

EPISEN

Ing2 2021-2022

Module BDM

BASES DE DONNÉES MULTIDIMENSIONNELLES ET NOMADES

- I. Contexte
- II. **Modélisation des entrepôts de données (DW)**
 - 1. Entrepôt et Magasin de données : définition
 - 2. Le Modèle Dimensionnel
 - 3. Le processus de modélisation
 - 4. **Autres caractéristiques du M.D. : Type, évolution**
- III. Conception physique des entrepôts de données
- IV. Alimentation des entrepôts de données
- V. Accès aux données de l'entrepôt
- VI. Perspectives et évolution

ENTREPÔTS DE DONNÉES

Type de modèle dimensionnel

Définition

La modélisation dimensionnelle consiste à mesurer un sujet d'analyse appelé 'Fait' par un ensemble d'indicateurs. Chaque indicateur est représenté selon des 'Dimensions'.

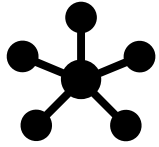
2 types de modèle dimensionnel :

- Modèle en étoile :

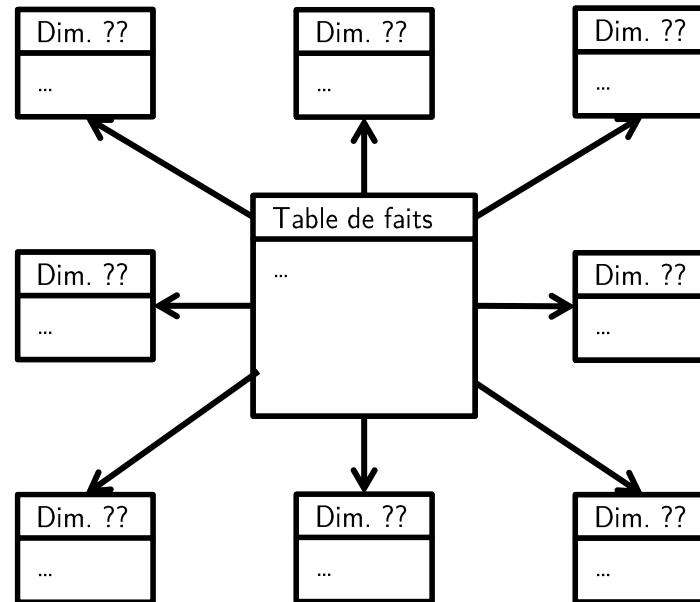
chaque dimension se réduit à une table et chaque ligne de la table de fait est reliée à chaque dimension

- Modèle en flocon :

une dimension est représentée selon une hiérarchie organisant les propriétés de la dimension



