

# **Transformation de modèles multidimensionnels : quid de l'ingénierie des connaissances ?**

Lucile Sautot

en collaboration avec S. Bimonte et L. Journaux

Novembre 2017

UMR TETIS, AgroParisTech

UR TSCF, IRSTEA

UMR LE2I, AgroSupDijon

# Plan

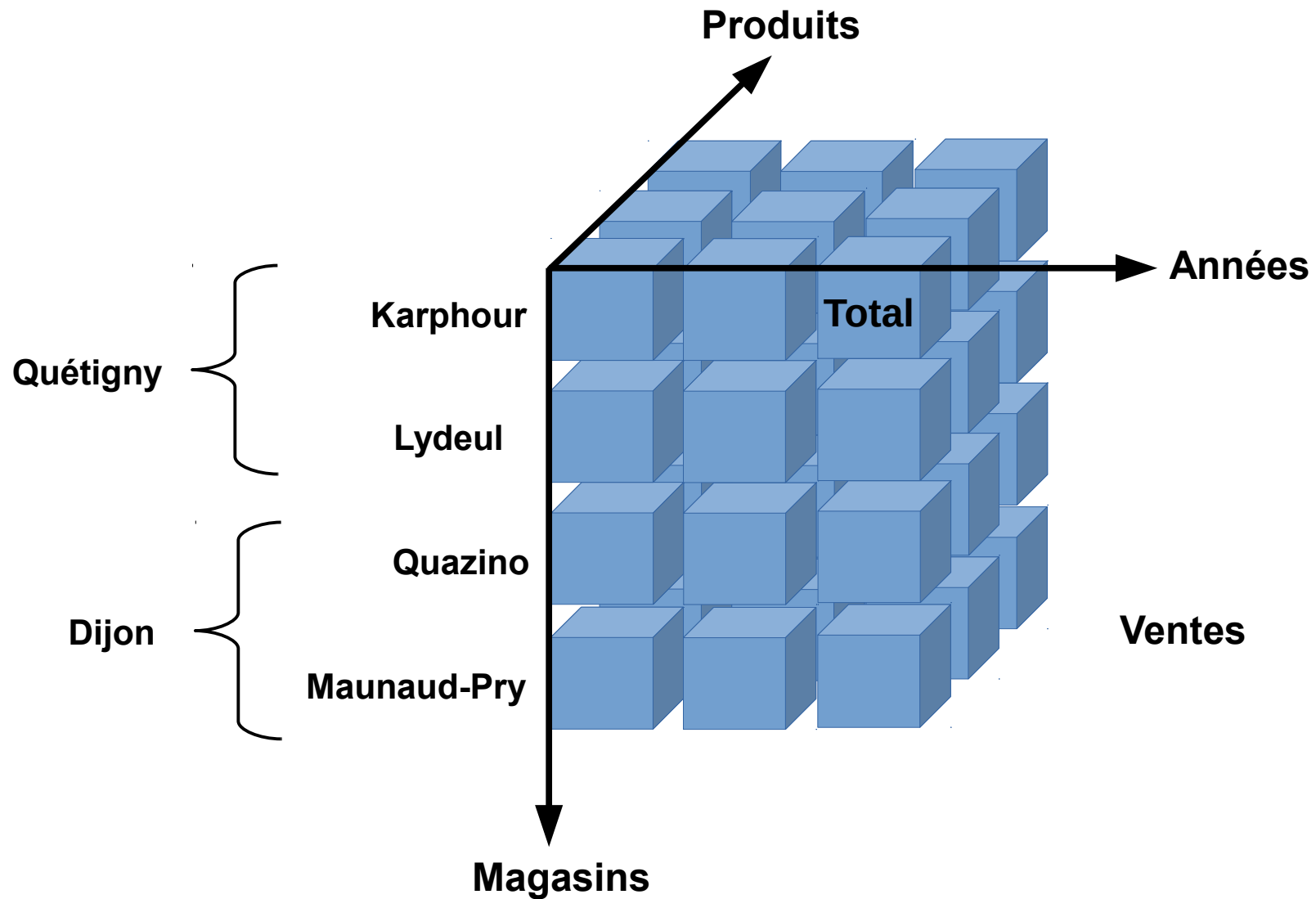
- Introduction
- Méthodologie proposée
  - Vue générale de la méthodologie
    - Calcul de nouvelles hiérarchies
    - Transformation du schéma multidimensionnel
    - Réduction du nombre de hiérarchies
  - Expérimentation
- La méthodologie proposée en question
- Conclusion et perspectives

# Plan

- Introduction
- Méthodologie proposée
  - Vue générale de la méthodologie
    - Calcul de nouvelles hiérarchies
    - Transformation du schéma multidimensionnel
    - Réduction du nombre de hiérarchies
  - Expérimentation
- La méthodologie proposée en question
- Conclusion et perspectives

