

# Entrepôts de données

D. Boukraâ – 2020/2021 Université de Jijel

# Chapitre 2 : Modélisation Multidimensionnelle

## INTRODUCTION

- Modèle de données, niveau de modélisation, normalisation

- Définition d'un modèle selon Torlone

« Un modèle de données est un ensemble de concepts et de relations entre concepts pour représenter une réalité données »

## INTRODUCTION

- Modèle de données, niveau de modélisation, normalisation
  - Niveaux de modélisation
    - *Niveau conceptuel* : concepts de la réalité, loin de la machine (e.g. entité/association, formalisme individuel (MCD), etc)
    - *Niveau logique* : premiers choix d'implémentation, plus proche à la machine (e.g. relationnel, hiérarchique, etc.).
    - *Niveau physique* : structures de stockage, d'optimisation, très proche à la machine (e. g. tables, index, etc)

## INTRODUCTION

- Modèle de données, niveau de modélisation, normalisation
  - **Normalisation**
    - Garantir l'intégrité des données.
    - Différentes formes normales (1NF, 2NF, 3NF, BCNF, etc)
    - 3ème forme normale satisfaisante
  - **La normalisation est**
    - ☺ Adaptée à une utilisation transactionnelle des données (base de données OLTP)
    - ☹ Inadaptée à une utilisation analytique (entrepôts de données)

## Modélisation multidimensionnelle

- **Principes de base** : la modélisation multidimensionnelle repose sur les principes suivants
  - Intérêt porté sur une partie des données /pas sur la totalité □ seulement les données nécessaires à la prise de décision.
  - Modèle de données intuitif → vision que portent les analystes et décideurs sur les données.
  - Tolérance de violation de certains principes de modélisation classiques (formes normales) en renforçant les contrôles d'intégrité.

## Modélisation multidimensionnelle

- **Vision analytique des données** → Objectif : analyser les performances et l'activité d'une entreprise

Comment savoir si mon entreprise est performante ? → ex : chiffre d'affaire  
Dans le cas de bonne (ou mauvaise performance) :

Comment savoir

- Quel produit à causé la bonne (ou mauvaise) performance ?
- Quelle période de temps ?
- Quelle région ?



## Modélisation multidimensionnelle

- **Vision analytique des données**

- Analyser un ensemble de **FAITS** qui se produisent dans la réalité par rapport à des **DIMENSIONS** pour comprendre et mieux expliquer les faits
- Ex :
  - Faits (facts) : ventes
  - Dimensions (dimensions) : produits, régions, temps
- La modélisation multidimensionnelle se base sur la dualité FAIT / DIMENSION (Fact/Dimension duality)

