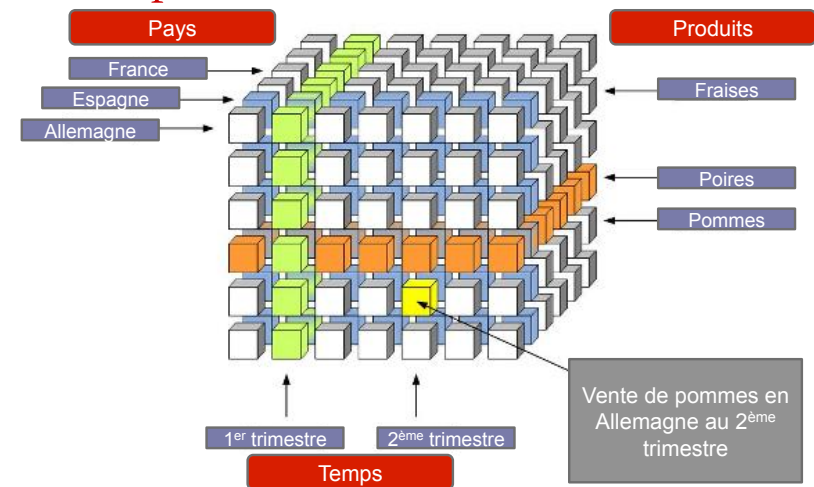


## Chapitre 2

# Conception logique

- 1) Introduction
- 2) Différents modèles
- 3) Caractéristiques

## Exemple



## Introduction

- **Modèle multidimensionnel**
  - Modéliser une base décisionnelle à partir de l'identification des faits à analyser et des dimensions d'analyses qui leur sont associées.
- Deux notions de base
  - **Les faits**
    - Donnée numérique (économique) à analyser servant de base à la définition des indicateurs.
    - Ex : nombre de ventes, chiffre d'affaires, quantité stockée...
  - **Les dimensions**
    - Axes d'analyse associés aux indicateurs.
    - Ex : Temps, clients, géographie, produit, fournisseur...

## Deux types de tables

- **Les tables de faits**
  - La table de faits est la clef de voûte du modèle dimensionnel où sont stockés les indicateurs de performances.
  - Les indicateurs étant les données les plus volumineuses d'un système d'information, il faut les rationaliser au sein des tables de faits.
- **Les tables de dimensions**
  - Entités complémentaires à la conception des tables de faits.
  - Jouent le rôle de référentiel à l'entrepôt.
  - Chaque axe du cube est représenté par une table de dimension.
  - Si la table de dimension n'existe pas on dit que la dimension est **dégénérée**.

## Les tables de faits

- Composition typique
  - Un ou plusieurs **attributs de dimension** servant à caractériser les attributs de faits et permettant de conduire les analyses.
  - Zéro ou plusieurs **attributs de faits**.
  - Eventuellement un attribut spécifique **clé primaire** s'il est nécessaire d'identifier chaque tuple.
  - Peu de colonnes mais beaucoup de lignes.

Ventes
Id_Date (CE)
Id_Produit (CE)
Id_Magasin (CE)
Quantité
Montant

## Différents types de faits

- **Fait additif**
  - Additionnable dans toutes les dimensions
  - Ex : chiffre d'affaires, quantité vendue
- **Fait semi-additif**
  - Additionnable seulement selon certaines dimensions
  - Ex 1 : le nombre de clients n'est pas additif sur la dimension produit
  - Ex 2 : le niveau de stock n'est pas additif sur la dimension temps
- **Fait non-additif**
  - Non additionnable quelque soit la dimension considérée
  - Exemple : la température est non-additive sur toute dimension

## Les tables de faits

- Définition d'une table de faits
  - Identifier les faits (mesures)
  - Identifier les différentes dimensions
  - Déterminer si ces faits peuvent être rassemblés dans une même table
  - Déterminer l'opportunité de matérialiser les faits calculés
- Choisir le grain de la table de faits
  - C'est le grain le plus fin de représentation qui va directement conditionner les possibilités d'analyse
  - Attention aux volumes de stockage et aux performances

## Les tables de dimensions

- Composition typique
  - La **clé primaire** de la dimension qui est une clé étrangère dans la table de faits.
  - D'autres attributs (textuelles ou numériques) appelés **membres** permettant de caractériser la dimension.
  - Peu de lignes mais beaucoup de colonnes.

Produit
Id_Produit (CP)
Description
Référence
Famille
Type
Gamme
Rayon
...