# OLAP

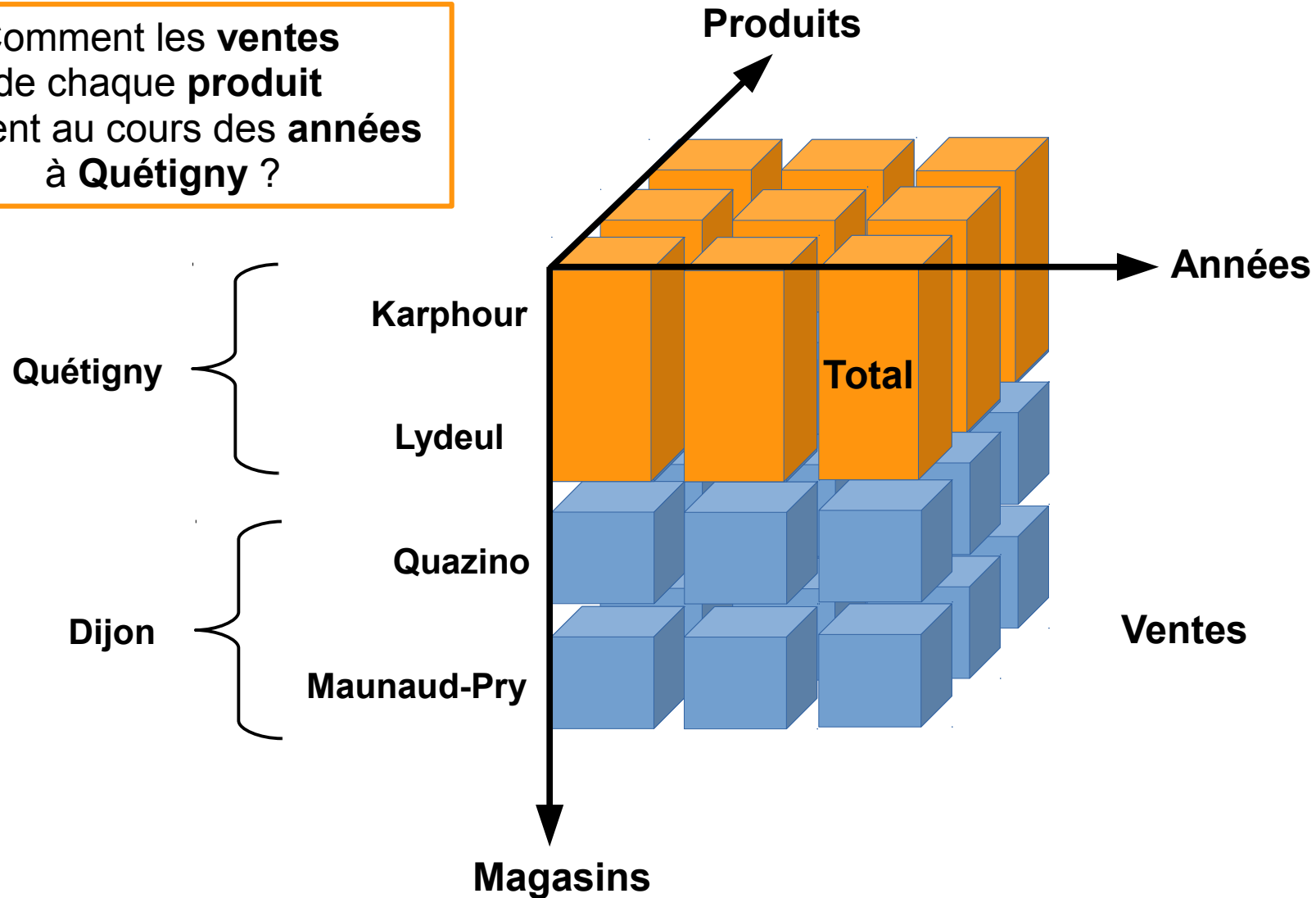
## *Modèle multidimensionnel*



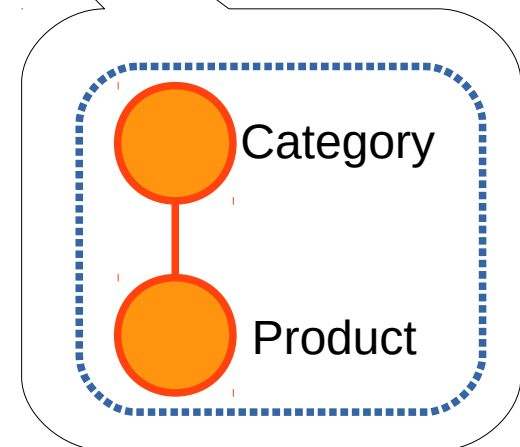
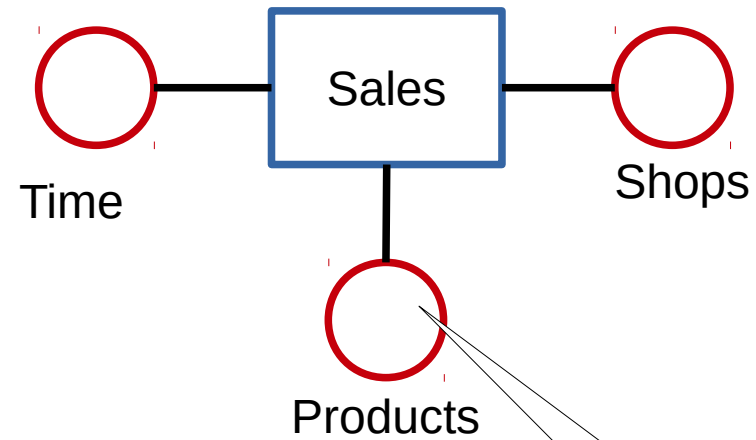
# OLAP

## *Modèle multidimensionnel*

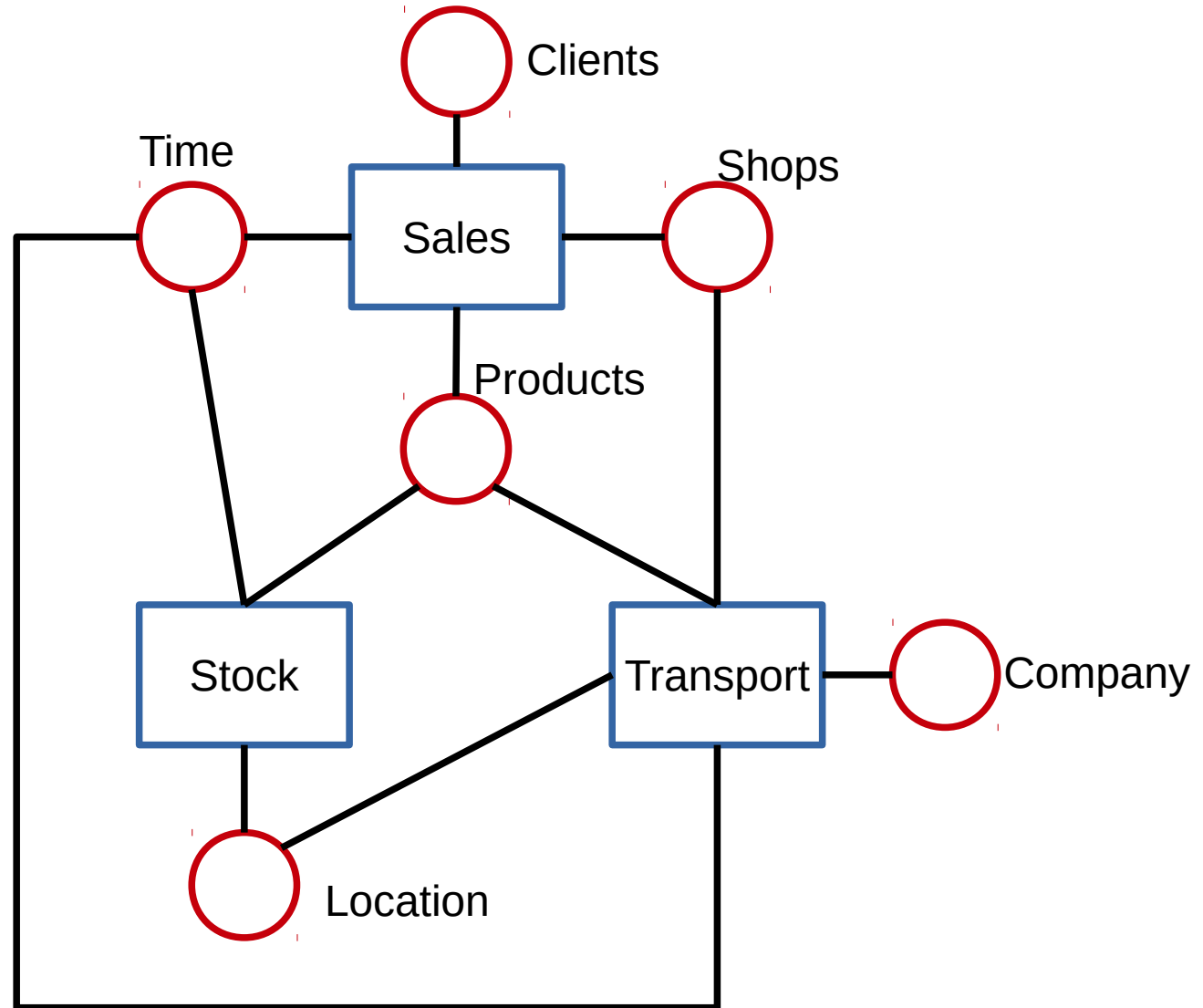
Comment les **ventes** de chaque **produit** évoluent au cours des **années** à **Quétigny** ?



