



Indicateur de performance

www.kapei-conseil.com



Clément LEFAURE
Associé / Directeur technique

clefaure@kapei-conseil.com
Mob : +33(0)7 82 02 95 00
20, Bd. Eugène Deruelle 69003 Lyon

Fondements de l'informatique décisionnelle

Session n°2

Conception d'un entrepôt décisionnel: Modélisation

Session N°2: Agenda

Conception d'un entrepôt décisionnel: Modélisation

- 1 – Processus de modélisation
 - Types de modèles de données
 - Etape par étape
- 2 – Au cœur du modèle multi-dimensionnel
 - Hypercubes, « slice » et « dice »
 - Les modèles dimensionnels
- 3 – Concepts avancés
 - Dimensions et tables de faits
 - Architecture « bus », matrices et dimensions conformes

Conception d'un entrepôt décisionnel: Modélisation

1. Processus de modélisation
2. Au cœur du modèle multi-dimensionnel
3. Concepts avancés

Processus de modélisation

Modéliser des données quèsaco? (2/3)

- Est-ce que vous vous promèneriez dans une ville inconnue sans carte ?



- Peut-être...
- Vous pouvez toujours utiliser les panneaux de signalisation des rues...

Processus de modélisation

Modéliser des données quèsaco? (3/3)

- Sans doute...
- Mais comment faire dans les villes dont les habitants parlent une autre langue ?
- On ne peut pas apprendre toutes les langues des pays dans lesquels on veut voyager à chaque fois...
- De la même façon il a fallut s'entendre sur **une façon de représenter graphiquement les bases de données indépendamment de leurs systèmes de gestion et des langages de programmation.**
- C'est l'intérêt d'un modèle de données



Processus de modélisation

Types de modèles de données

- La modélisation d'applications décisionnelles s'inscrit dans la démarche classique de modélisation:

1. Modèle Conceptuel de données (MCD)

1. Utilisateurs "métier", Analystes



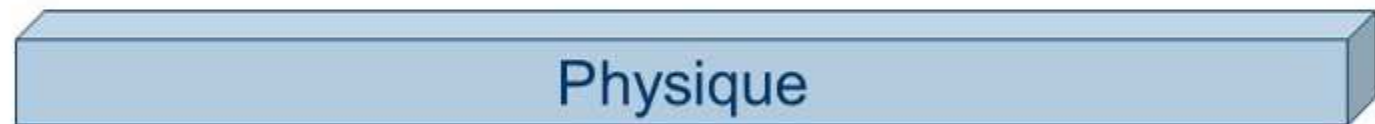
2. Modèle Logique de Données (MLD)

1. Analystes « métiers », Architectes



1. Modèle Physique de Données (MPD)

1. Architectes, DBAs, Développeurs



Processus de modélisation dimensionnelle (1/7)

- Tout part d'une **étude de besoin** auprès des utilisateurs (key-users) de la futures application
- Ce recueil de desiderata est souvent effectué au moyen d'**atelier de travail** (workshops)
- Le résultat se présente souvent sous une forme de **liste de questions** que les utilisateurs se posent au cours de leurs analyses.
- Par exemple:
 - Quel est le nombre de bien vendus par compagnie ?
 - Quel est le revenu moyen pour chaque produit ?
 - Quel est le dernier niveau de stock pour un mois donné ?
 - Quel le coût minimal pour un type de produit spécifique dans le dernier mois, présenté par région ?

Processus de modélisation dimensionnelle (2/7)

- On essaie ensuite d'analyser cette liste de questions
- L'objectif est d'identifier:
 - *Ce qui doit être mesuré* (indicateurs)
 - **Les AXES d'analyses** (dimensions)
 - Les niveaux de regroupement (agrégats)
- Par exemple:
 - Quel est le nombre de *bien vendus* par **COMPAGNIE** ?
 - Quel est le *revenu moyen* pour chaque **PRODUIT** ?
 - Quel est le dernier *niveau de stock* pour un **MOIS** donné ?
 - Quel le *coût minimal* pour un **TYPE DE PRODUIT** spécifique dans le dernier **MOIS**, présenté par **REGION** ?

Processus de modélisation dimensionnelle (3/7)

- Qu'est-ce qui doit être mesuré ?
 - Les indicateurs clés de performance (Key Performance Indicateur)
 - Ce sont des valeurs numériques
- Par exemple:

Chiffre d'Affaire
Volume
Marge
Résultat

Processus de modélisation dimensionnelle (4/7)

A mesurer par...

➤ Rechercher les circonstances

Quoi?

Où?

Qui?

Chiffre d'Affaire
Volume
Marge
Résultat

Comment?

Quand?

Processus de modélisation dimensionnelle (5/7)

A mesurer par...

Quoi?
Vehicule

Où?
Marché

Qui?
Concessionnaire

Chiffre d'Affaire
Volume
Marge
Résultat

Niveau de suivi
Comment?

Date
Quand?