## Les tables de dimensions

- Caractéristiques des membres
  - Les membres d'une table de dimension sont généralement dépendants et organisés en hiérarchie (dépendances fonctionnelles).
    - Référence → Famille
    - Famille → Type
    - Type → Gamme
  - Contrainte à respecter : pas de circuit dans une hiérarchie.
  - Les membres guident les opérations ROLL/UP et DRILL/DOWN.
  - Un membre peut être caractérisé par des attributs spécifiques (qui ne guident pas les agrégations mais qui peuvent servir pour des sélections).

## Chapitre 2 Conception logique

### 1) Introduction

### 2) Différents modèles

### 3) Caractéristiques

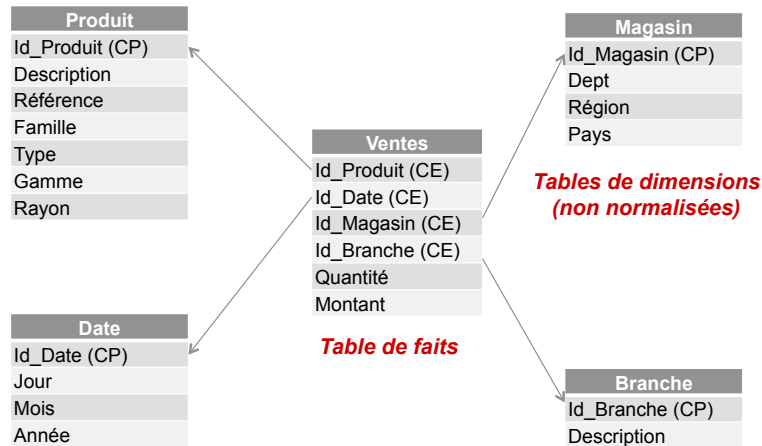
## Les tables de dimensions

- Normalisation
  - Une table de dimension non normalisée peut être décomposée totalement ou partiellement en utilisant les dépendances fonctionnelles entre ses attributs.
  - Rappel sur la normalisation 3FN par décomposition :
    - Choisir une DF dont la partie gauche PG n'est pas une superclé
    - Construire une nouvelle table Tn sur les attributs de la partie gauche PG et de la partie droite PD de la dépendance
    - Supprimer de la table initiale Ti, les attributs de PD
    - Spécifier une dépendance d'inclusion  $T_i[PG] \subseteq T_n[PG]$
    - Itérer le processus tant que possible (ou tant que souhaité)
  - Attention : La normalisation réduit les volumes et limite les incohérences mais peut altérer les performances (nécessité de jointures).

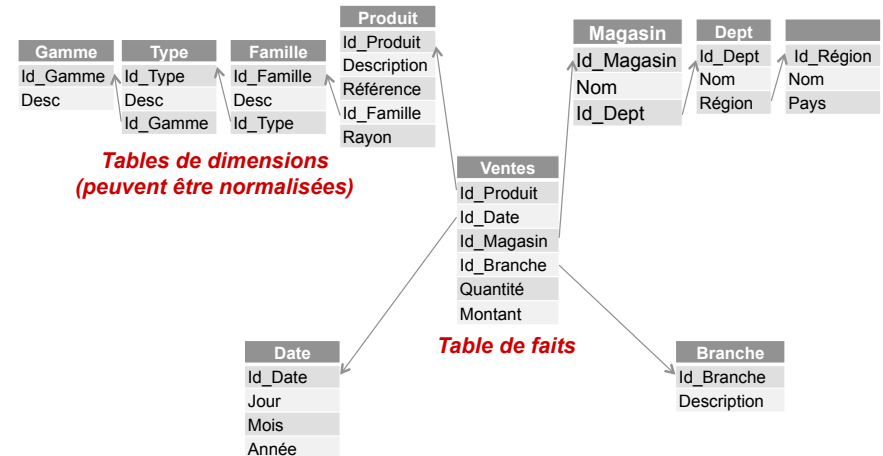
## Modélisation logique

- Trois grands types de modèles
  - **Schéma en étoile**
  - **Schéma en flocon**
  - **Schéma en constellation**