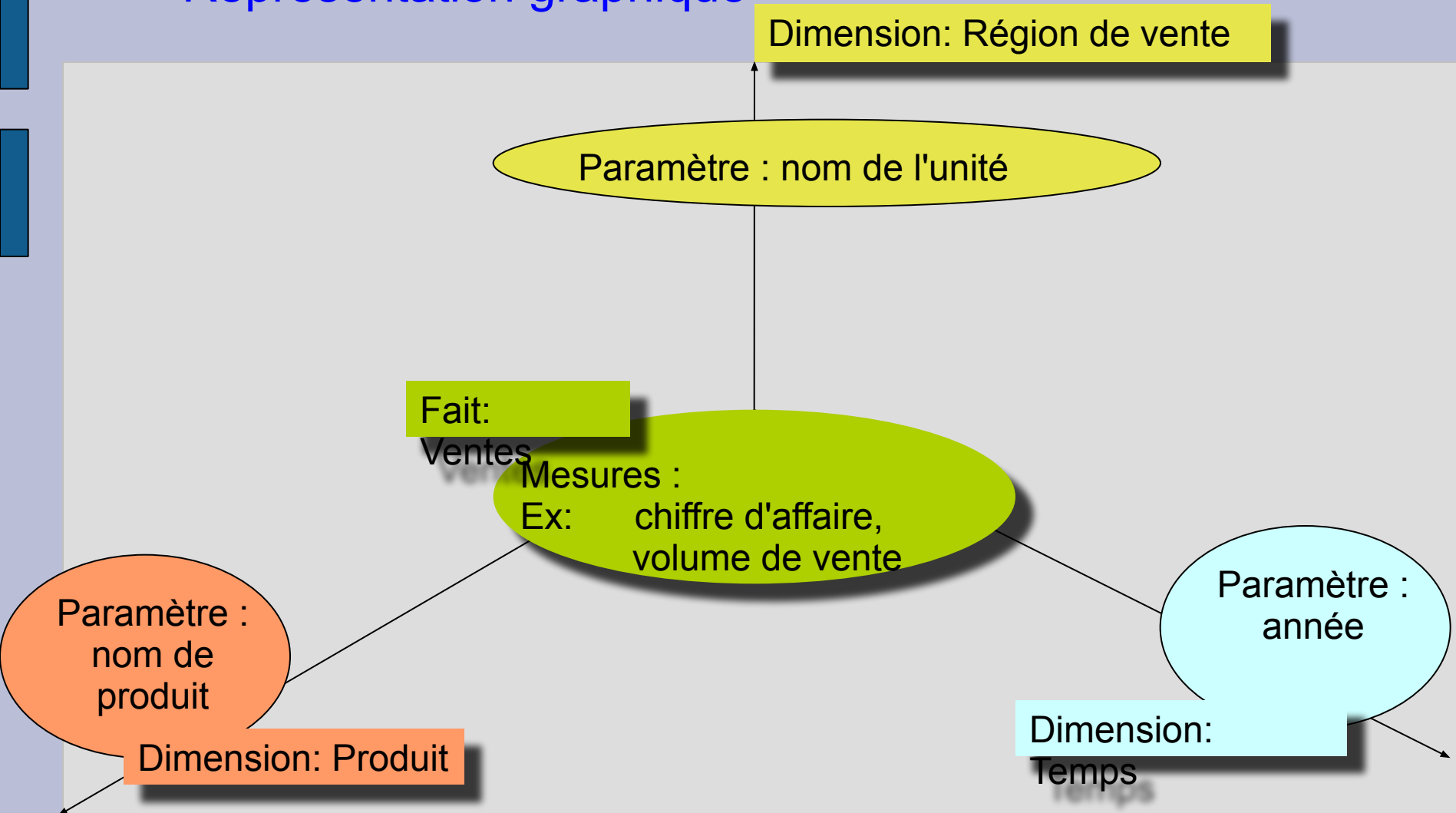


## Modélisation multidimensionnelle

- Vision analytique des données
  - Un fait est analysable selon des INDICATEURS appelés **MESURES**
  - Une dimension explique le fait par des **Paramètres**
  - Une dimension peut être **hiérarchisée** pour permettre différents groupements de données