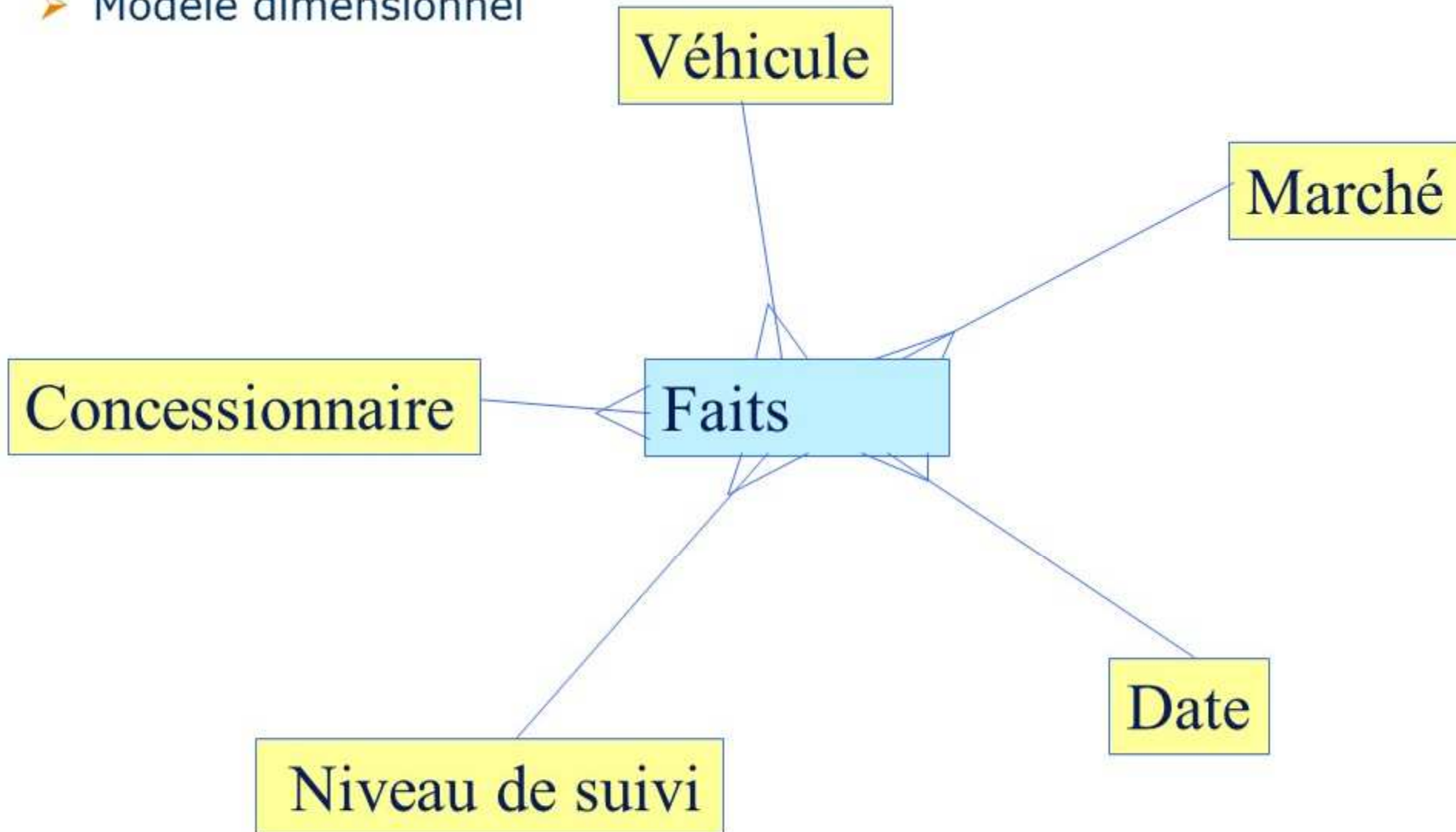
Processus de modélisation dimensionnelle (6/7)

- Une étoile apparaît....



Processus de modélisation dimensionnelle (7/7)

➤ Modèle dimensionnel



Conception d'un entrepôt décisionnel: Modélisation

1. Processus de modélisation
2. Au cœur du modèle multi-dimensionnel
3. Concepts avancés

Au cœur du modèle multi-dimensionnel

Rappels sur les dimensions

➤ Les dimensions :

- Contiennent des données **qualitatives** (champs texte ou date)
- Correspondent à un axe d'analyse métier
- Sont qualifiées par leur attributs
- Représentent en général un volume de données faible

➤ Exemples:

- *Produit:*
 - *couleur du produit, packaging...*
- *Magasin:*
 - *Géographie, Type de magasin, Type de tarif...*
- *Client:*
 - *Genre du client, CSP...*
- *RH:*
 - *Organisation de l'entreprise, type de poste...*
- *Temps:*
 - *Jour, mois, trimestre, année...*