## Schéma en étoile



## Schéma en flocon



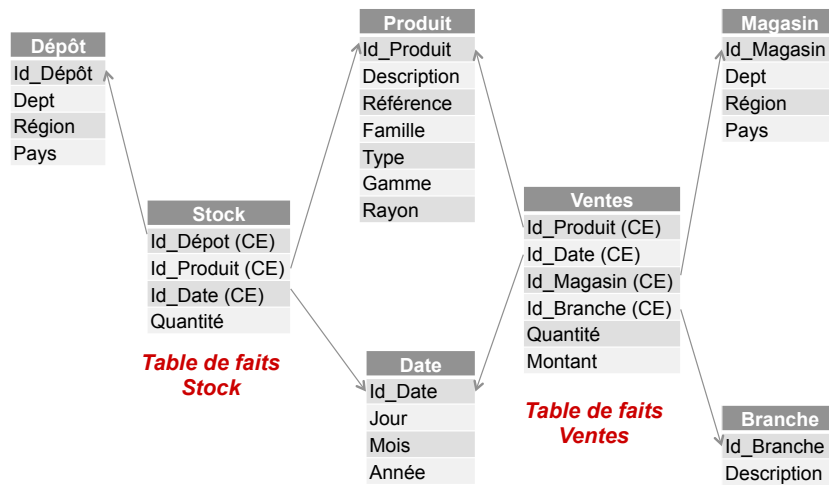
## Schéma en étoile

- **Caractéristiques**
  - Une table de faits centrale reliée par clés primaires aux tables de dimensions.
  - Chaque dimension est décrite par une seule table dont les attributs représentent les diverses granularités possibles.
- **Avantages**
  - Facilité de navigation, gestion des agrégats.
  - Peu de jointures.
- **Inconvénients**
  - Toutes les dimensions ne concernent pas toutes les mesures.
  - Redondances dans les dimensions.

## Schéma en flocon

- **Caractéristiques**
  - Une table de faits centrale reliée par clés primaires aux tables de dimensions.
  - Les dimensions sont décrites par une succession de tables (à l'aide de clefs étrangères) représentant la granularité de l'information.
- **Avantages**
  - Economie de place de stockage.
  - Évite les redondances.
- **Inconvénients**
  - Complexité du modèle.
  - Plus de jointures.

## Schéma en constellation



## Chapitre 2 Conception logique

- 1) Introduction
- 2) Différents modèles
- 3) Caractéristiques

## Schéma en constellation

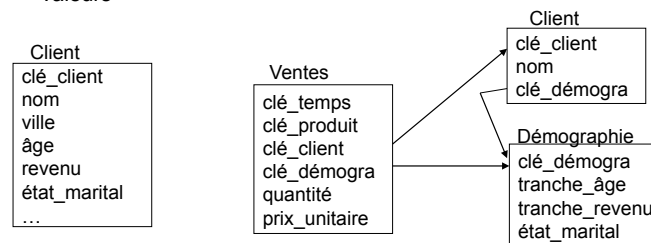
- Caractéristiques
  - Plusieurs tables de faits à dimensions partagées.
- Avantages
  - Meilleure gestion des données creuses.
- Inconvénients
  - Complexité du modèle.
- Conclusion
  - Faibles différences entre les trois schémas en terme de performances.

## La dimension temps

- Dimension toujours présente, souvent reliée à toute table de fait.
- Différencier les dates et les périodes.
- Diverses possibilités pour repérer le temps :
  - Attribut de type DATE géré avec les fonctions SQL.
  - Codification spécifique gérée directement par l'application.
  - Table Temps + Calendrier
    - Informations supplémentaires : événements, jours fériés, vacances, période fiscale, saison haute, saison basse...
- Possibilité de gérer des échelles de temps différentes (ex : calendrier civil / fiscal) avec deux tables de dimensions différentes.

## Grande dimension

- Pb des tables de dimensions avec de nombreux attributs et de nombreux tuples (Clients, Produits...)
- Solutions :
  - Évoluer vers un schéma en flocon
  - Introduire une nouvelle table de dimension contenant des intervalles de valeurs



## Ecrasement de l'ancienne valeur

- Renoncer à suivre les situations passées.
- Plutôt correction d'informations erronées.

Temps	Fait de Vente	Client
#T:JJ:MM:AA:Even	#T:#C:#P:Prix	#C: #V:Nom:SitMarital
1201:14:02:99:St Valentin	200:100:77:100	100:Didier:Divorcé:Marlé
	201:100:77:100	
	202:100:77:100	
	202:100:66:10000	
	568:100:77:20	
	1115:100:77:100	
	1116:100:77:100	
	1117:100:66:50000	
	1200:100:77:100	

## Evolution d'une dimension

- Les tables de dimension doivent pouvoir être mises à jour même si leur évolution est généralement lente
  - Ex : Changement d'adresse d'un client, d'appellation d'un produit...
- Solutions
  - Ecrasement de l'ancienne valeur
  - Versionnement
  - Valeur initiale / valeur courante
  - Valeur antérieure / valeur courante

## Versionnement

- Clé étendue d'un numéro de version.
- Partitionnement automatique de l'historique.