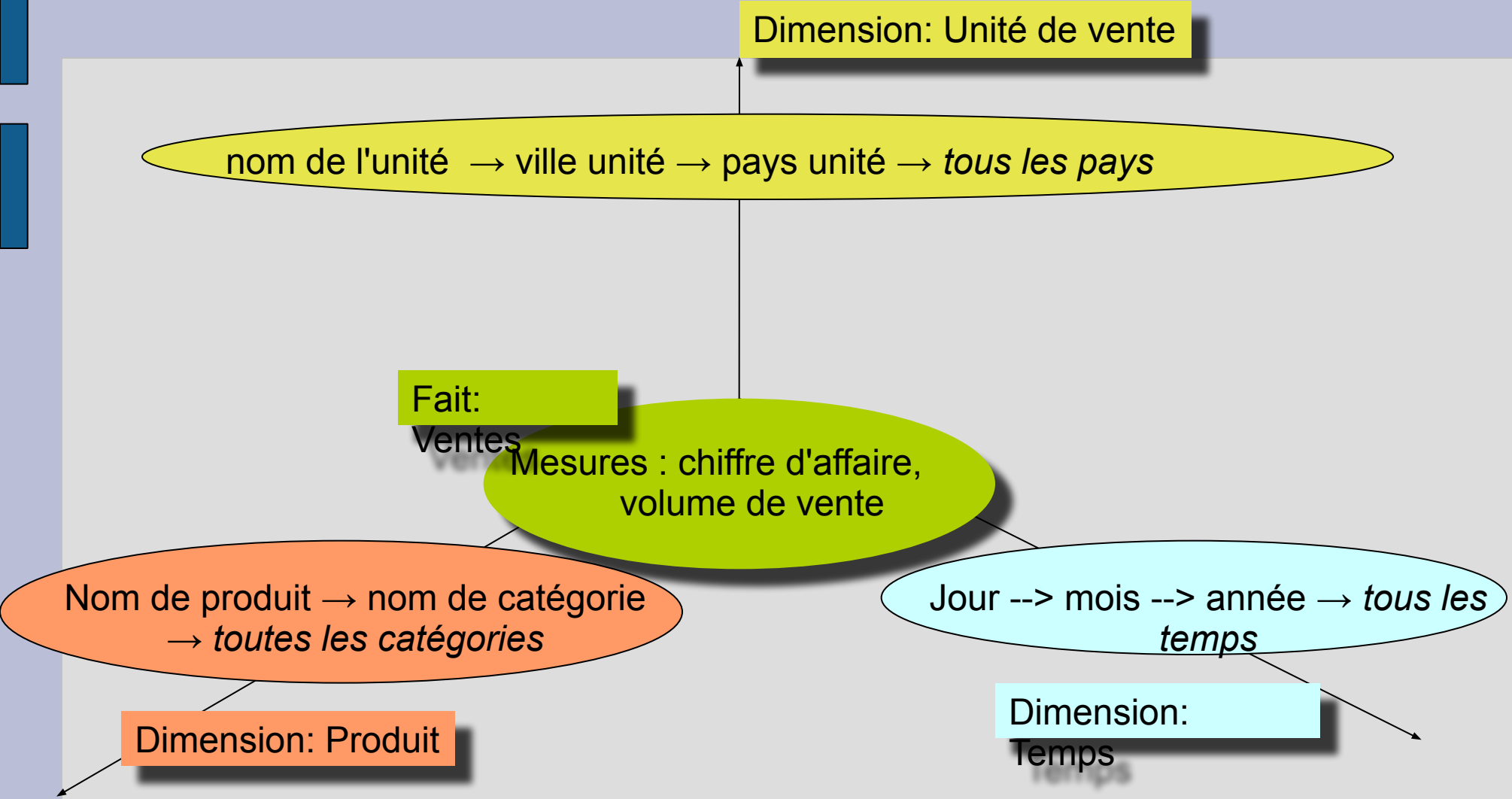
# Modélisation multidimensionnelle

- Représentation graphique



# Modélisation multidimensionnelle

- Hiérarchies de dimensions



## Modélisation multidimensionnelle

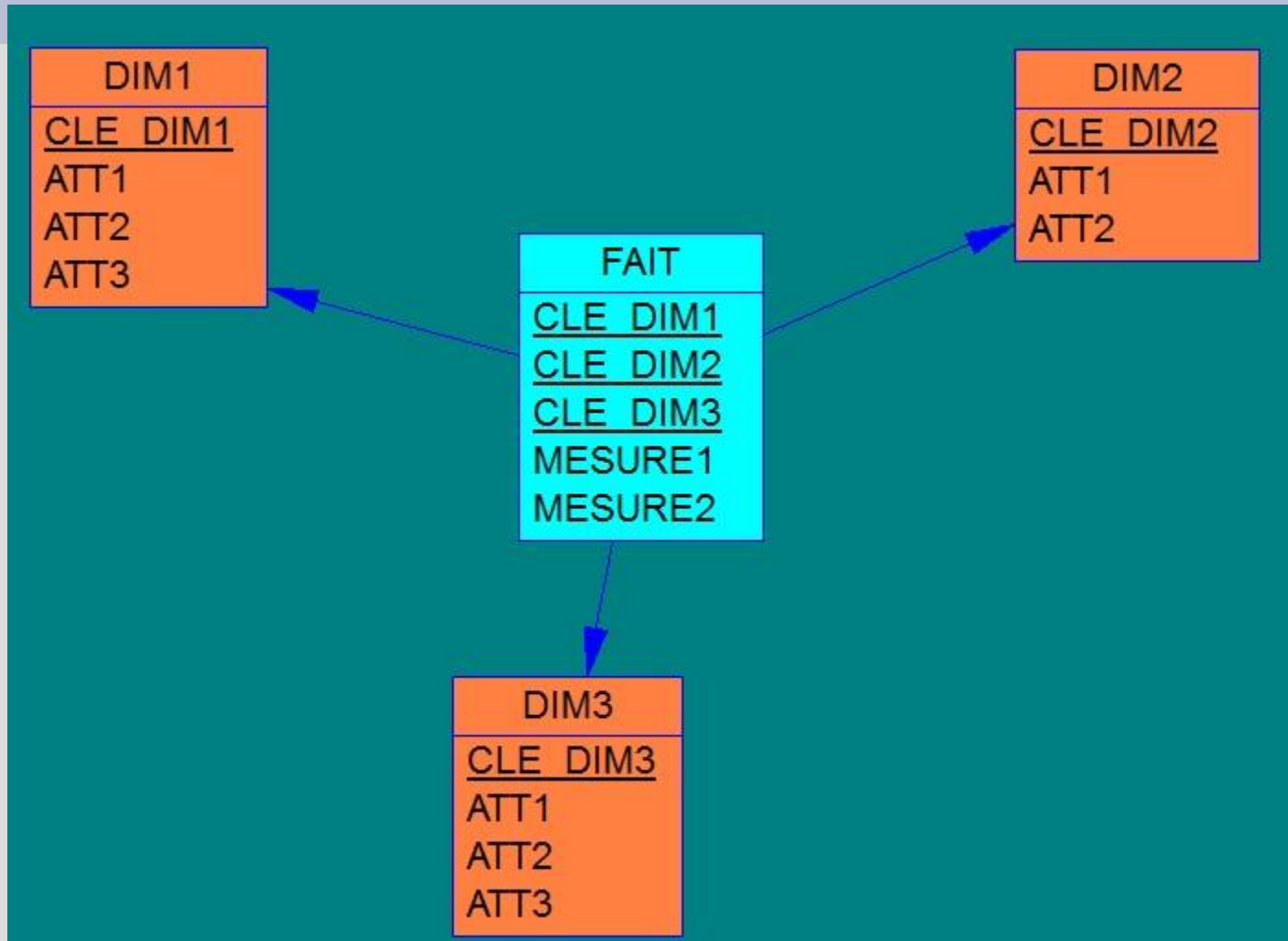
- **Modèle multidimensionnel au niveau conceptuel**
  - Absence de modèle multi-dimensionnel standard
  - Accord sur la dualité Fait/Dimension
- **Modèles logiques**
  - Dominance de la technologie relationnelle
  - Schéma en étoile (Kimball)
  - Variantes du schéma en étoile (flocon de neige, mixte)