Au cœur du modèle multi-dimensionnel

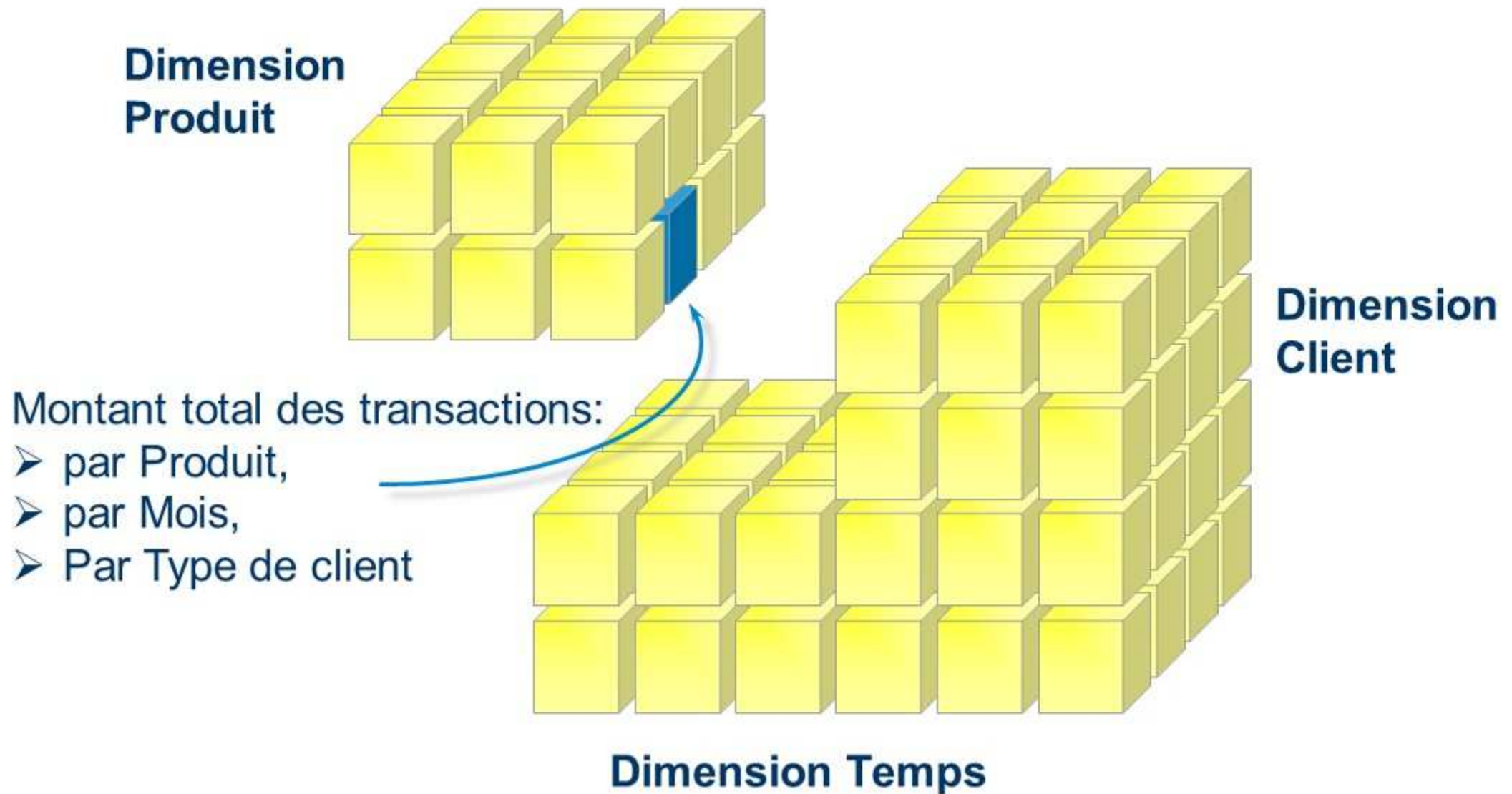
Rappels sur les tables de faits

➤ Les tables de faits :

- Contiennent des données **quantitatives** (indicateurs numériques)
- Contiennent des informations détaillées
- Représentent la volumétrie la plus importante de l'entrepôt
- Sont historisées
- Exemples:
 - *Quantité en stock dans le magasin X pour le produit Y à la date Z*
 - *Montant de vente pour le commercial X de l'agence Y sur le mois M*
 - *Fréquence de visite sur le site X du client Z*
 - *Salaire, augmentation pour le salarié X du service Y par le responsable Z*

