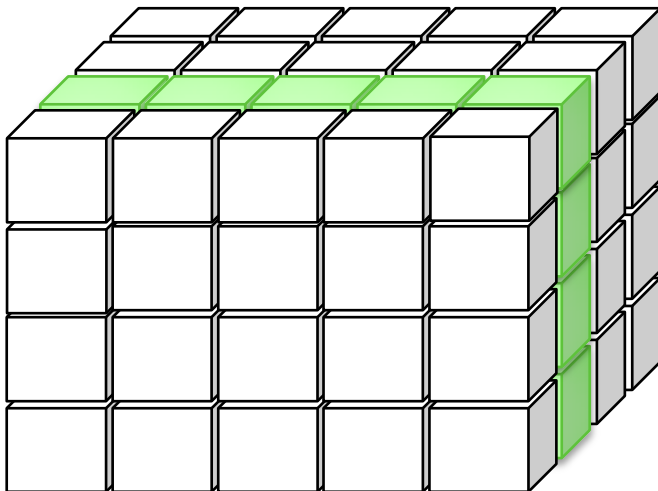
# OPÉRATIONS OLAP - SLICE

## Slicing

- Extraction d'une tranche d'informations : Sélection d'une dimension pour passer à un sous-cube

	MAZDA	BMW
JAN	200	86
FEV	178	90

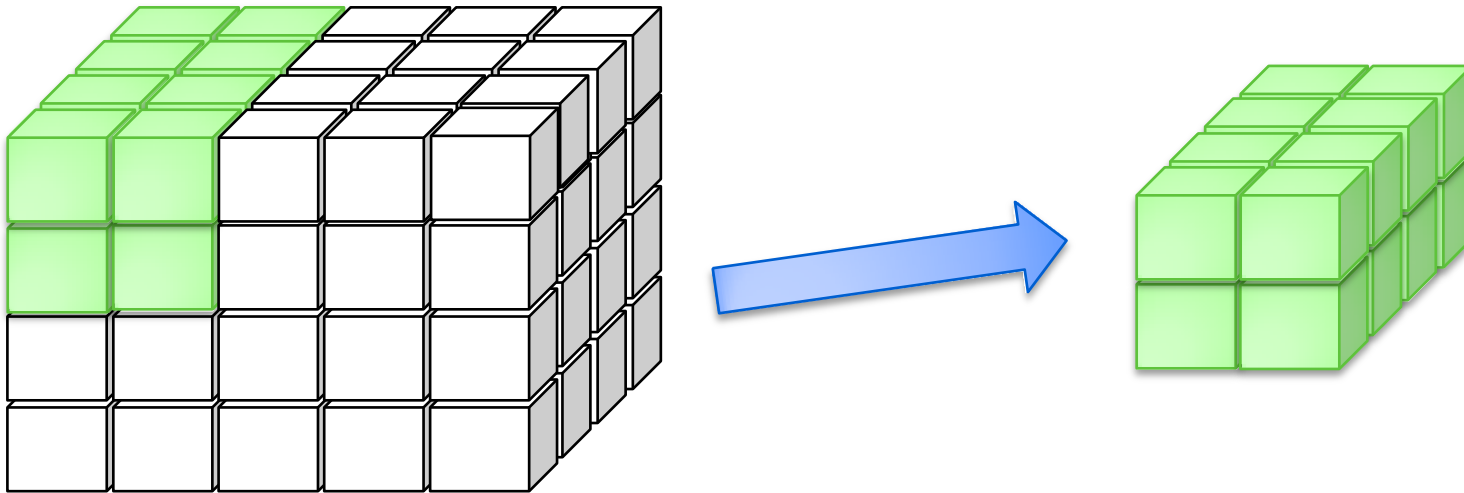
	MAZDA
JAN	200
FEV	178



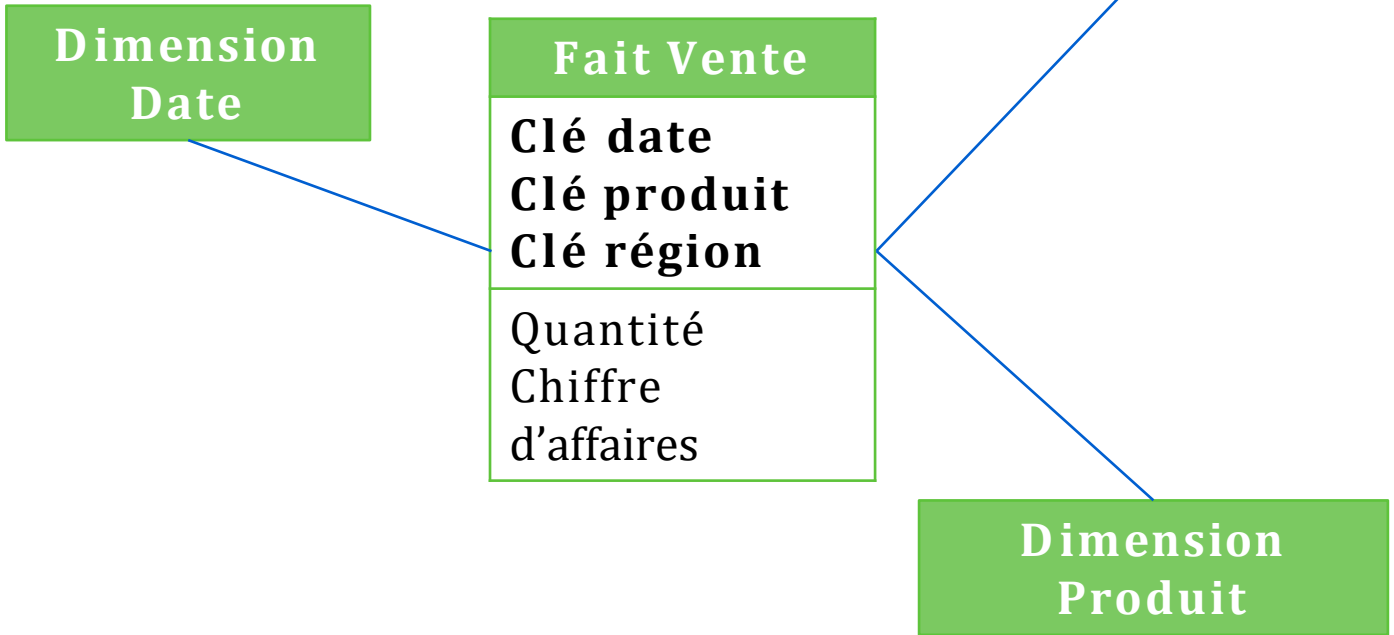
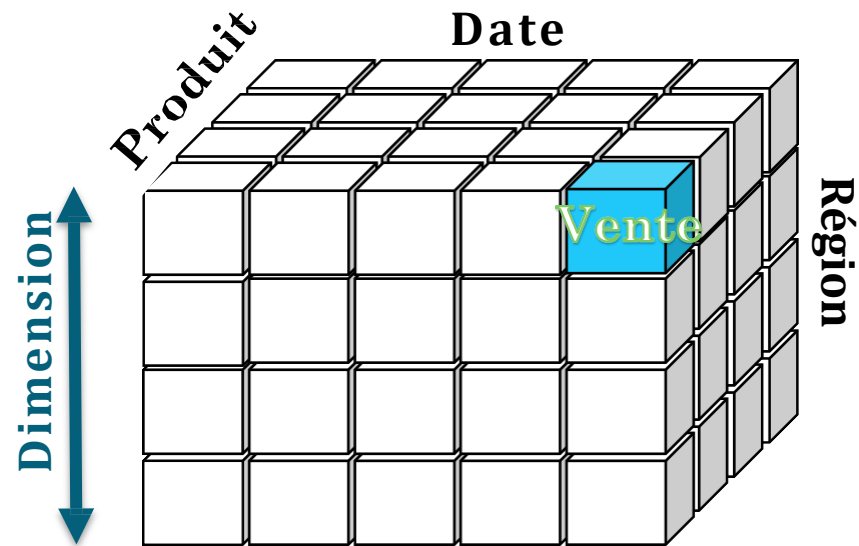
# OPÉRATIONS OLAP - DICE