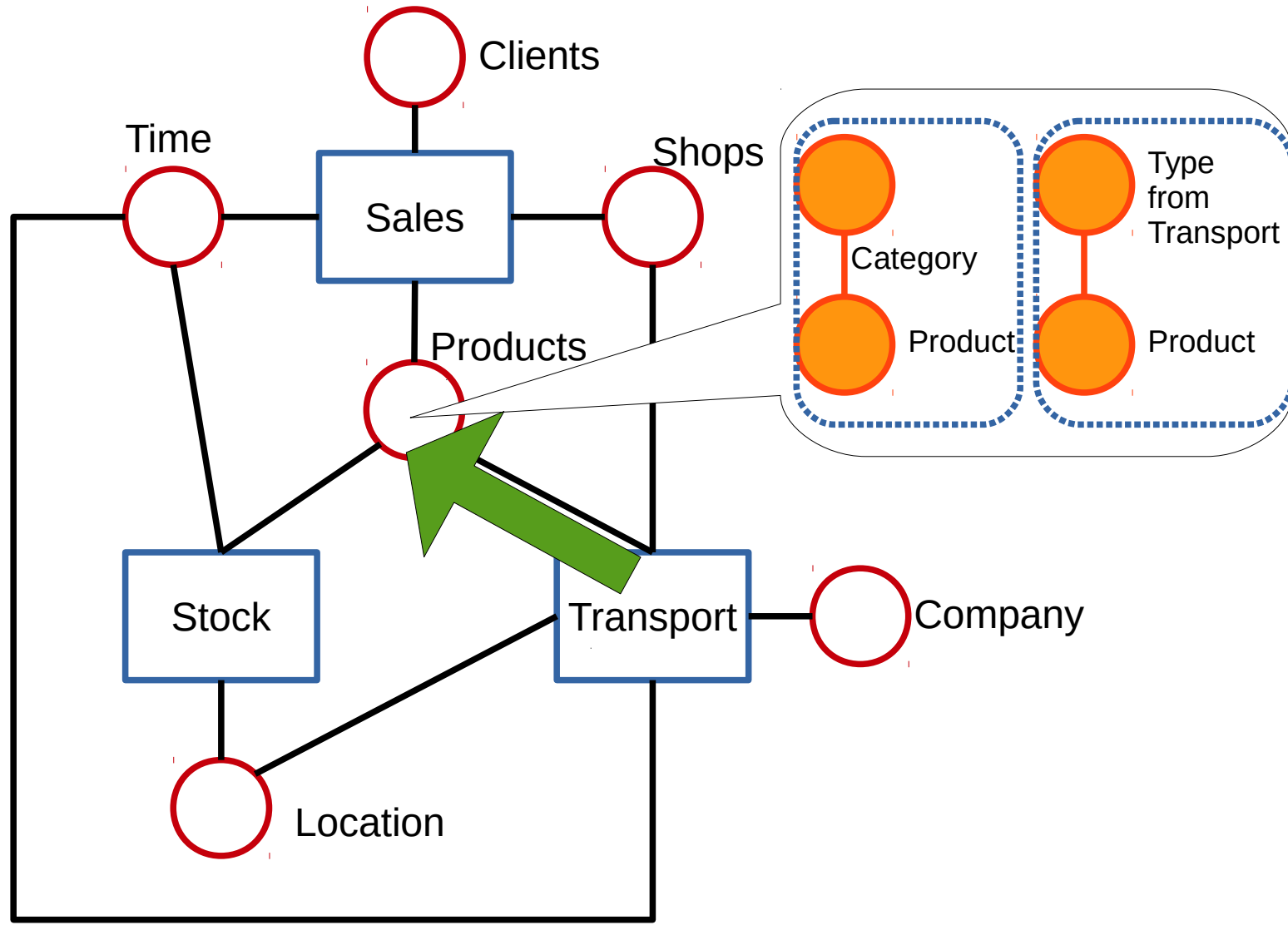
# Schéma en étoile



# Schéma en constellation



# Objectif





# État de l'art

**Favre et al., 2006** : Définition de règles par les utilisateurs pour construire de nouvelles hiérarchies

**Leonhardi et al., 2010** : Utilisation du clustering sur les faits pour définir de nouvelles dimensions

**Ben Messaoud et al., 2004** : Agrégation de faits avec la CAH\*

**Bentayeb, 2008** : Création de niveaux dans une hiérarchie avec le k-means

**Ceci et al., 2011** : Utilisation de la CAH pour l'intégration de données continues dans une dimension

# État de l'art

**Favre et al., 2006** : Définition de règles par les utilisateurs pour construire de nouvelles hiérarchies

**Leonhardi et al., 2010** : Utilisation du clustering sur les faits pour définir de nouvelles dimensions

**Ben Messaoud et al., 2004** : Agrégation de faits avec la CAH

**Bentayeb, 2008** : Création de niveaux dans une hiérarchie avec le k-means

**Ceci et al., 2011** : Utilisation de la CAH pour l'intégration de données continues dans une dimension

**Construction de nouvelles hiérarchies par les utilisateurs**

# État de l'art

Favre et al., 2006 : Définition de règles par les utilisateurs pour construire de nouvelles hiérarchies

**Leonhardi et al., 2010** : Utilisation du clustering sur les faits pour définir de nouvelles dimensions

**Ben Messaoud et al., 2004** : Agrégation de faits avec la CAH

**Bentayeb, 2008** : Création de niveaux dans une hiérarchie avec le k-means

**Ceci et al., 2011** : Utilisation de la CAH pour l'intégration de données continues dans une dimension

**Utilisation du clustering**

# État de l'art

**Favre et al., 2006** : Définition de règles par les utilisateurs pour construire de nouvelles hiérarchies

**Leonhardi et al., 2010** : Utilisation du clustering sur les faits pour définir de nouvelles dimensions

**Ben Messaoud et al., 2004** : Agrégation de faits avec la CAH

**Bentayeb, 2008** : Création de niveaux dans une hiérarchie avec le k-means

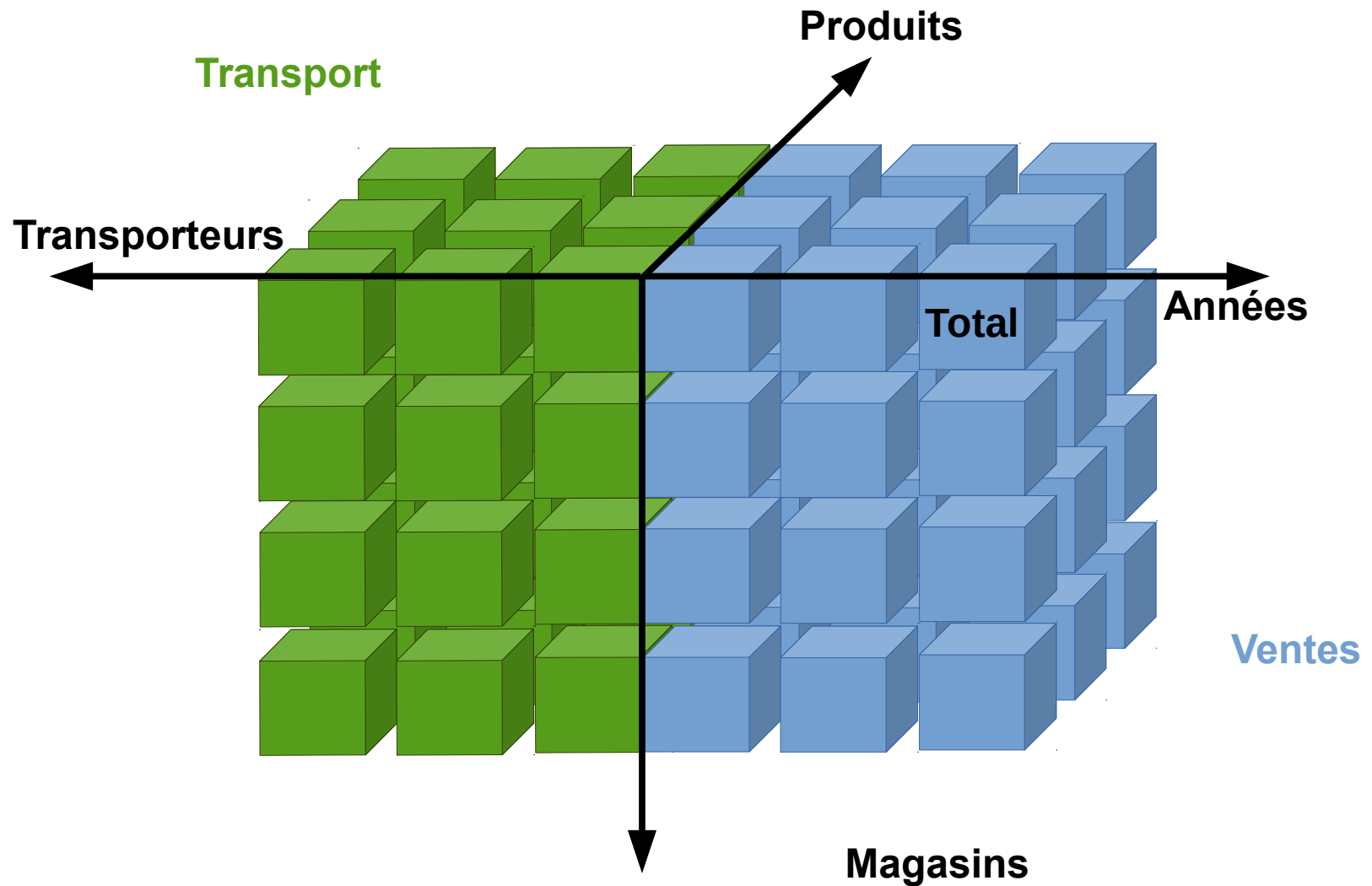
**Ceci et al., 2011** : Utilisation de la CAH pour l'intégration de données continues dans une dimension

**Données quantitatives**

# Problématique

Comment enrichir une dimension  
avec des données factuelles  
dans un schéma multidimensionnel en constellation  
?