Au cœur du modèle multi-dimensionnel

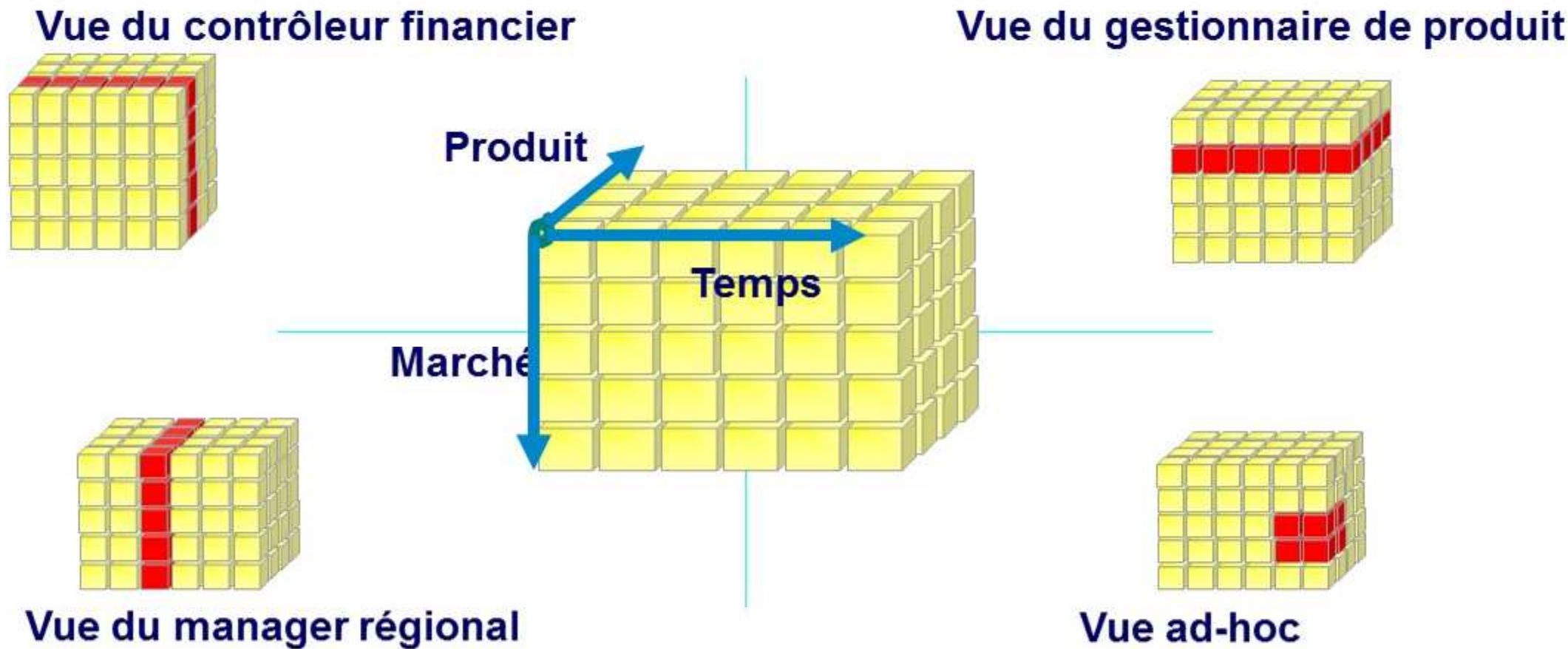
Fonctionnement d'un hypercube: axes d'analyse



Au cœur du modèle multi-dimensionnel

Fonctionnement d'un hypercube: « slicing » & « dicing »

- Les outils multi-dimensionnel permettent aux utilisateurs de voir les données sous tous les angles.
- La possibilité de sélectionner différents angles d'analyse est appelée "slice" et "dice" par les anglo-saxons



Au cœur du modèle multi-dimensionnel

Modèle en « étoile »

- Un schéma en étoile est une structure dimensionnelle qui représente une seule table de faits reliée directement aux dimensions
- Avantages:
 - Facilite la navigation lors de la restitution
 - Performances:
 - nombre de jointures limité
- Inconvénients:
 - Redondances dans les dimensions
 - Alimentation complexe
 - Le métier à modéliser n'est pas
 - toujours aussi simple



Au cœur du modèle multi-dimensionnel

Modèle en « flocon »

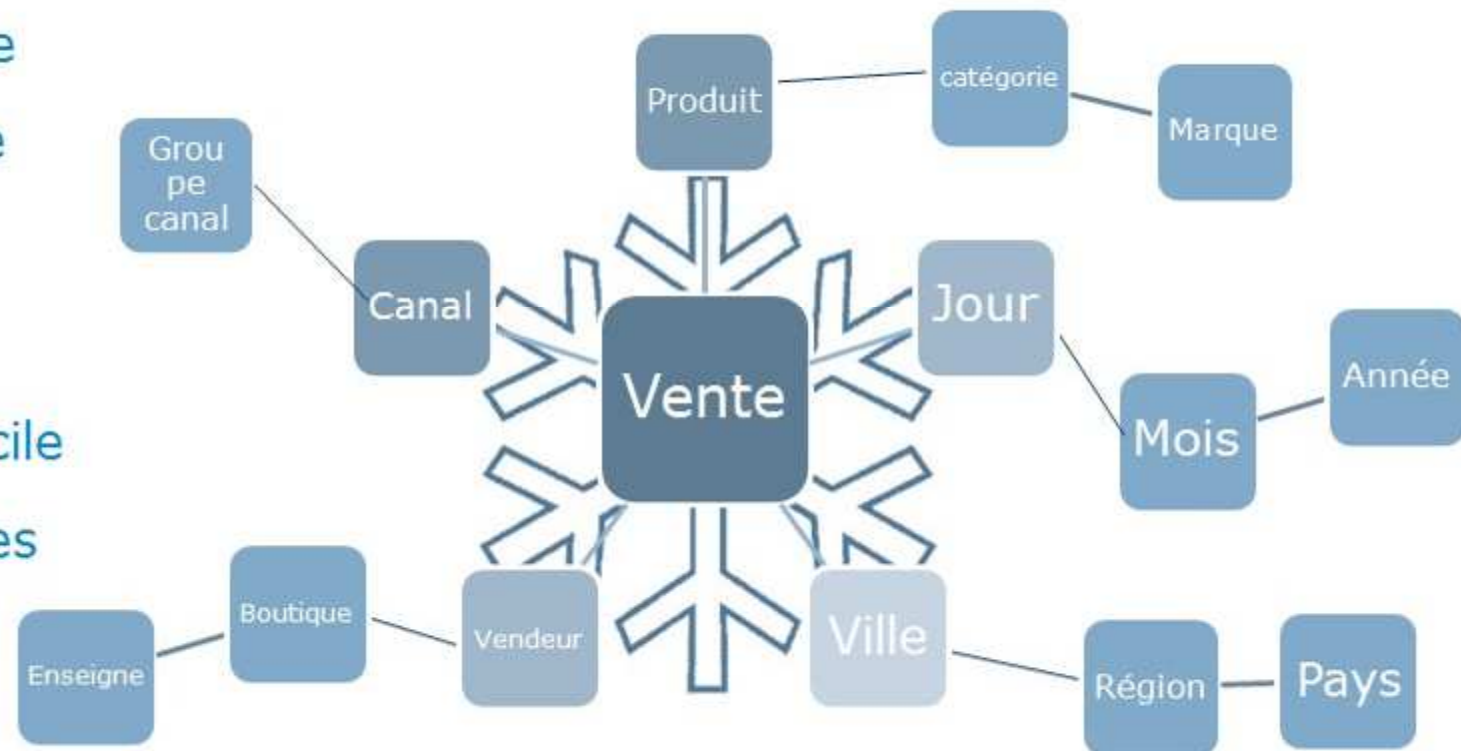
- Un schéma en flocon est en schéma en étoile mais avec certaines dimensions sur plusieurs niveaux (plus normalisées)