## Schéma en étoile

- Principe du schéma en étoile (star schema)
  - Basé sur le paradigme relationnel
  - Composé d'
    - **Une** table de faits contenant les mesures
    - **Plusieurs** tables de dimension contenant les paramètres de description des mesures

## Schéma en étoile

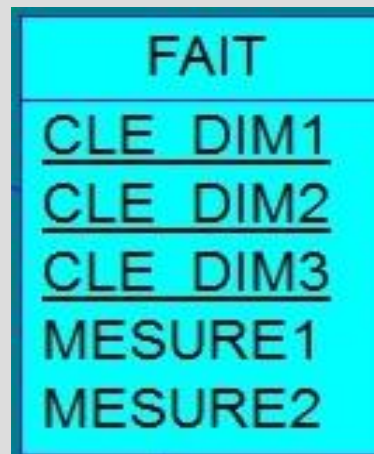
- Représentation graphique



## Schéma en étoile

- Table de faits

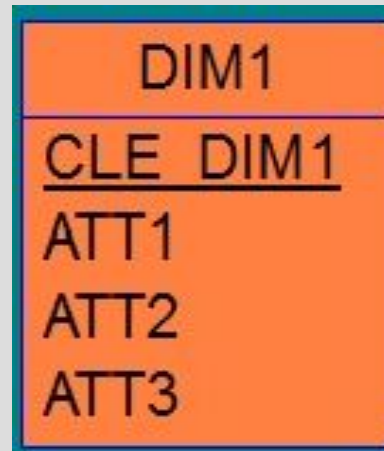
- Mesures de base ou calculées.
- Mesures au niveau le plus bas de détail ou agrégées.
- Une clé étrangère par dimension.
- Peut contenir un attribut d'identification (ex : numéro du ticket de caisse).



## Schéma en étoile

- Table de dimension

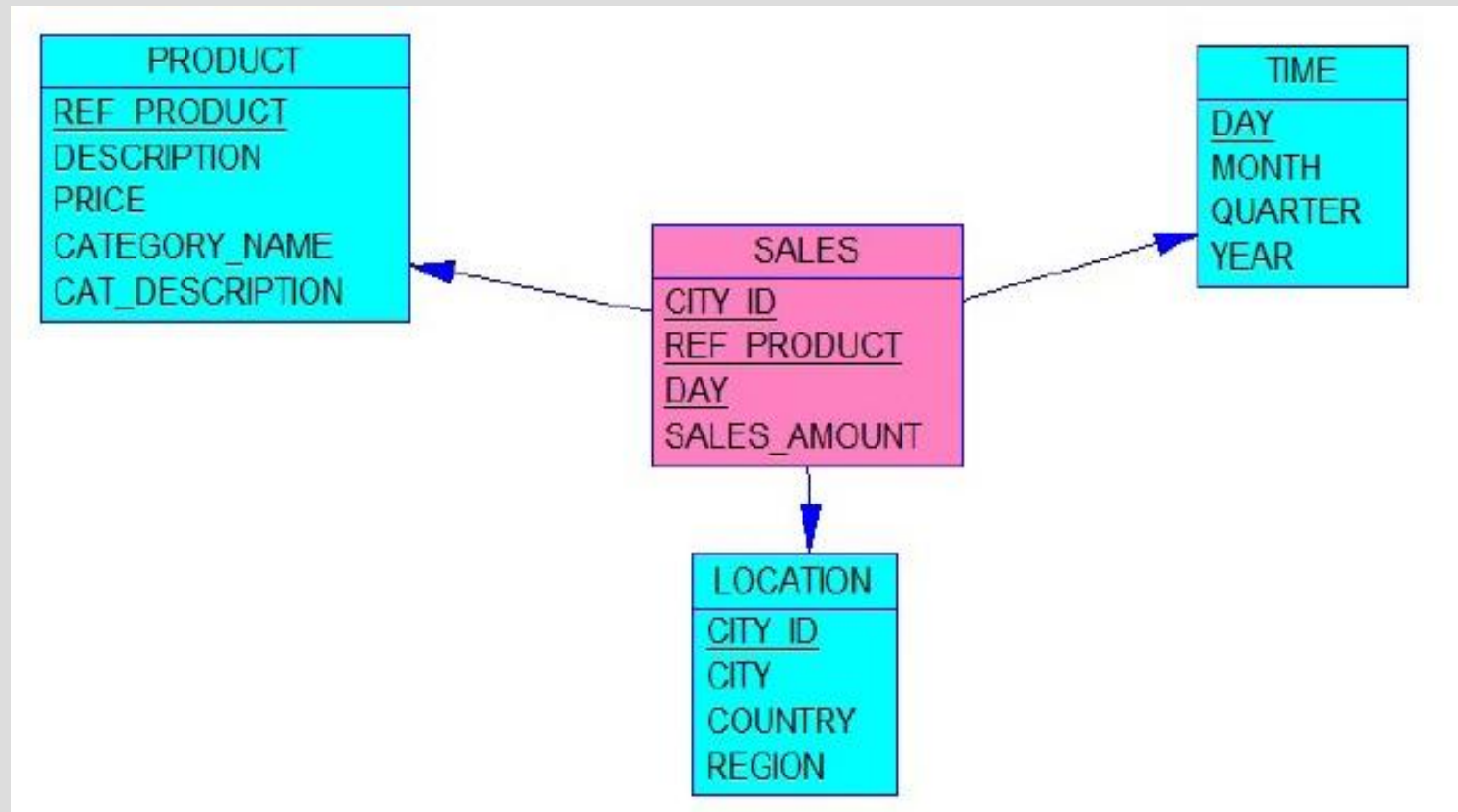
- Une clé identificatrice, en général artificielle
- Un ensemble d'attributs descriptifs
- Relation dénormalisée par rapport à la 3ème FN