# Problématique

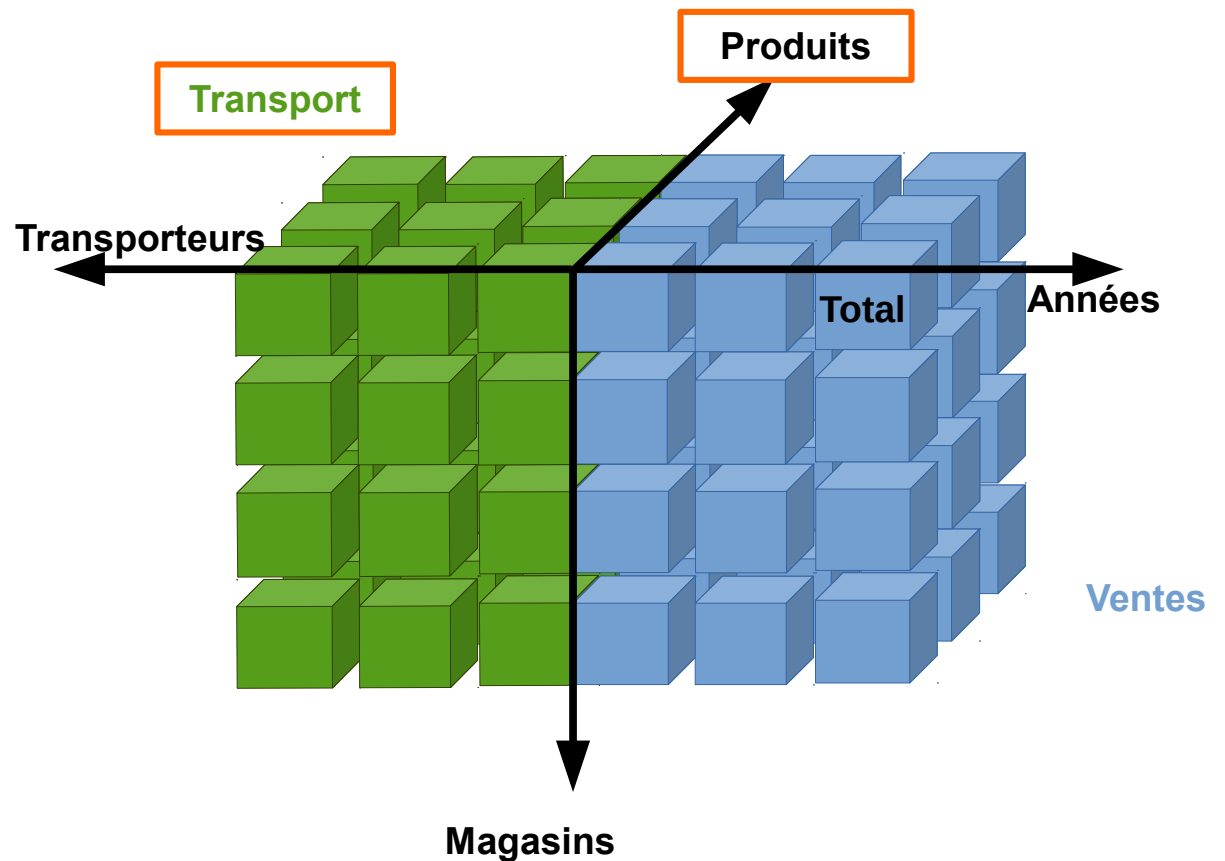


# Plan

- Introduction
- Méthodologie proposée
  - Vue générale de la méthodologie
    - Calcul de nouvelles hiérarchies
    - Transformation du schéma multidimensionnel
    - Réduction du nombre de hiérarchies
  - Expérimentation
- La méthodologie proposée en question
- Conclusion et perspectives