- Avantages:

- Réduction du volume
- Maintenance + facile

- Inconvénients:

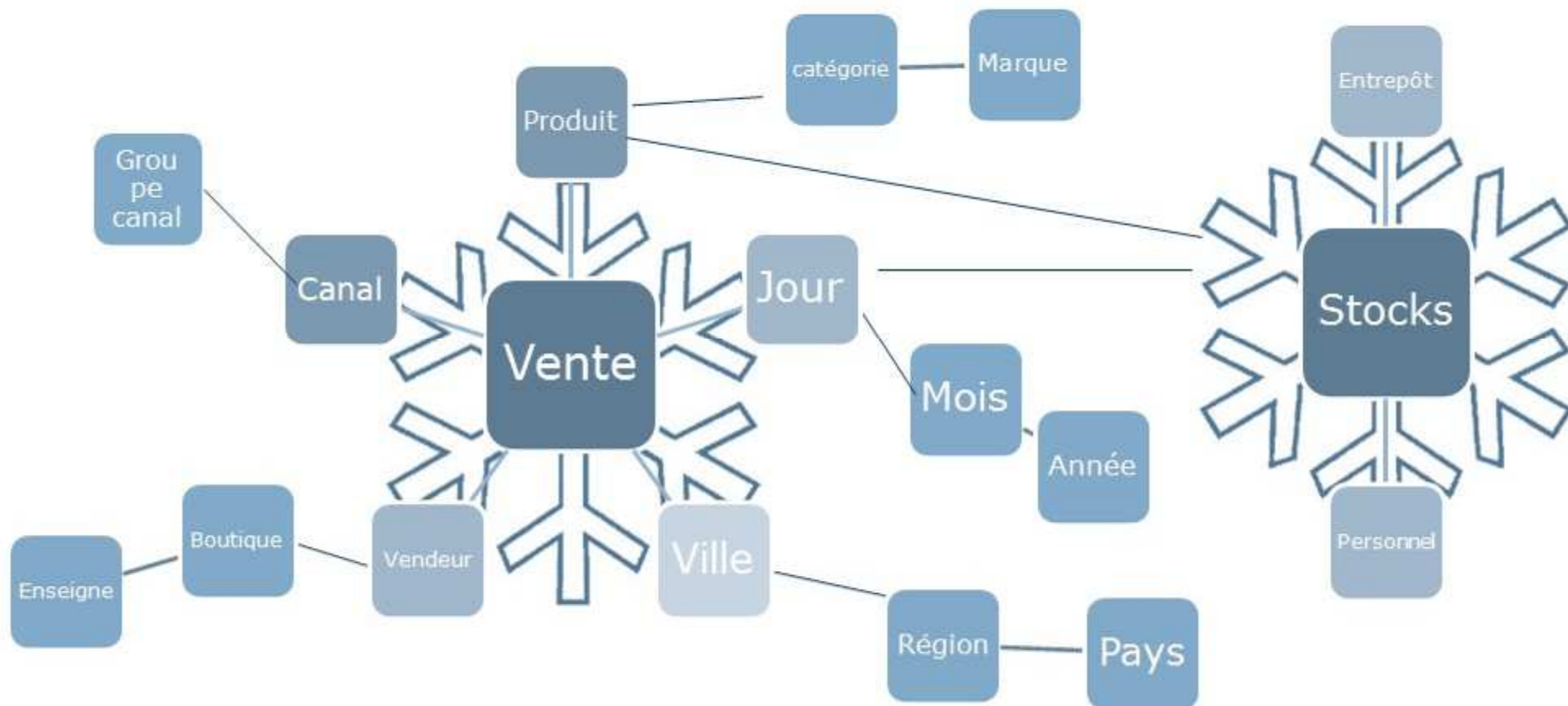
- Navigation plus difficile
- Nombreuses jointures



Au cœur du modèle multi-dimensionnel

Modèle en « constellation »

- Un schéma en constellation est en schéma en flocon mais avec des dimensions communes sur plusieurs tables de faits