Temps	Fait de Vente	Client
#T:JJ:MM:AA:Even	#T:#C:#P:Prix	#C: #V:Nom:SitMarital:DateEffet
1201:14:02:99:St Valentin	200:100:1:77:100	100:1:Didier:Célibataire:10
	201:100:1:77:100	100:2:Didier:Marlé:203
	202:100:1:77:100	100:3:Didier:Divorcé:567
	202:100:1:66:10000	100:4:Didier:Marlé:1118
	568:100:2:77:20	
	1115:100:3:77:100	
	1116:100:3:77:100	
	1117:100:3:66:50000	
	1200:100:4:77:100	

## Valeur initiale / Valeur courante

- L'ancienne valeur n'est utile que pendant un certain temps pour étudier les effets d'une transition.

Temps	Fait de Vente	Client
#T:JJ:MM:AA:Event	#T:#C:#P: Prix	#C: #V:Nom:SMcour:SMorig:DateEffet
1201:14:02:99:St Valentin	200:100:77:100	100:Didier:Marie:Célibataire:1118
	201:100:77:100	102:Paul:Célibataire:NULL:NULL
	202:100:77:100	
	202:100:66:10000	
	568:100:77:20	
	1115:100:77:100	
	1116:100:77:100	
	1117:100:66:50000	
	1200:100:77:100	

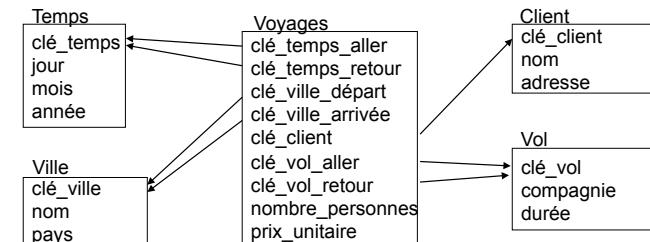
Produit
#P:Descr
66:Bague
77:Fleur

Entrepôts de données

25

## Dimension à plusieurs rôles

- Il est fréquent qu'une dimension joue plusieurs rôles relativement à la même table des faits.
- Introduction dans la table de faits d'une clé étrangère différente pour chaque rôle.



Entrepôts de données

27

## Valeur antérieure / Valeur courante

- L'ancienne valeur n'est utile que pendant un certain temps pour étudier les effets d'une transition.

Temps	Fait de Vente	Client
#T:JJ:MM:AA:Event	#T:#C:#P: Prix	#C: #V:Nom:SMcour:SMant:DateEffet
1201:14:02:99:St Valentin	200:100:77:100	100:Didier:Marie:Divorcé:1118
	201:100:77:100	102:Paul:Célibataire:NULL:NULL
	202:100:77:100	
	202:100:66:10000	
	568:100:77:20	
	1115:100:77:100	
	1116:100:77:100	
	1117:100:66:50000	
	1200:100:77:100	

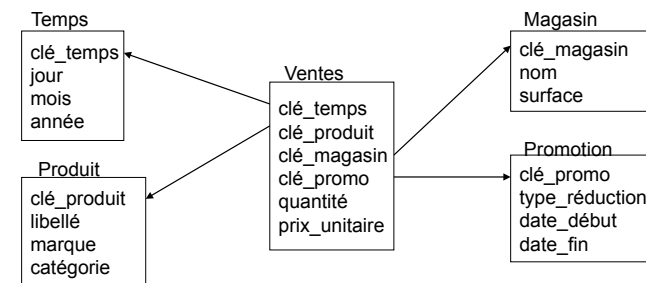
Produit
#P:Descr
66:Bague
77:Fleur

Entrepôts de données

26

## Dimension causale

- Dimension qui provoque le fait ou établit l'existence d'un fait.



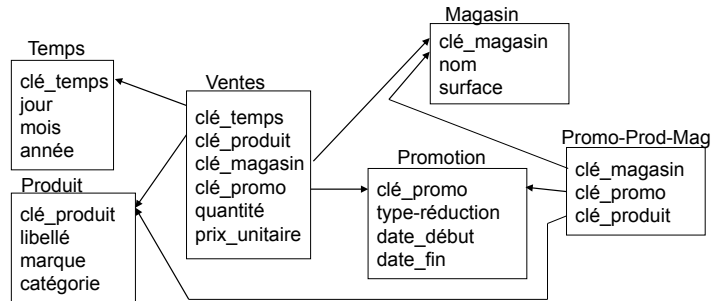
- Ce schéma en étoile ne permet toutefois pas de déterminer les produits qui étaient en promotion et qui n'ont pas été vendus.