# Vue générale de la méthodologie

Choix d'un fait source  
et d'une dimension cible



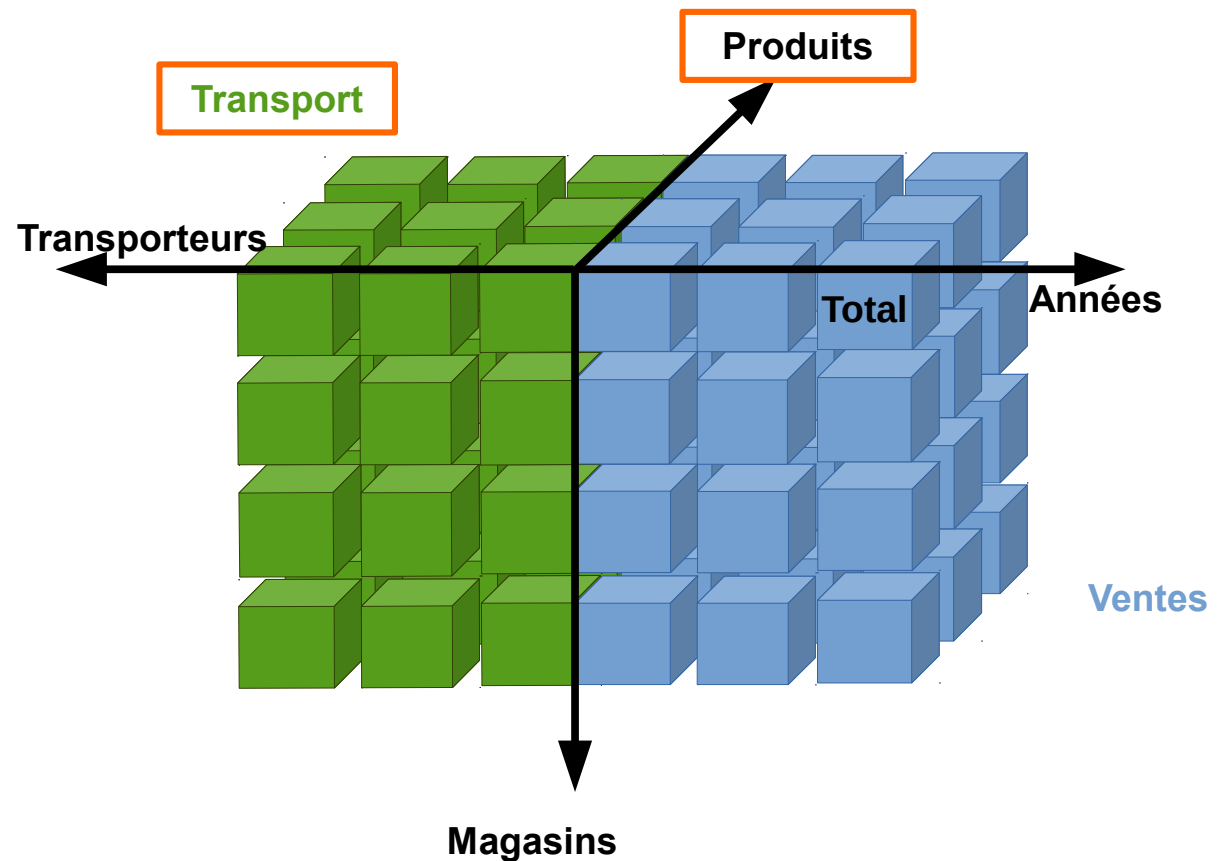


# Vue générale de la méthodologie

Choix d'un fait source  
et d'une dimension cible



Calcul de  
nouvelles hiérarchies

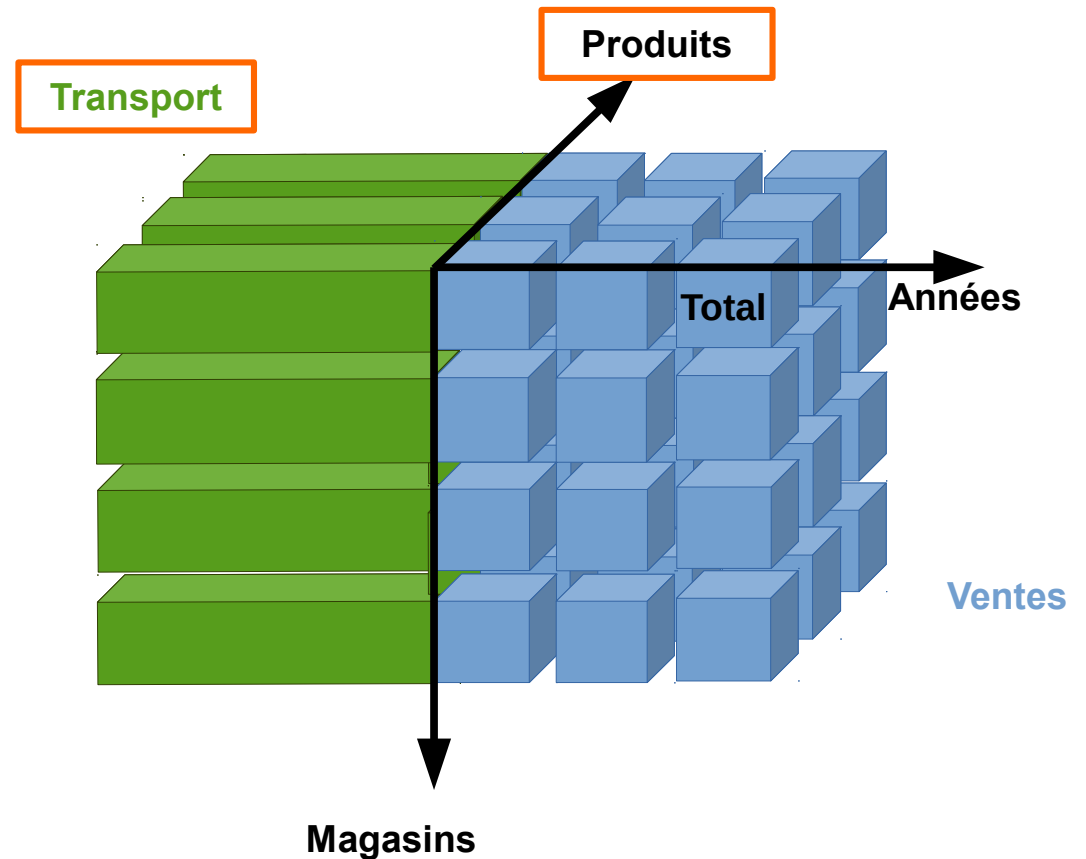


# Vue générale de la méthodologie

Choix d'un fait source  
et d'une dimension cible



Calcul de  
nouvelles hiérarchies

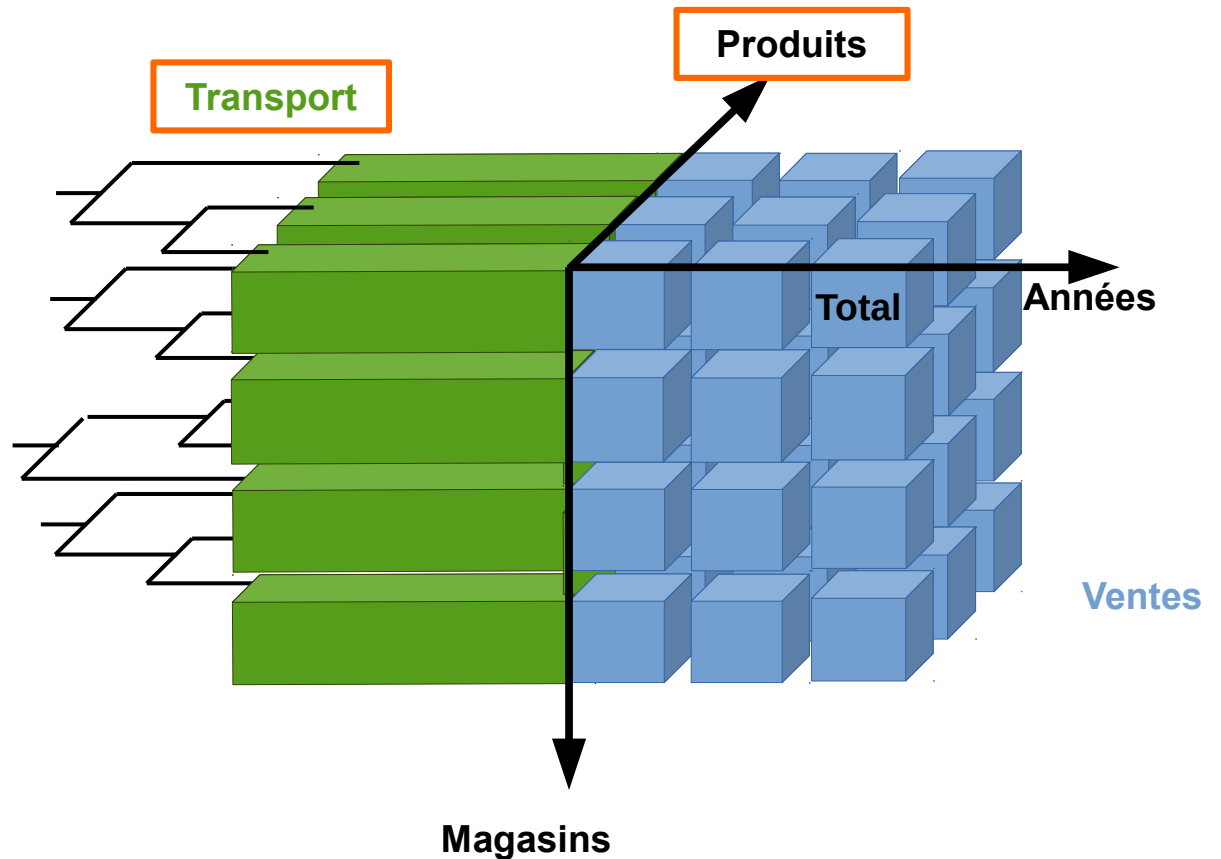


# Vue générale de la méthodologie

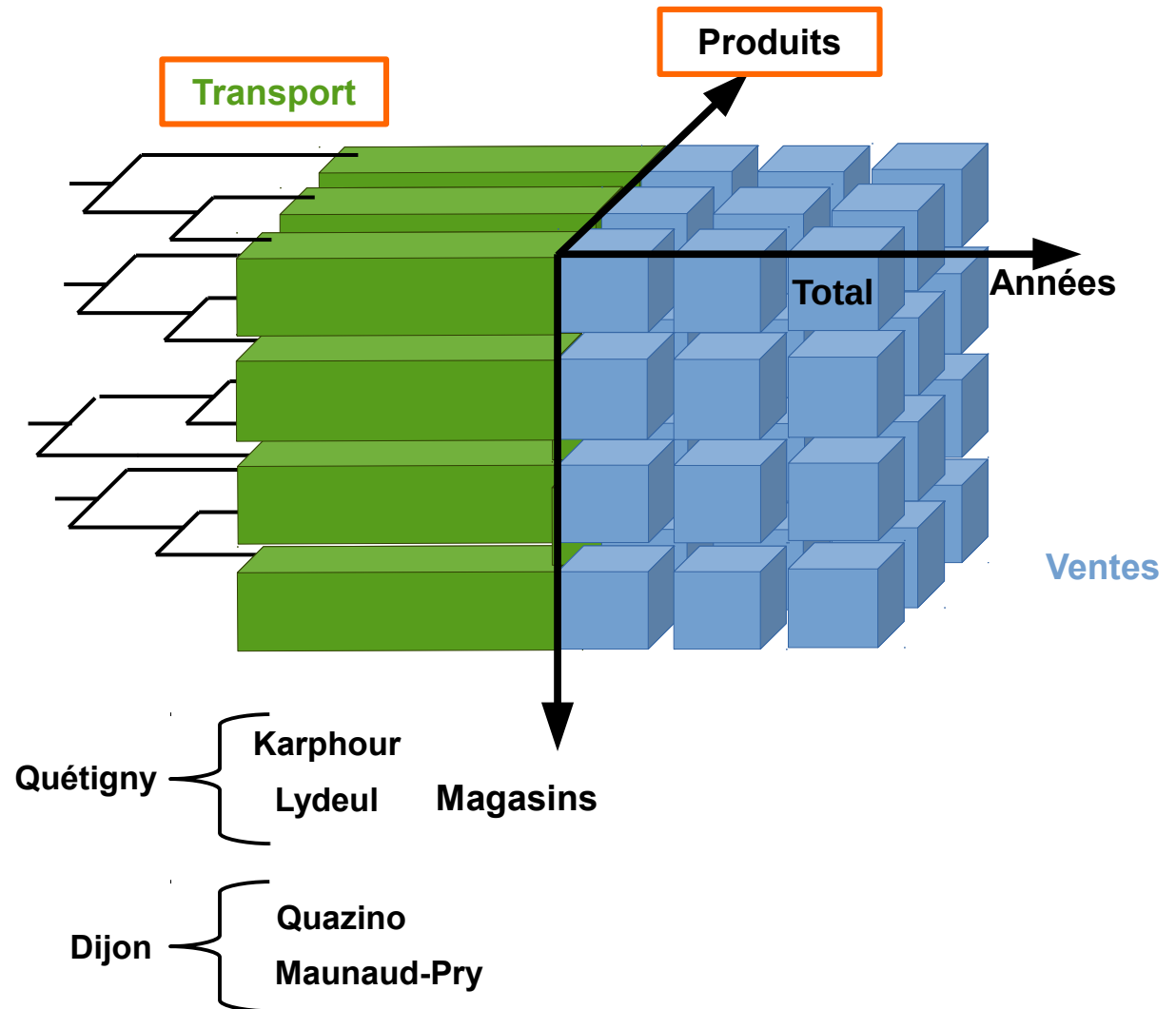
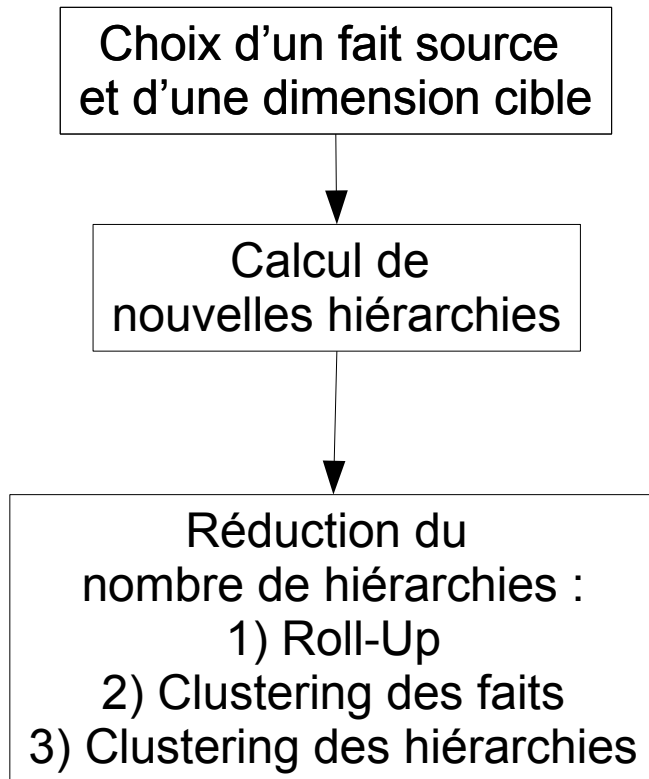
Choix d'un fait source  
et d'une dimension cible