Conception d'un entrepôt décisionnel: Modélisation

1. Processus de modélisation
2. Au cœur du modèle multi-dimensionnel
- 3. Concepts avancés**

Au cœur du modèle multi-dimensionnel

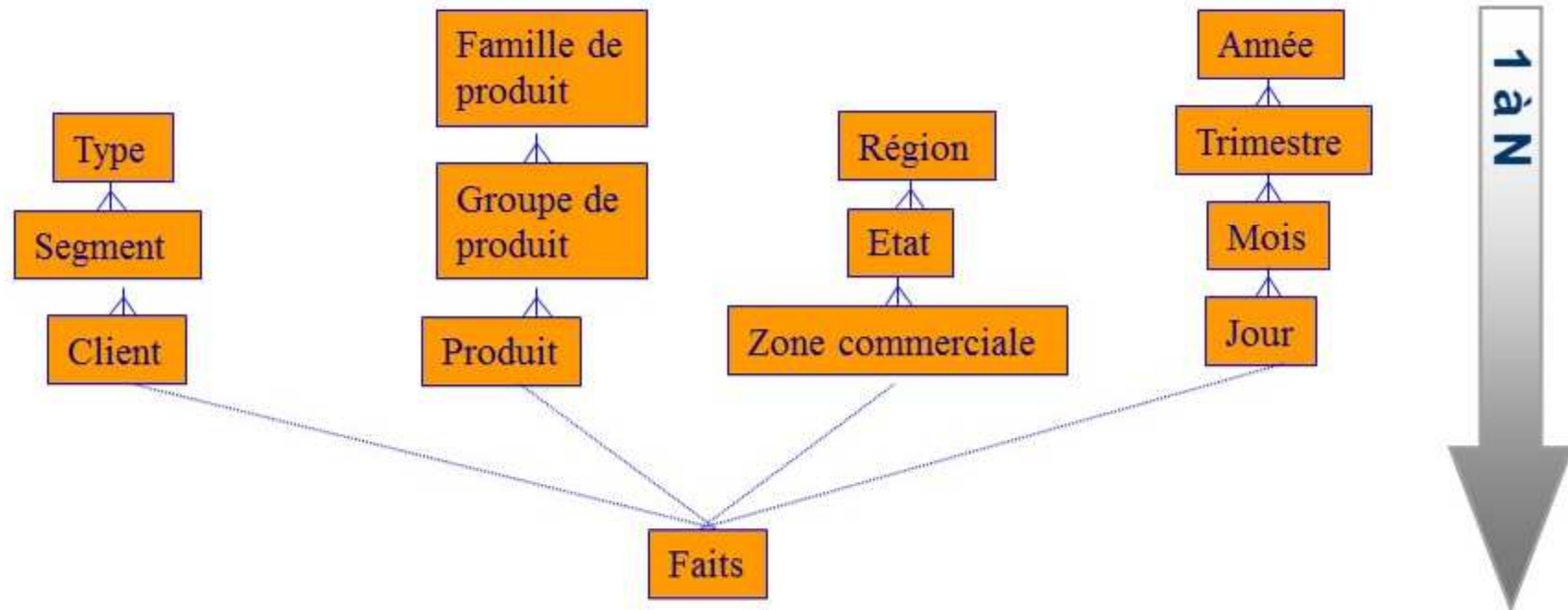
Dimensions: concepts avancés (1/2)

- Les attributs d'une dimension sont souvent organisés en **hiérarchies**:
 - Année → Trimestre → Mois → Jour
 - Monde → Zone commerciale → Pays → Ville
 - Agence → Service → Employé
 - Catégorie → Gamme → Famille de produit → Produit
- Les dimensions peuvent évoluer dans le temps (**Slowly Changing Dimensions** ou SCD)
 - **SCD type 0**: Non gestion de la modification
 - **SCD type 1**: Ecrasement de l'ancienne valeur
 - **SCD type 2**: Historisation de l'ancienne valeur (avec un booléen ou une fourchette de dates)
 - **SCD type 3**: Utilisation d'une colonne supplémentaire pour la nouvelle version (limité)
- « **Junk** » **dimension** (dimensions fourre tout):
 - Pour éviter de créer plein de dimensions avec un seul attribut, on les regroupe dans une même dimensions, même s'ils n'ont aucun rapport entre eux
- Gestion de membre inconnu:
 - **Valeur « Unknown »**: une ligne « joker » est créé dans la dimension
 - **Membre inféré**: La dimension est alimentée automatiquement quand une nouvelle valeur est détectée lors de l'alimentation de la table de fait

Au cœur du modèle multi-dimensionnel

Dimensions: concepts avancés (2/2)

- Chaque dimension peut contenir des **hiérarchies** sur plusieurs niveaux:
 - EX: Un « Client » appartient à un « Segment », et les « Segments » sont organisés en « Type »
- Ces hiérarchies correspondant aux **niveaux de granularité** (ou « niveau d'agrégation ») disponible pour les faits