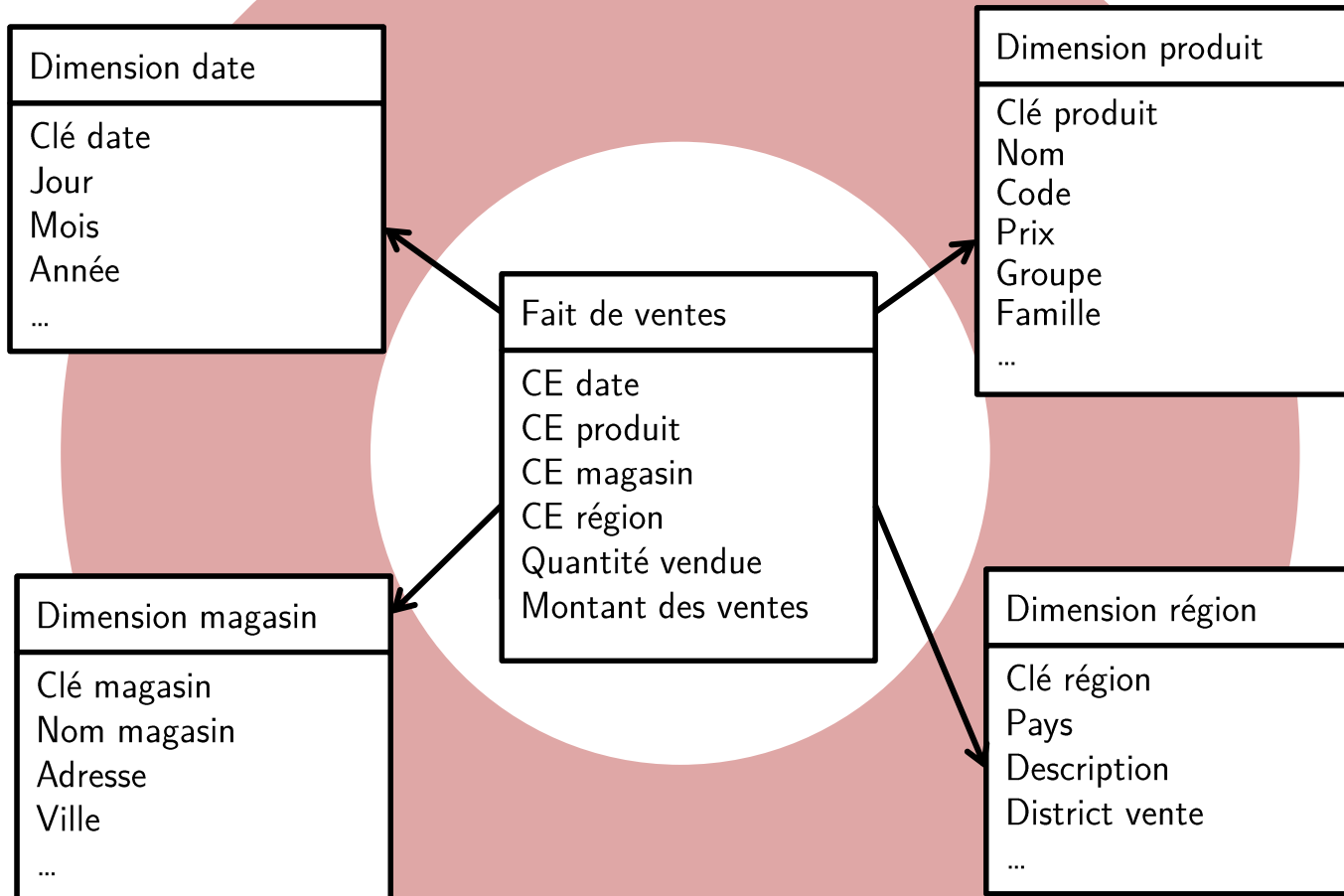
Modèle en étoile



Modèle en étoile

Définition

Un modèle en étoile est une structure dimensionnelle qui représente une table de faits entourée par un seul cercle de dimensions.