## Schéma en étoile

- **Exemple:** analyser les montants des ventes par période de temps (time), par produit (product) et par endroit (location)



## Schéma en étoile

- Granularité (niveau de détail) de la table de dimension
  - Une dimension peut être hiérarchisée → contenir plusieurs niveaux hiérarchiques
  - Granularité : niveau de détail d'une dimension
  - Ex :
    - *Date: jour, mois, année, décennie*
    - *Produit : produit, catégorie*
    - *Zone : ville, wilaya, pays, région*
  - La définition de la granularité dépend des besoins d'analyse et de la disponibilité des données détaillées.

## Schéma en étoile

- Granularité de la table de faits
  - Suit la granularité de toutes les dimensions → dépend des besoins d'analyse
  - Exemples de granularité de la table de fait : analyser le volume de vente par
    - Produit, Jour, Ville
    - Catégorie, Jour, Ville
    - Catégorie, Mois, Ville
    - Catégorie, Mois, Wilaya
    - Catégorie, Mois, Pays
    - Catégorie, Années, Pays
    - ...

## Schéma en étoile

- Granularité de la table de faits

- Chaque niveau de granularité est traduit par un groupement de données

- Ex :

- **SELECT SUM(quantite\*prix) ...**

- **GROUP BY Produit, Jour, Ville**

- **SELECT SUM(quantite\*prix) ...**

- **GROUP BY Catégorie, Jour, Ville**

- ...

Dans le cas de la granularité la plus fine, la table de faits est jointe aux dimensions sans groupements

## Schéma en étoile

- Additivité des mesures

- Définit si les valeurs peuvent être regroupées (en particuliers additionnées) selon toutes les dimensions à différents niveaux de granularité.
- L'additivité ne concerne pas la possibilité de regrouper mais le sens donné aux groupement.
- **Mesures additives :**
  - Ex : Volume de ventes additif par rapport à toutes les dimensions
- **Mesures semi-addititives**
  - Ex : solde d'un client non additif par rapport au temps mais additif / région
- **Mesure non-additive :**
  - Ex : prix unitaire non additif par rapport à aucune dimension

## Schéma en étoile

- Caractéristiques du schéma en étoile

- **Avantages**

- Facilité de navigation (peu de tables).
- Requêtes simples (peu de jointures).