Au cœur du modèle multi-dimensionnel

Tables de faits: concepts avancés

- Constitution d'une table de fait:
 - Une collection de **liens vers des dimensions**. EX: Magasin de Lyon, Produit A, Mois X
 - Une collection de mesures (indicateurs). EX: CA, quantité en stock...
- Notion de **granularité** sur chaque axe:
 - Le niveau de granularité **sur chaque axe** est très important
 - EX: l'état des stocks peut être représenté dans plusieurs tables de faits:
 - Le stock fin de mois par famille de produits au niveau de chaque magasin
 - Le stock de fin d'année par produit dans chaque pays
 - → Ces tables auront les mêmes indicateurs mais des granularités différentes
- Types de tables de faits courants :
 - Instantané périodique
 - Instantané de transaction
 - Instantané récapitulatif

Au cœur du modèle multi-dimensionnel

Clés « métier » et clés « de substitution »

➤ Deux types de clés à utiliser dans les modèles dimensionnels

	Business key	Surrogate key
En français	Clé « métier »	Clé de substitution
Identifiant	Dans le système transactionnel source	Dans le système décisionnel
Destinataire	Métier: utilisateur final (elle est parlante pour lui)	Technique: technicien de maintenance
Utilité		Permet de tracer les différentes version d'un même objet métier
Création	Par copie dans le système source	Par incrémentation automatique
Convention de nommage	« Code », « NO », « NB »	« Id », « Clé technique », « TK »

Au cœur du modèle multi-dimensionnel

Matrice de bus

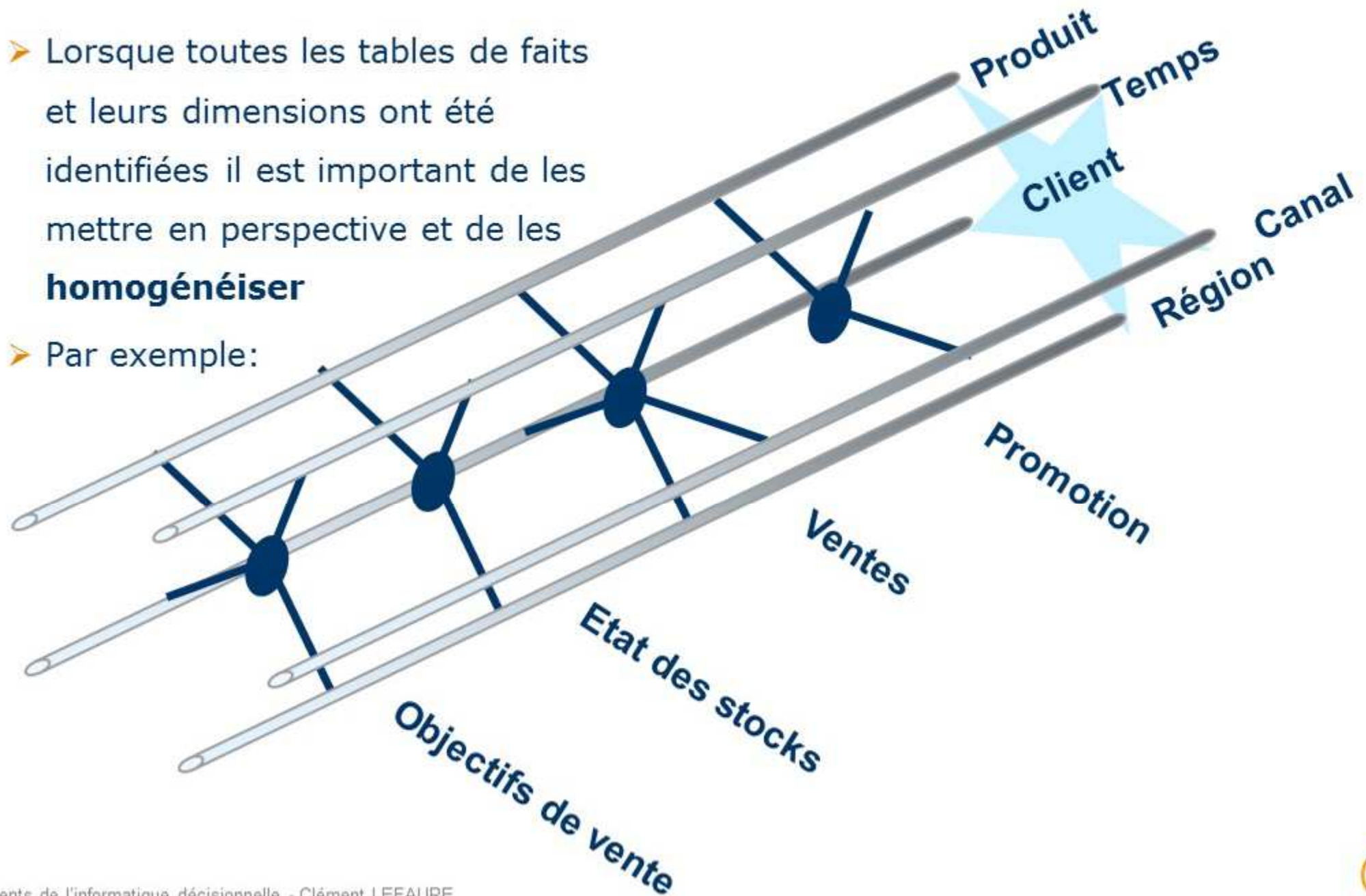
- **La « matrice de bus » (ou « matrice des besoins »):**
 - C'est un des documents permettant de spécifier un entrepôt de données multidimensionnel
 - Elle représente le schéma en étoile (ou en flocon) sous forme de tableau
 - Il met en relation les dimensions et les mesures des tables de faits
 - **En précisant la granularité** du lien
- **Exemple:**