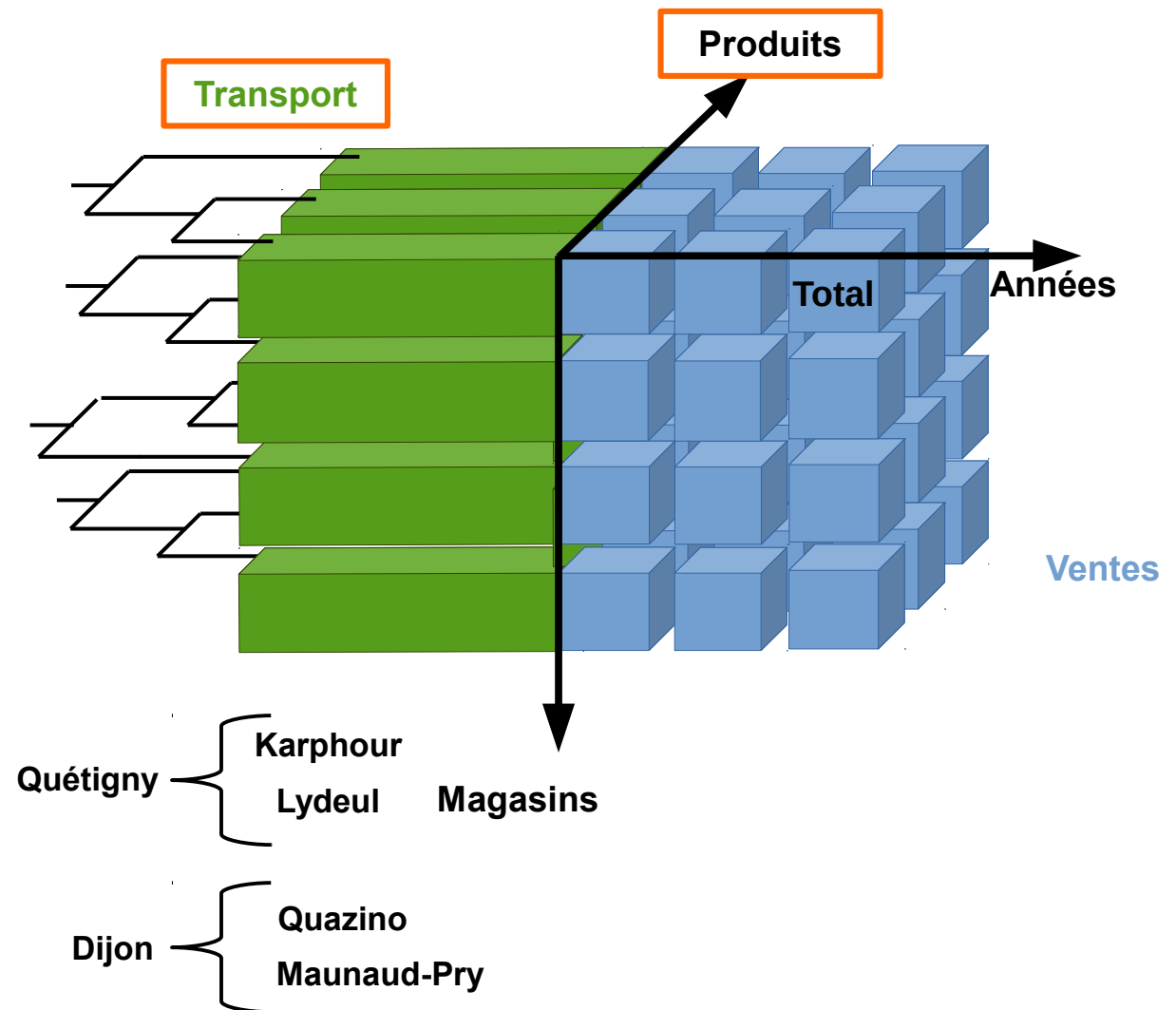
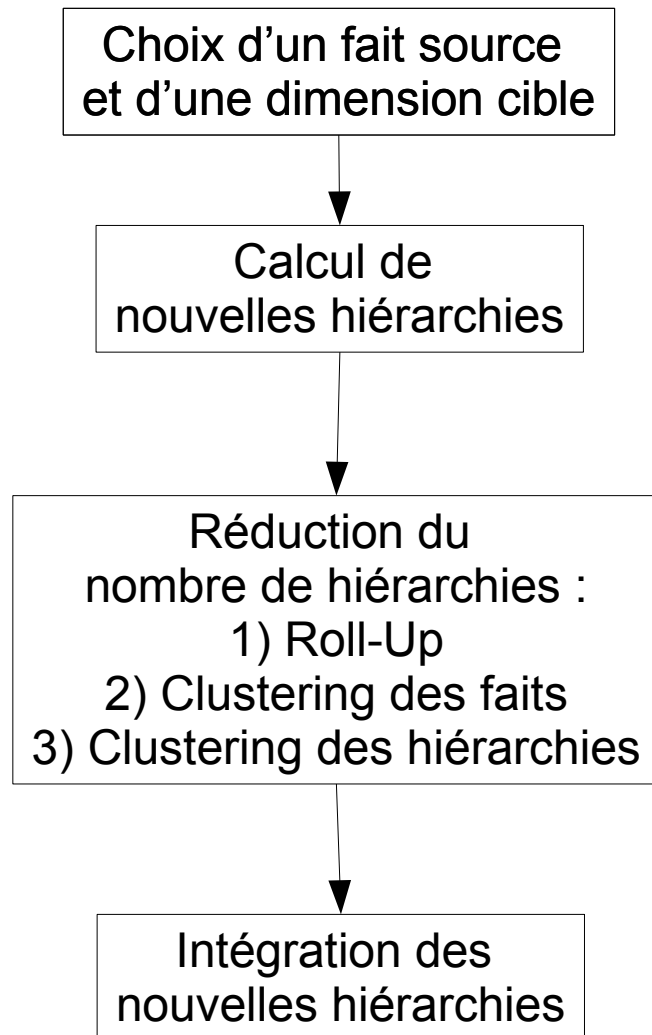
Calcul de  
nouvelles hiérarchies