- **Inconvénients**

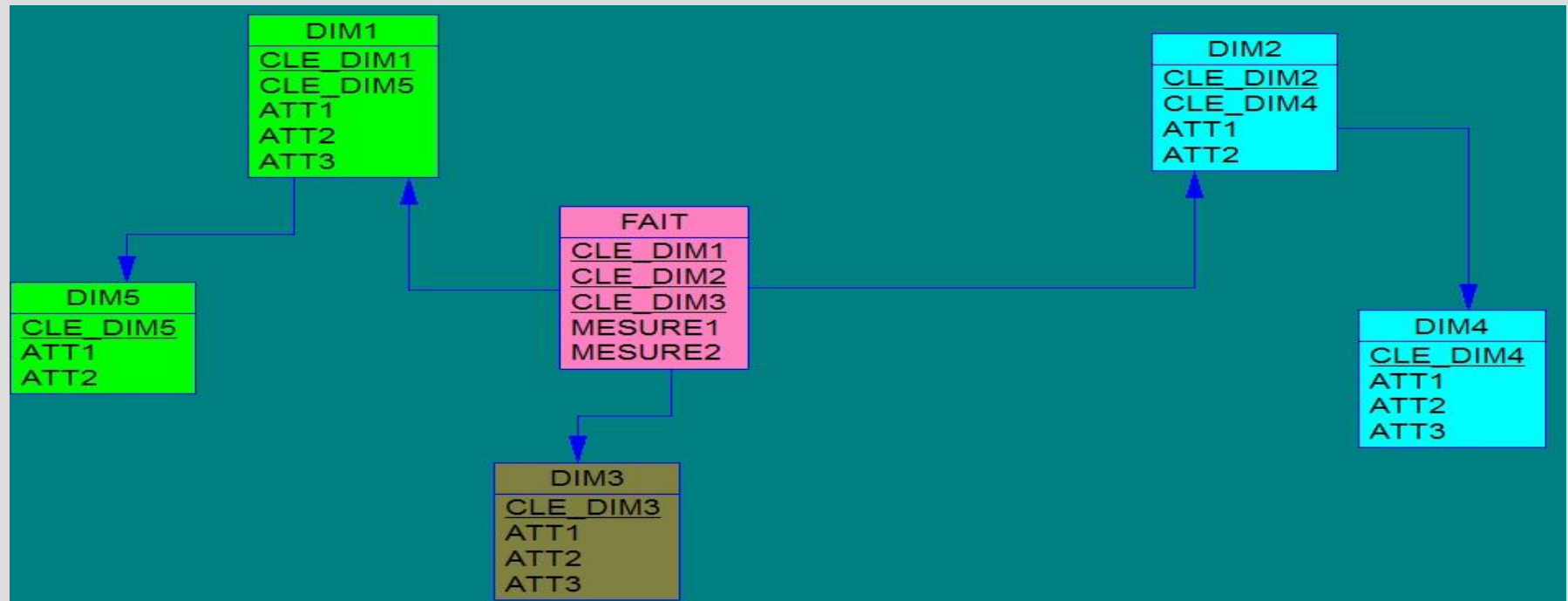
- Alimentation complexe, surtout lors de l'ajout de nouvelles instances de dimensions.
- Redondance dans les dimensions.
- Hiérarchies implicites.

## Schéma en étoile

- Variantes du schéma en étoile

- Le schéma en flocon de neige (snow-flake schema)

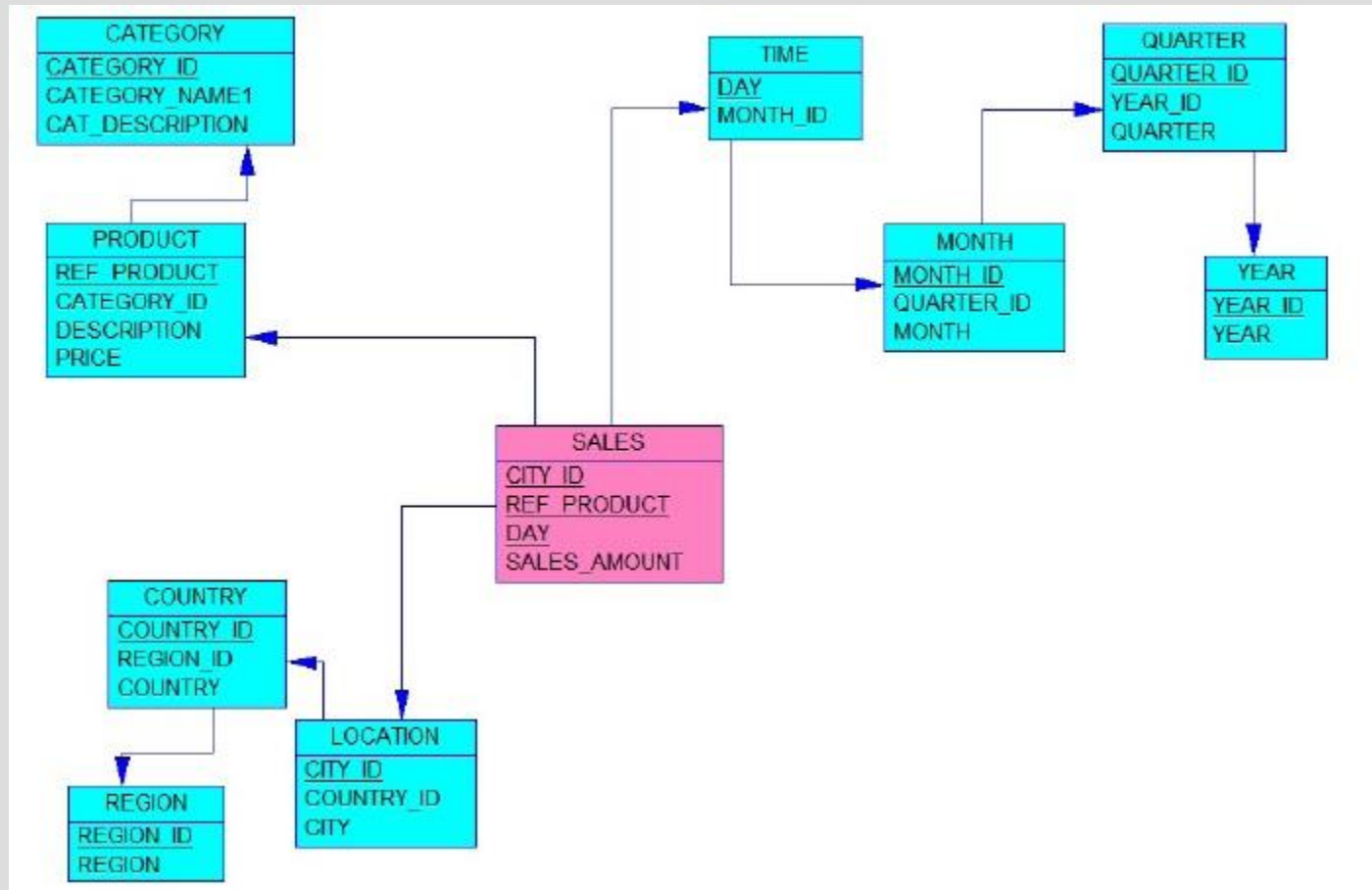
- Normalisation de toutes les dimensions (3NF)
    - Explicitation des hiérarchies