Modèle en étoile

Avantages

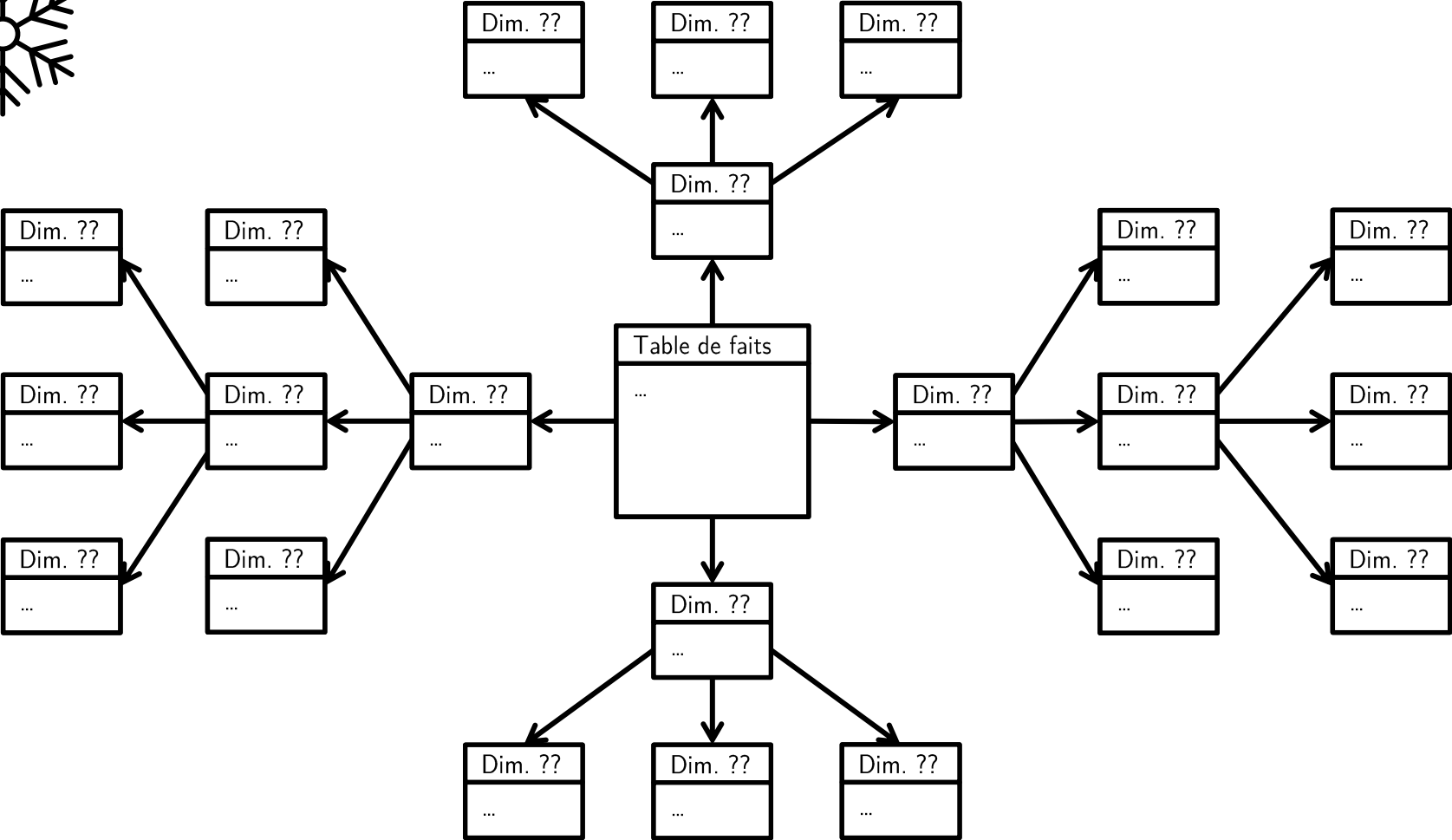
- Facilité de navigation
- Nombre de jointures limité

Inconvénients

- Redondance dans les dimensions
- Toutes les dimensions ne concernent pas les mesures

- Dans un modèle en étoile, toute dimension à niveaux multiples est aplatie en une seule dimension.
- Le modèle en étoile est conçu pour répondre à des requêtes inhérentes à la structure dimension-fait.