# Vue générale de la méthodologie



# Vue générale de la méthodologie



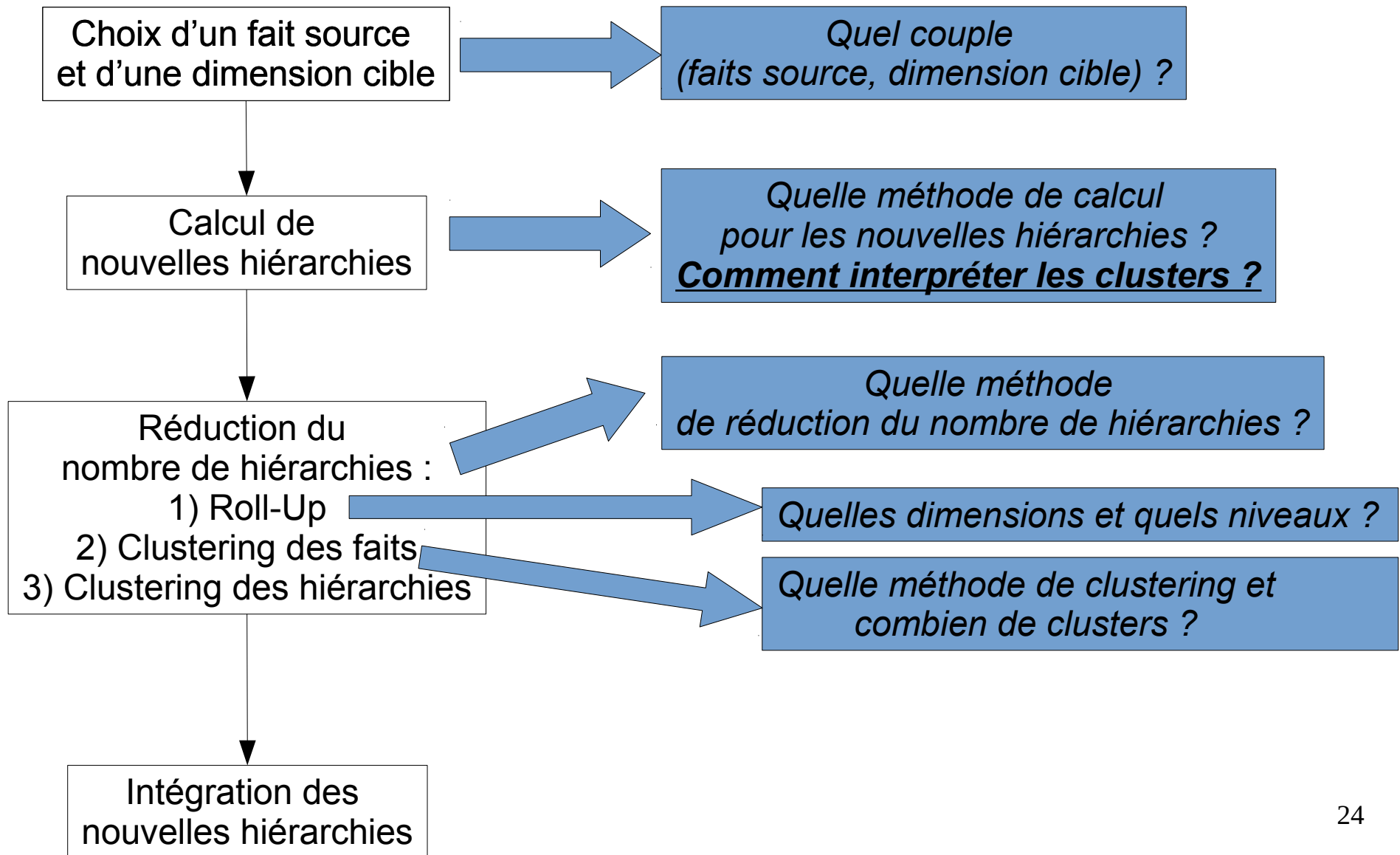
# Expérimentation

- Sur le **benchmark TPC-DS**
  - Nombre de membres très variables d'une dimension à l'autre : entre 5 et 1 920 800
- Prototype développé sous Matlab
  - Limite : au maximum  $10^8$  cases dans un tableau
- Des temps de calcul acceptables.

# Plan

- Introduction
- Méthodologie proposée
  - Vue générale de la méthodologie
    - Calcul de nouvelles hiérarchies
    - Transformation du schéma multidimensionnel
    - Réduction du nombre de hiérarchies
  - Expérimentation
- **La méthodologie proposée en question**
- Conclusion et perspectives

# Mise en oeuvre pratique





# Plan

- Introduction
- Méthodologie proposée
  - Vue générale de la méthodologie
    - Calcul de nouvelles hiérarchies
    - Transformation du schéma multidimensionnel
    - Réduction du nombre de hiérarchies
  - Expérimentation
- La méthodologie proposée en question
- Conclusion et perspectives

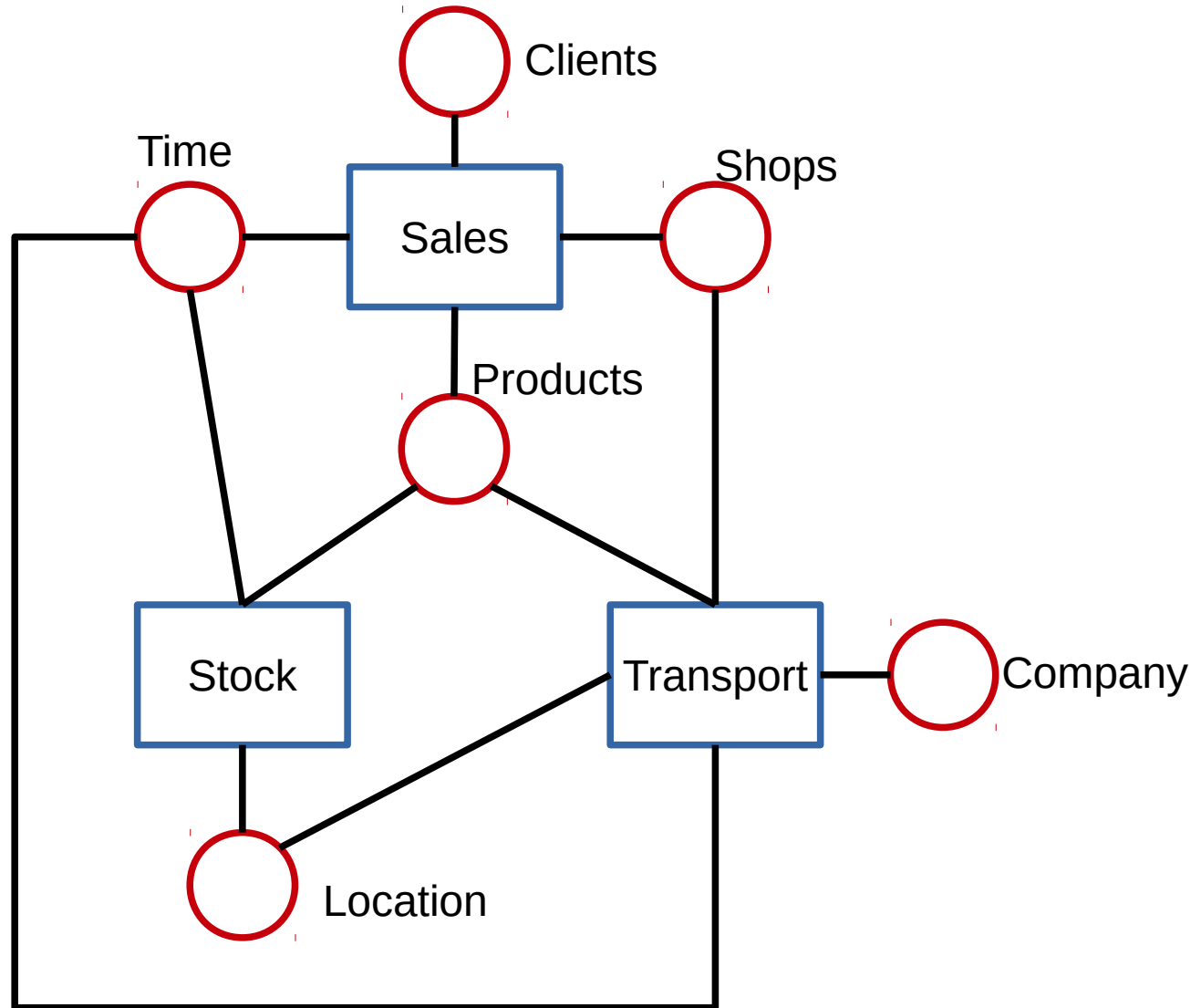
# Conclusion

- Une méthodologie pour intégrer des données factuelles au sein d'une dimension.
- Une expérimentation qui montre que la mécanique fonctionne.
- Reste l'utilisation de la méthodologie !