## Dice

Extraction d'un bloc de données : Sélection de deux ou plusieurs dimensions



# CUBE VS SCHÉMA EN ÉTOILE



## LANGUAGE MDX

- Le langage de requête d'un cube s'appelle le **MultiDimensional eXpressions Language (MDX)**.
- MDX a été adopté par une grande majorité de fournisseurs OLAP et est devenu la norme pour les systèmes OLAP

# LANGAGE MDX – SYNTAXE

```
SELECT
    { [Measures].[Sales Amount],
      [Measures].[Tax Amount] } ON
COLUMNS,
    { [Date].[Fiscal].[Fiscal Year].&[2002],
      [Date].[Fiscal].[Fiscal Year].&[2003] }
ON ROWS
FROM [Adventure Works]
WHERE ( [Sales Territory].[Southwest] )
```

## Requête MDX

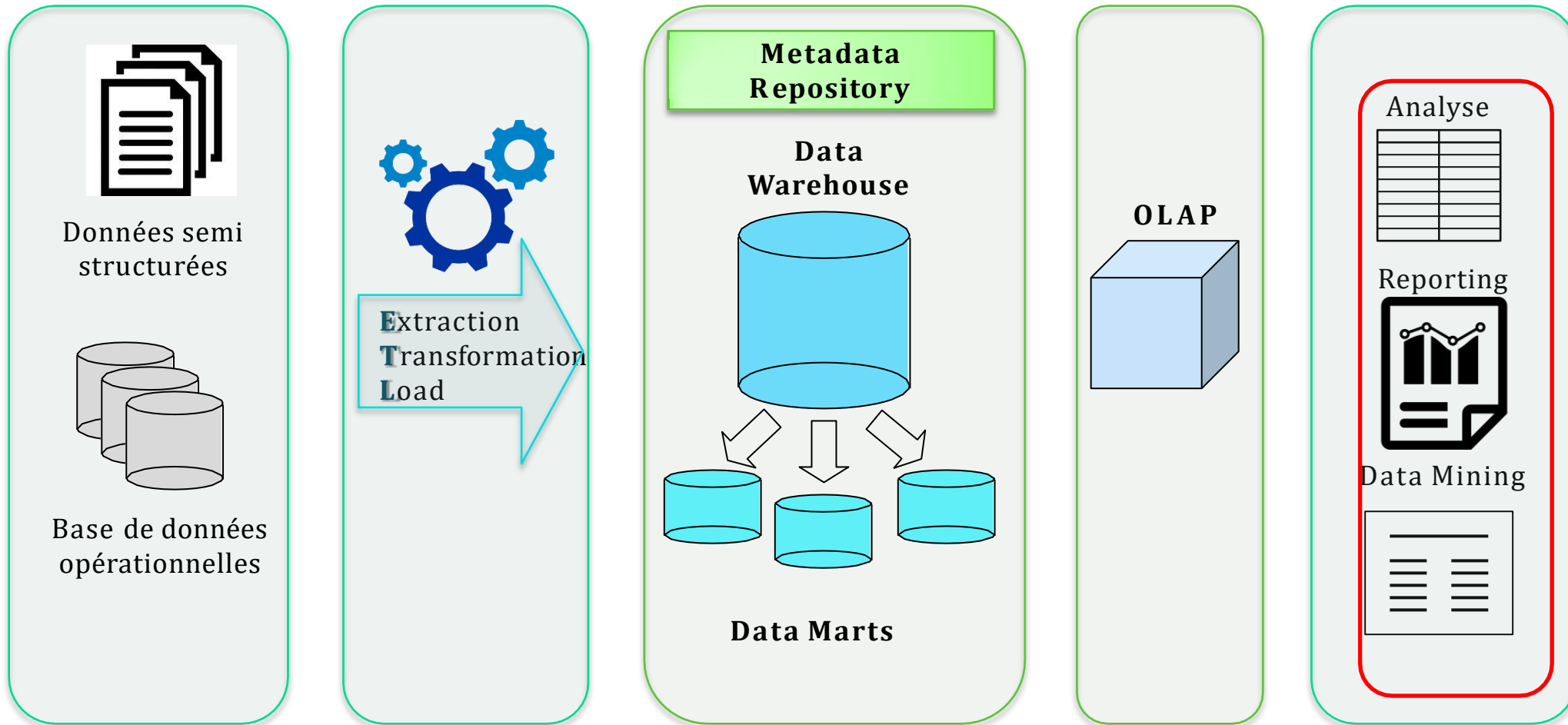
**Clause FROM** : Sélectionne un cube

**Clause SELECT** : Définit les axes

**Clause WHERE** : Définit un slice

# RESTITUTION DES DONNÉES REPORTING

# ARCHITECTURE



## KEY PERFORMANCE INDICATORS - KPI

- Les indicateurs clés de performance (KPI) définissent un ensemble de chiffres clés utilisés pour mesurer par rapport à une cible, à un benchmark ou à une date.
- Les indicateurs de performance clés sont utilisés en Business Intelligence pour évaluer l'état actuel de l'activité, évaluer la performance stratégique et prescrire un plan d'action.



## TABLEAU DE BORD

Un tableaux de bord fournit des affichages visuels d'informations importantes consolidées et organisées sur un seul écran, de sorte que les informations puissent être visualisées d'un seul coup d'œil et facilement explorées.

# TABLEAU DE BORD

Un tableau de bord:

- **Utilise des composants visuels** (graphiques, barres de performance, jauges, compteurs, ...) pour mettre en évidence, en un coup d'œil, les données et les exceptions qui nécessitent une intervention
- **Est transparent pour l'utilisateur**, ce qui signifie qu'il nécessite une formation minimale et qu'il est extrêmement facile à utiliser
- **Combine les données de divers systèmes** en une vue unique, résumée et unifiée de l'entreprise
- Activer l'analyse ou l'accès au détail aux sources de données sous-jacentes ou aux rapports
- **Présente une vue dynamique** du monde réel avec des mises à jour de données rapides

# TABLEAU DE BORD



# SECTEURS D'APPLICATIONS

## Banque & Assurance

- Profiling des Clients

## Commerce

- Ciblage des clients
- Promotions

## Santé

- Epidémiologie

## Ressources Humaines

- Gestion et affections des profils

## Télécommunication

- Analyse des appels

## Transport

- Gestion de la logistique

# APPLICATION DÉCISIONNELLE

**Objectif**

Gagner 2 points de part de marché

**Processus**

Développer le marché de la distribution spécialisée

**Activité**

Prospecter de  
nouveaux clients

Développer la  
clientèle existante

Suivre l'action  
commerciale

**Objectif  
opérationnel**

Développer les implantations  
dans 100 nouvelles agences

Passer le CA moyen de 150 à  
200 milles DH

Favoriser les promotions si  
CA/agence

**Indicateur**

- Nombre de visite  
- Taux de nouveaux clients

- CA moyen/commande  
- Nombre de client < 200M DH

- Taux d'offres promotionnelles  
- % CA généré

# OUTILS BI

## Data Mining/ML



## BI/Visualization



## ETL/DW



## Data Processing



## Data Storage