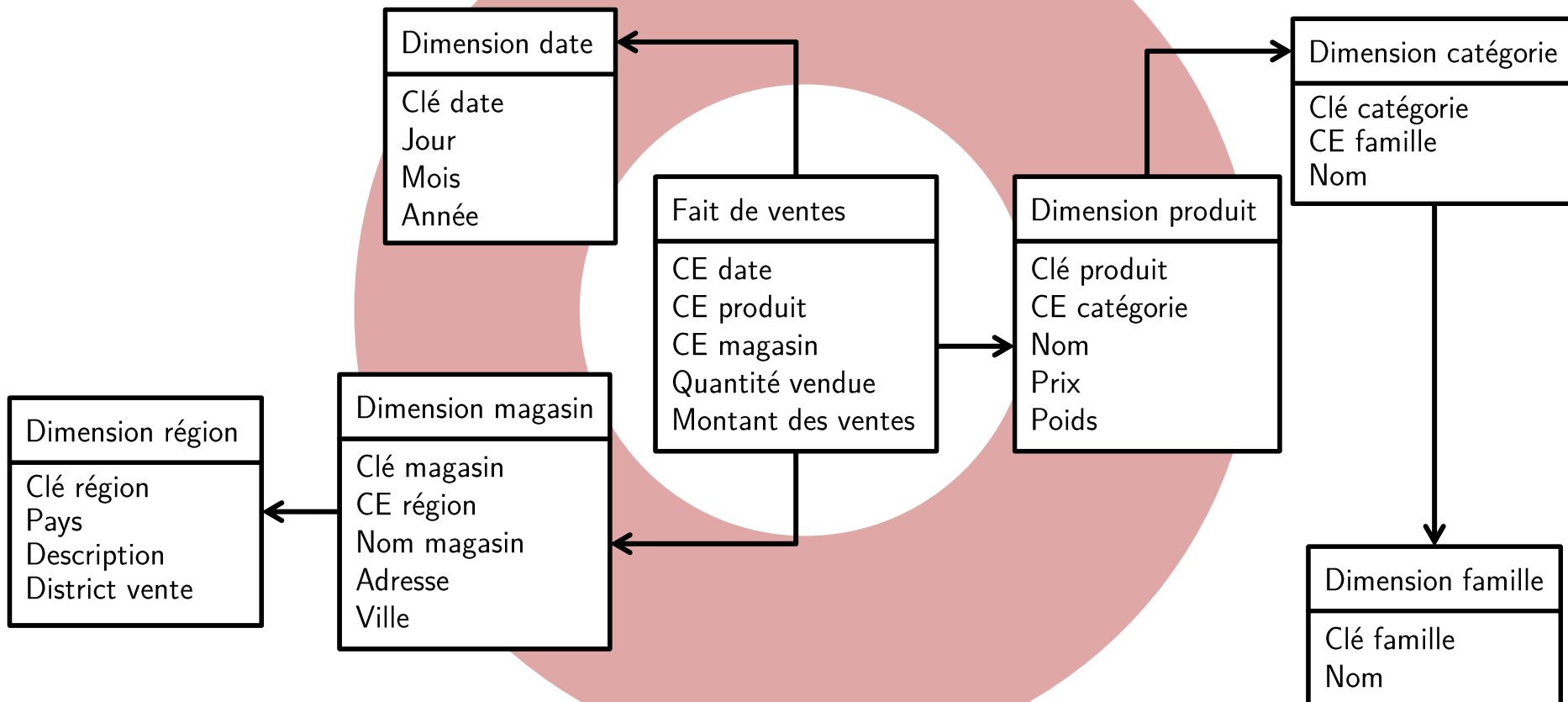
Modèle en flocon



Modèle en flocon

Définition

Un modèle en flocons de neige est aussi une structure dans laquelle une table de faits est entourée par un seul cercle de dimensions. Cependant une dimension peut être à niveaux multiples.



Modèle en flocon

Avantages

- Normalisation des dimensions
- Économie d'espace disque

Inconvénients

- Modèle plus complexe (jointure)
- Requêtes moins performantes

Une dimension contient des membres organisés en hiérarchie :

- Chacun appartient à un niveau hiérarchique (ou niveau de granularité) particulier
- Granularité d'une dimension : nombre de niveaux hiérarchiques

La table de dimension de niveau hiérarchique le plus bas est reliée à la table de fait. On dit qu'elle a la granularité la plus fine

Gestion des clés

Table de faits

- **Clé principale :**
 - Typiquement une clé composée.
Ex: (idDateTransaction, idClient, idProduit)
- **Clés étrangères :**
 - Jamais nulles
 - Utiliser plutôt une valeur spéciale dans la table de dimension;
Ex: une ligne "Aucun spécial" dans la dimension RabaisSpécial si le client n'a eu droit à aucun rabais lors d'une transaction)

La gestion des clés

Table de dimension

- Au moins une clé commune avec le système OLTP :
 - la **clé fonctionnelle** (opérationnelle) utilisée pour la mise à jour de la dimension, elle identifie de manière unique chaque ligne de la dimension
- Une autre clé primaire se substitue à cette clé
 - La **clé technique** dite « clé de substitution » (**surrogate key**)
 - Générée par une séquence du moteur de la base de données en général.
 - Cependant, elle peut parfois être calculée.