Entrepôts de données

28

## Table d'existence

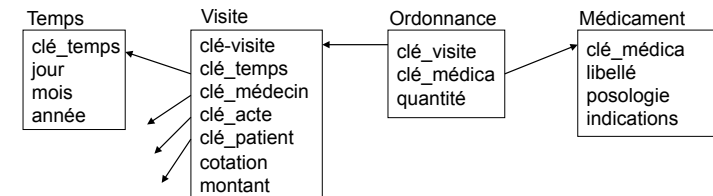
- Introduction d'une table d'existence permettant d'indiquer les produits en promotion et les magasins concernés.



- Cette table d'existence apparaît comme une table de suivi d'événements (table de faits sans fait).

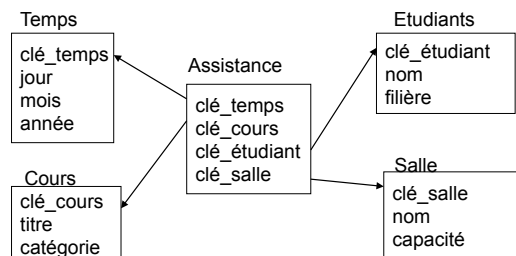
## Faits de fait

- Enregistrement de faits secondaires relativement à un fait primaire.
- La table primaire joue le rôle de table de dimension pour la table secondaire.
- Chaque tuple de la table secondaire référence le tuple concerné de la table primaire.



## Table de faits sans fait

- Tuple dans la table des faits avec uniquement les valeurs des attributs de dimension.



- Possibilité de compter des occurrences
  - Quels sont les étudiants les moins assidus, quelles sont les salles les plus utilisées...?

## Décomposition d'une table de faits

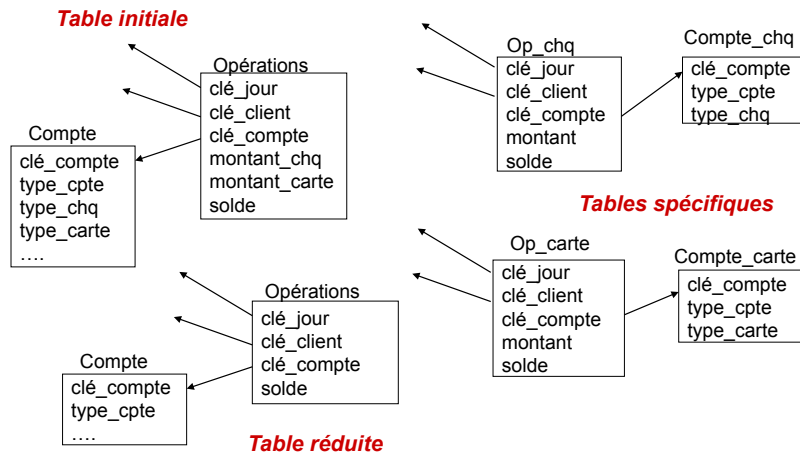
- Lorsque la table contient plusieurs faits qui se partagent à priori les mêmes dimensions mais qui présentent certains aspects spécifiques (compte chèque / compte épargne, assurance automobile / assurance habitation...).

### Solution

- Une table de faits réduite
  - Table de faits réduite aux mesures communes
  - Table de dimensions réduite aux attributs communs
- Tables de faits spécifiques par produit hétérogène
  - Table de faits avec les mesures propres au produit
  - Table de dimensions avec les attributs propres au produit



## Décomposition d'une table de faits

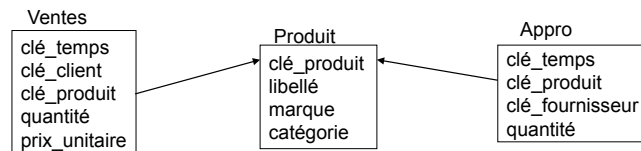


## Approche

- Etude des besoins
  - Décrire les objectifs visés, situer les types d'analyse à effectuer et les types de rapports à produire
  - Identifier les sources et évaluer les difficultés d'intégration, les volumes de données concernés (approximatifs)...
- Conception du schéma
  - Approche guidée par les besoins
  - Approche guidée par les sources
- Vérification du schéma

## Dimension conforme partagée

- Une dimension est dite **conforme** lorsqu'elle est utilisable de la même manière par plusieurs tables de faits différentes.
- Possibilité de partager une dimension conforme et de mener en conséquence des analyses utilisant simultanément les faits de ces différentes tables.



- Ex : Sur une période donnée, analyser les ventes des produits qui ont été approvisionnés par un unique fournisseur.