	Dimensions				
		Temps	Produit	Géographie	Commercial
Tables de faits	Stocks	Mois	Produit	Région	
	Ventes	Jour	Produit	Magasin	Agent
	Valorisation	Jour	Produit		
	Budget	Mois	Famille	Magasin	Agent

Au cœur du modèle multi-dimensionnel

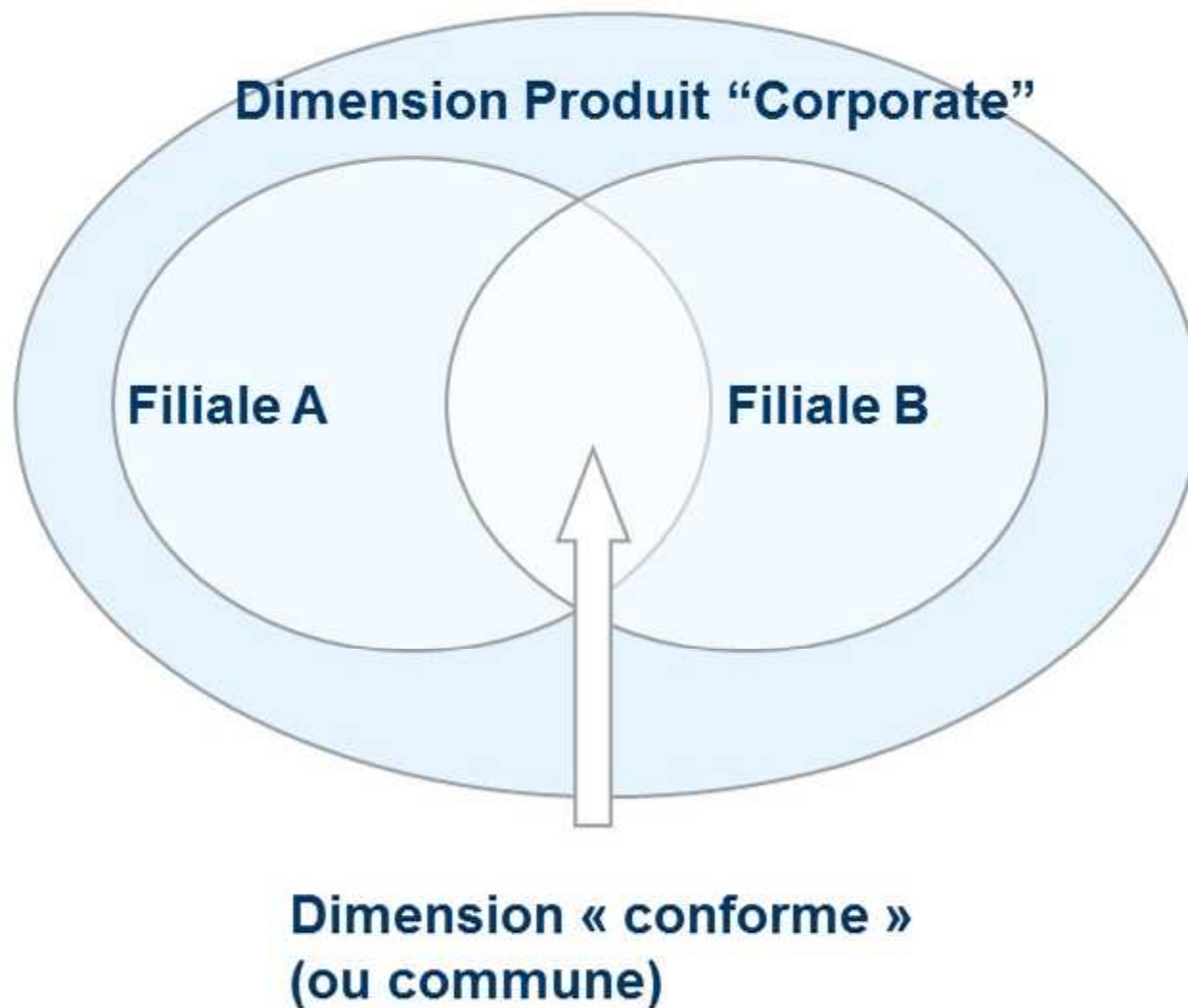
Architecture en bus

- Lorsque toutes les tables de faits et leurs dimensions ont été identifiées il est important de les mettre en perspective et de les **homogénéiser**
- Par exemple:



Au cœur du modèle multi-dimensionnel

Conformité des dimensions



➤ Création d'une dimension commune:

- Imaginons qu'une compagnie comporte deux filiales avec des dimensions différentes
- Il est généralement intéressant d'agréger les données sur une dimension commune
- Dans ce cas on liste des attributs communs aux deux filiales. L'agrégation est alors possible sur cette « **dimension commune** »

- « ***Entrepôts de données: Guide pratique de modélisation dimensionnelle*** » 2eme Ed. par Ralph Kimball & Margy Ross, 2011
- « ***Slowly changing dimension*** », Wikipédia
 - http://en.wikipedia.org/wiki/Slowly_changing_dimension

Exercice:

- Transformer ce schéma de base de donnée relationnel en base multi-dimensionnelle