Ex : dans une dimension Calendrier, la clé fonctionnelle est la date, la clé technique est la valeur entière de la date

 - En général inconnue des utilisateurs

Avantages :

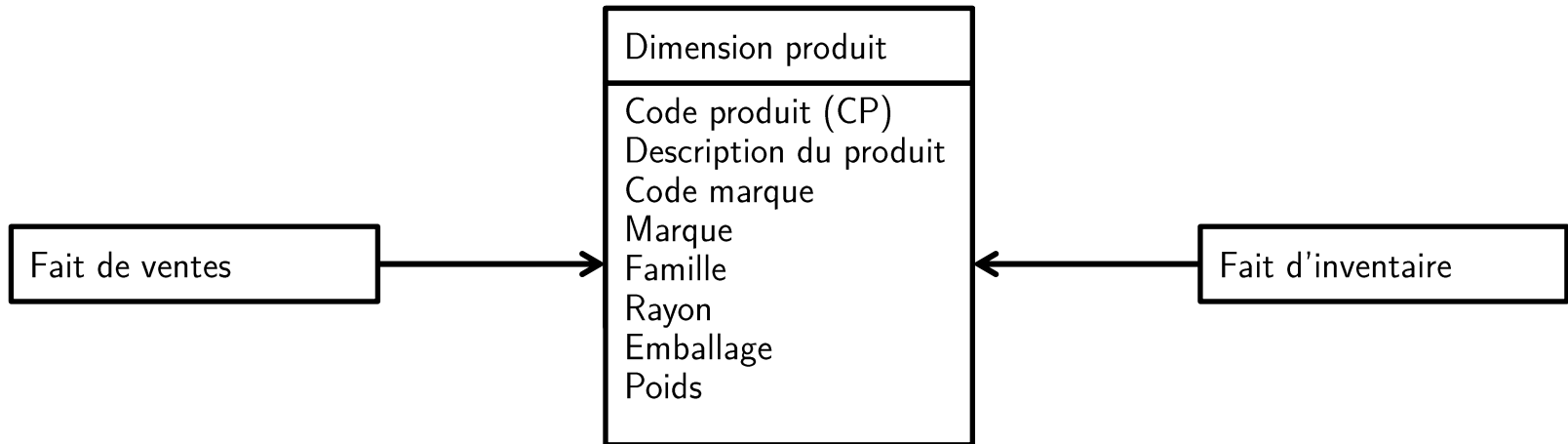
- ✓ **Performance** : accès par index et jointures accélérés (clé entière, de petite taille);
- ✓ **Robustesse** : ne change jamais contrairement à une clé naturelle;
- ✓ **Cohérence** : permet de faire simplement référence à une entrée de la table depuis une autre table (celle de faits par exemple).

Dimensions conformes

- Également appelées « master dimensions » ou « common reference dimensions »;
- Dimensions potentiellement partagées par des tables de faits modélisant des processus d'affaires différents;
- Avantages:
 - ✓ **Cohérence** : les différentes tables de faits sont filtrées et étiquetées de manière de cohérente;
 - ✓ **Intégration** : permet de faire simplement référence à une entrée de la table depuis une autre table (celle de faits par exemple).
 - ✓ **Productivité** : favorise l'extension de l'entrepôt d'une itération de développement à l'autre.