## Schéma en étoile

- Variantes du schéma en étoile

- Le schéma en flocon de neige: Exemple





## Schéma en étoile

- Variantes du schéma en étoile

- **Avantages**

- Volume réduit par / schéma en étoile
    - Hiérarchies explicites
    - Facilité d'alimentation

- **Inconvénients**

- Requêtes difficiles (beaucoup de tables)
    - Requêtes complexes (jointures)

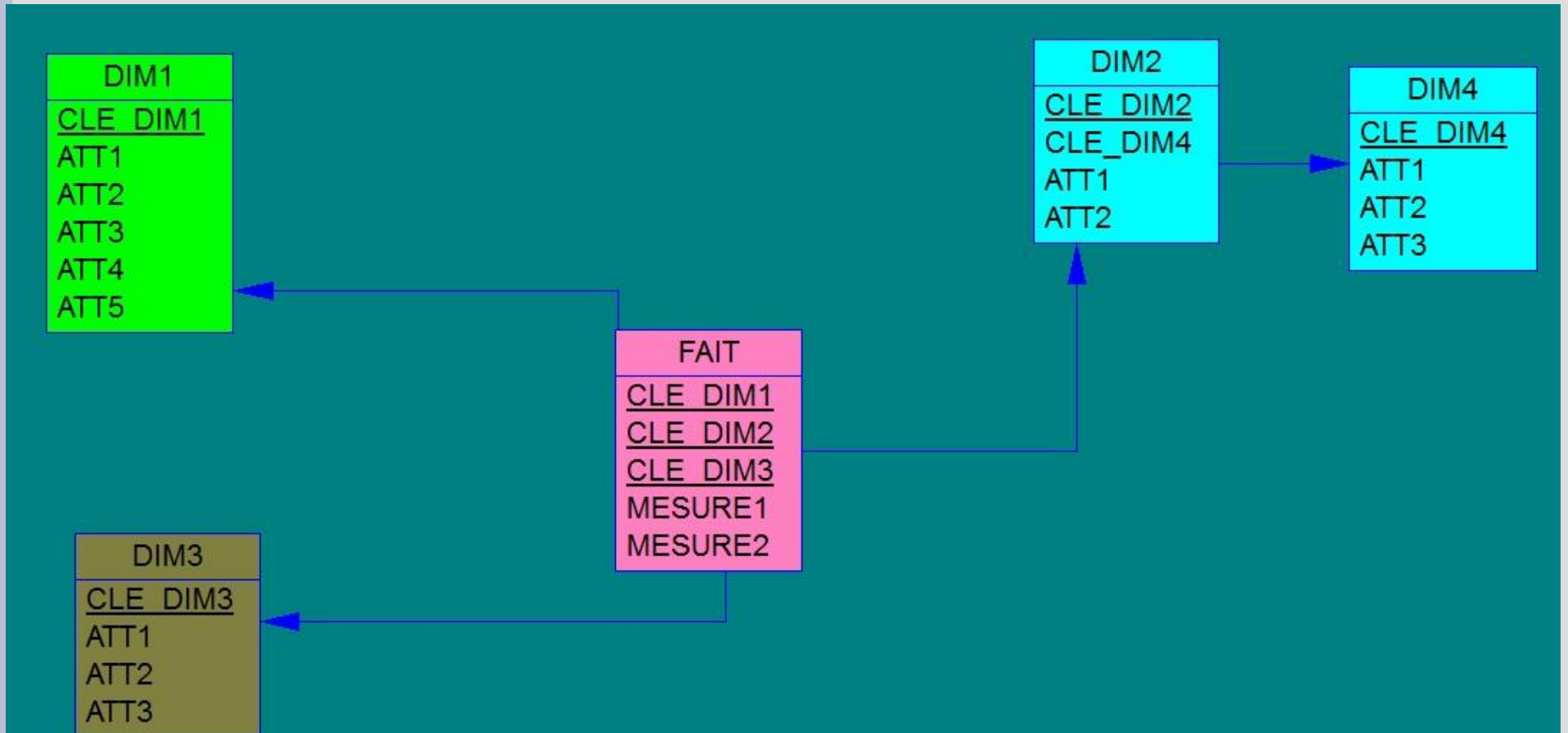
## Schéma en étoile

- Variantes du schéma en étoile
  - Le schéma mixte (étoile /flocon de neige)
    - Combine les deux schémas précédents (étoile & flocon de neige)
    - Normalisation de certaines dimensions seulement
    - Le choix des dimensions à normaliser dépend de l'utilisation.