Exemple : dimension temporelle

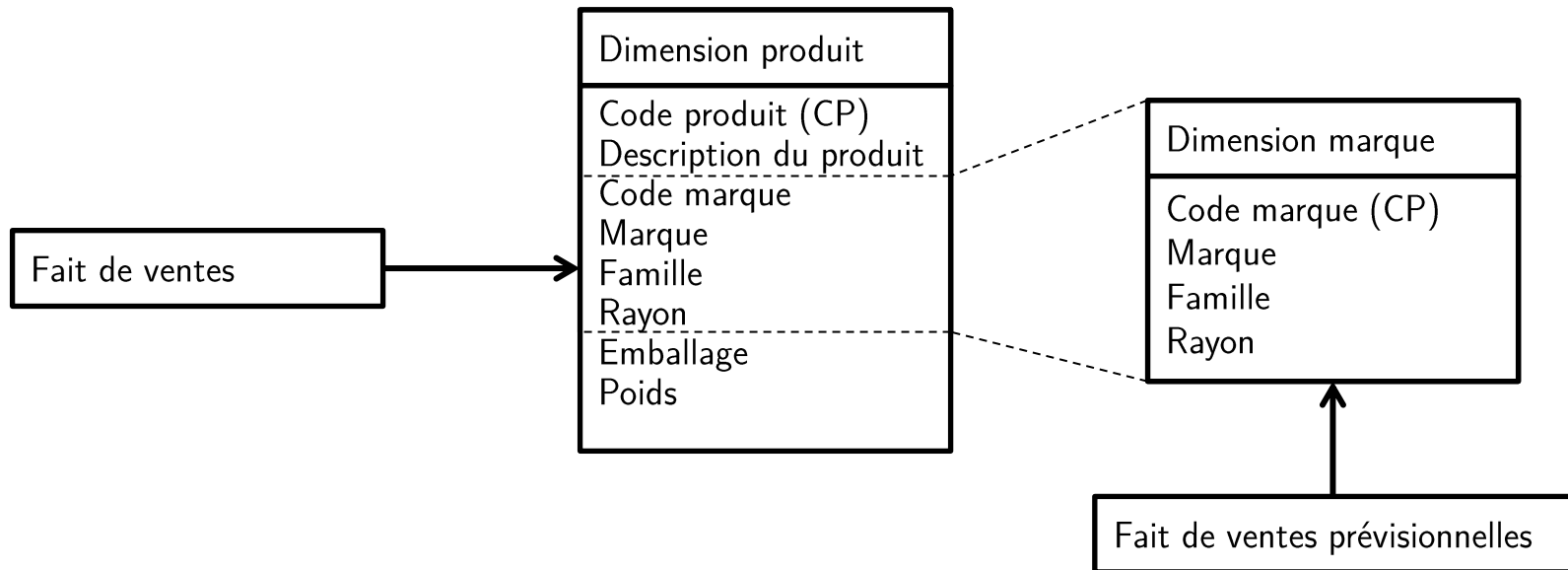
Dimensions conformes



Dimensions conformes

- Des dimensions peuvent être conformes si les attributs d'une table de dimension sont un sous-ensemble des attributs d'une autre table;

Ex: dimensions représentant des niveaux de granularité différents



Dimensions dégénérées

- Clés de la table de faits n'ayant aucun attribut autre qu'elles-mêmes;
- Correspondent souvent à des identifiants dans les systèmes sources;
Ex : no de commande, no de billet, etc.
- Il faut toujours laisser ces colonnes dans la table de faits:
 - Elles sont la colle qui « tient ensemble » les items d'une ligne de la table de faits;
 - Permettent de répondre à des questions plus générales comme « quel est le nombre moyen de lignes correspondant à une même commande ? »
 - Permettent également de retracer la provenance d'une ligne à une source de données.

Fait de ventes
Date (CE)
Produit (CE)
Magasin (CE)
Quantité vendue
Montant des ventes
No de transaction (DD)

Variation des dimensions

- Les attributs descriptifs des tables de dimension sont choisis pour être 'relativement' stables.
- Néanmoins, ils peuvent parfois évoluer dans le temps.
 - Un client peut se marier, avoir des enfants, changer d'adresse, ...
 - Un produit peut changer de nom ou de rayon,
 - Les employés peuvent être engagé, promus, mutés, ou bien licenciés...

Dimensions à évolution lente

- Slowly Changing Dimensions (SCD);
- Même si elles sont plus statiques que les faits, les dimensions peuvent également changer:
Ex : adresse/status d'un client, catégorie d'un produit, etc.
- Plusieurs stratégies d'historisation:
 - SCD Type 1
 - SCD Type 2
 - SCD Type 3

Dimensions à évolution lente

SCD de type 1 : Écrasement de l'ancienne valeur

Renoncer à suivre les situations passées

Clé produit	Description du produit	Rayon
12345	Intelli-Kids 1.0	logiciel

Après modification dans la table dimension

Clé produit	Description du produit	Rayon
12345	Intelli-Kids 1.0	Jeux éducatifs

Avantages

- Facile à mettre en œuvre

Inconvénients

- Perte de la trace des valeurs antérieures des attributs
- Perte de la cause de l'évolution dans les faits mesurés

Dimensions à évolution lente

SCD de type 2 : ajouter une ligne

Soit la dimension Client :

Clé client	Clé fonctionnelle	Nom	Ville	Tranche d'âge
4655	RG45	Dupond	Toulouse	30-40 ans

... et la table Faits de commandes liée à la table dimension Client :

Clé client	Date	Clé produit	Qté	Montant
4655	10/10/2010	15865	12	14,80
4655	05/01/2011	25240	8	65,45

Lors de sa dernière commande, le client signale qu'il a déménagé à Bordeaux.
Comment mettre à jour la BD ?

Dimensions à évolution lente

SCD de type 2 : ajouter une ligne

Solution : Un nouveau client est créé dans la dimension.

Clé client	Clé fonctionnelle	Nom	Ville	Tranche d'âge	Valide début	Valide fin
4655	RG45	Dupond	Toulouse	30-40 ans	10/10/2010	02/04/2012
9667	RG45	Dupond	Bordeaux	30-40 ans	03/04/2012	31/12/9999

Et la nouvelle commande est enregistrée dans la table de faits.

Clé client	Date	Clé produit	Qté	Montant
4655	10/10/2010	15865	12	14,80
4655	05/01/2011	25240	8	65,45
9667	03/04/2012	58642	6	78,5

« Valide début » et « valide fin » indiquent l'intervalle de temps pendant lequel l'enregistrement est valide.

Dimensions à évolution lente

SCD de type 3 : ancienne et nouvelle valeur

On maintient deux colonnes, avec l'ancienne et la nouvelle valeur.

Clé produit	Description du produit	Ancien Rayon	Nouveau rayon
12345	Intelli-Kids 1.0	Jardinage	logiciel

↓

Clé produit	Description du produit	Ancien Rayon	Nouveau rayon
12345	Intelli-Kids 1.0	logiciel	Jeux éducatifs

Lors d'une modification, on déplace le contenu de la colonne « nouvelle valeur » dans la colonne « ancienne valeur », puis on écrase la valeur de la colonne « nouvelle valeur ».

Dimensions à évolution lente

- On a vu les types suivants:
 - **SCD Type 1:**
 - Écraser l'ancienne valeur avec la nouvelle;
 - À utiliser seulement pour faire des corrections ou lorsque l'ancienne valeur n'est pas significative pour les besoins d'affaires;
 - **SCD Type 2:** L'approche la plus employée
 - Ajouter une ligne dans la table de dimension pour la nouvelle valeur;
 - Permet de faire des analyses historiques;
 - À utiliser lorsque l'ancienne valeur a une signification analytique ou si le changement est une information en soi.
 - **SCD Type 3:** L'approche la moins employée
 - Avoir deux colonnes dans la table de dimension correspondant à l'ancienne et la nouvelle valeur.
 - Profondeur de l'historique est d'un seul changement (on peut ajouter plus de colonnes pour plus de profondeur);
 - À utiliser lorsqu'on veut vouloir comparer les faits avec l'ancienne ou la nouvelle valeur;
- Il existe aussi des schémas hybrides, combinant les Types 2 et 3.