## Schéma en étoile

- Variantes du schéma en étoile

- Le schéma mixte (étoile /flocon de neige)

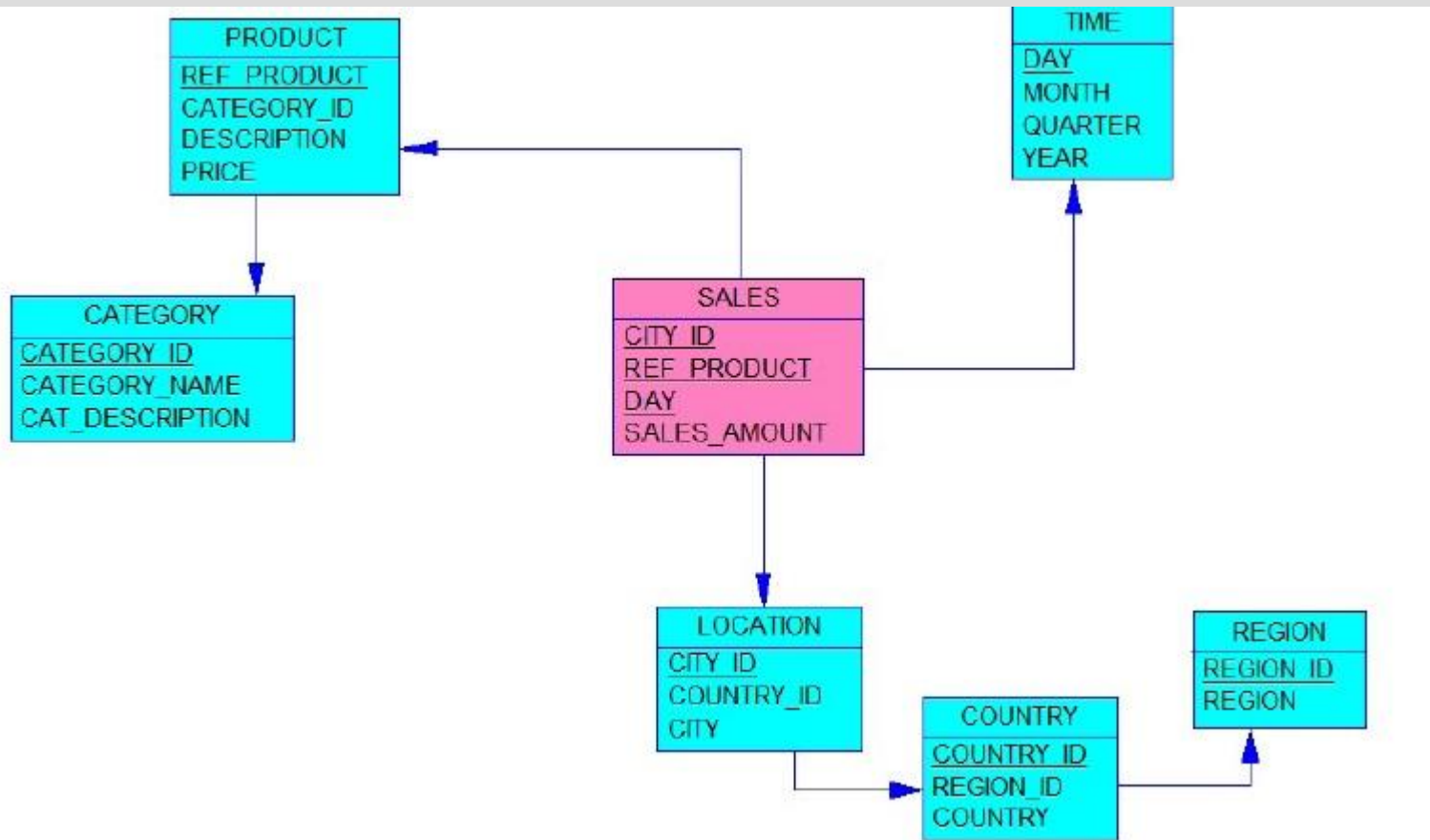


## Schéma en étoile

- Variantes du schéma en étoile

- Le schéma mixte (étoile / flocon de neige)

### Exemple



## Schéma en étoile

- Extension du schéma en étoile

- La constellation de faits : faits dépendant de quelques dimensions seulement (conformité des faits)