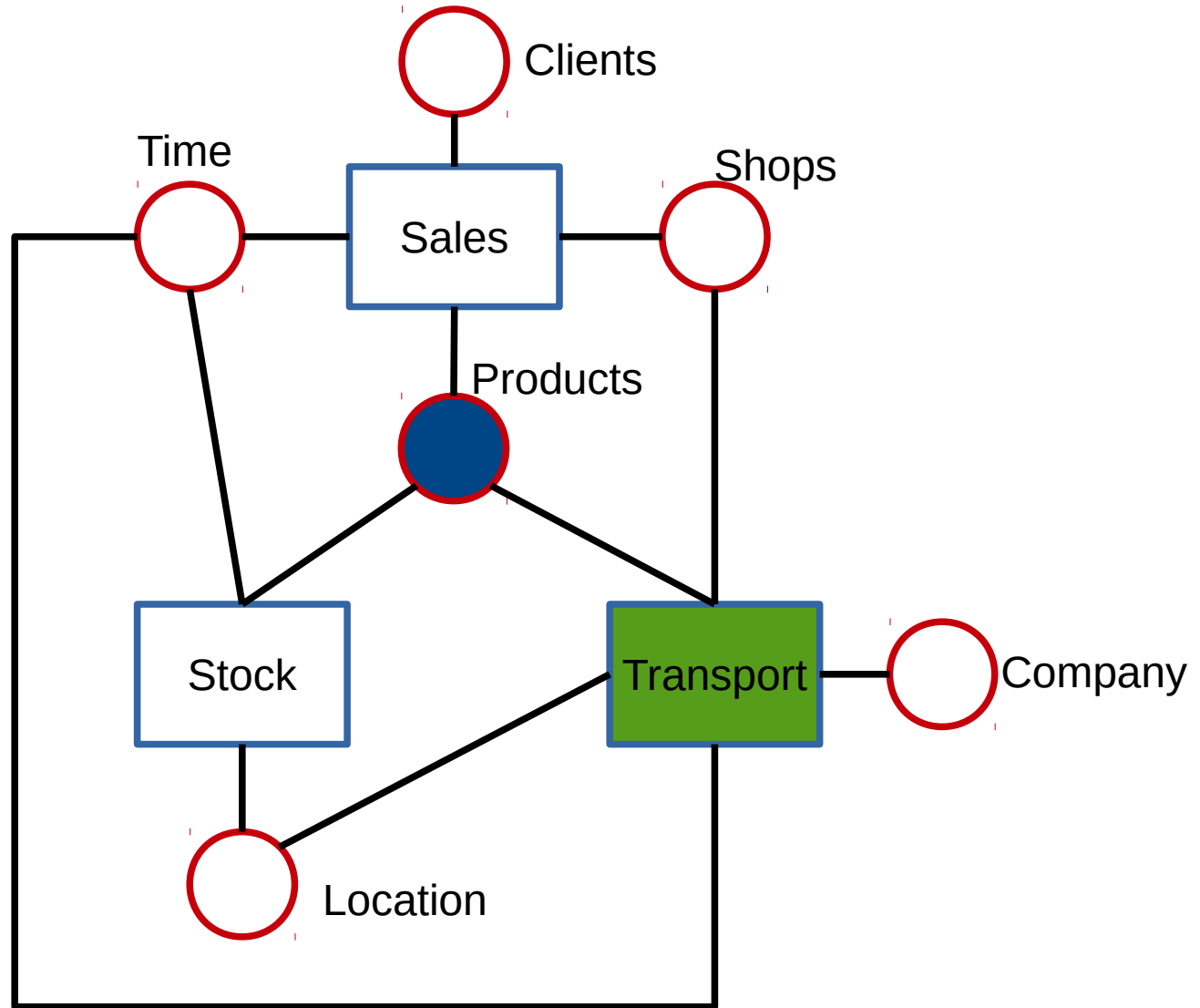
C'est fini !

Merci de votre attention !

# Schéma en constellation



# Schéma en constellation



# Calculer de nouvelles hiérarchies dans une table de faits

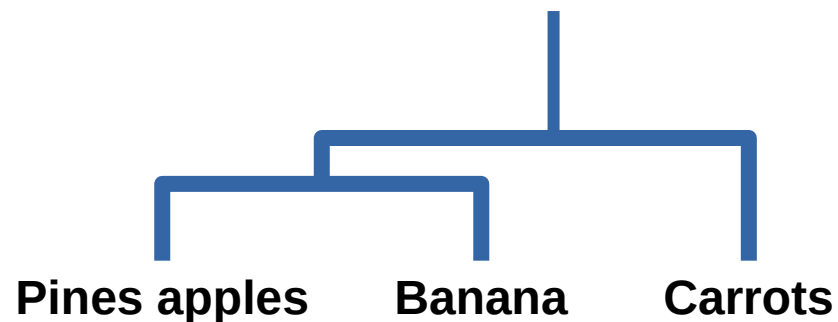
Objets

Descriptions des objets

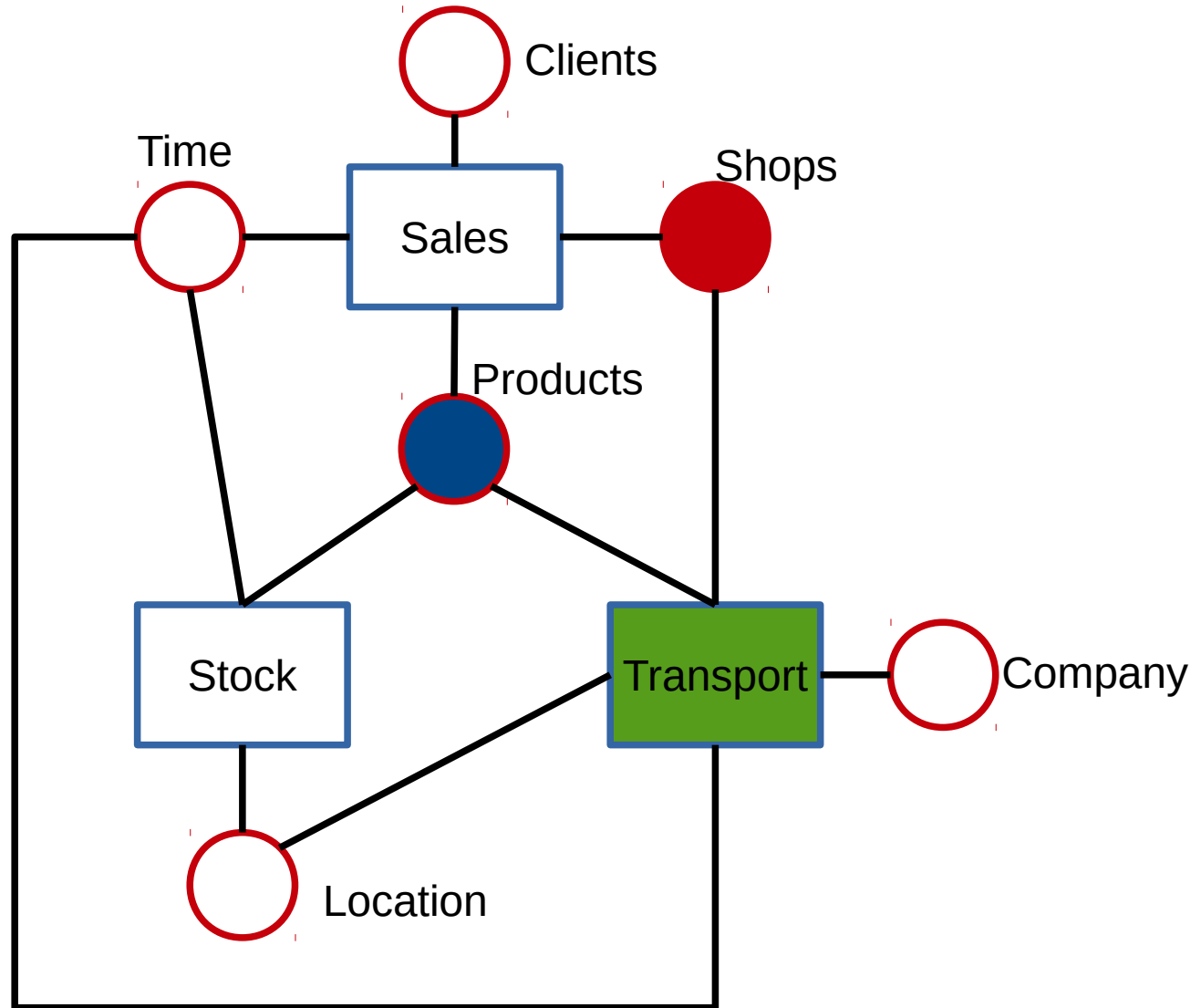
Products.Product	Transport	
	Average distance (km)	Total of weight (kg)
Pine apples	12,8	500 000
Carrots	6	8 000 000
Banana	38,4	100 000
...	...	...

# Calculer de nouvelles hiérarchies dans une table de faits

Objets	Descriptions des objets	
<b>Products.Product</b>	<b>Transport</b>	
	<b>Average distance (km)</b>	<b>Total of weight (kg)</b>
<b>Pine apples</b>	<b>12,8</b>	<b>500 000</b>
<b>Carrots</b>	<b>6</b>	<b>8 000 000</b>
<b>Banana</b>	<b>38,4</b>	<b>100 000</b>
...	...	...



# Schéma en constellation





# Partitions de la table de faits

**“Partitionneur”**

**Objets**

**Descriptions des objets**

Shops. Franchise	Shops. Shop
Organic Ultra	Vie Claire
Organic Ultra	Vie Claire
Organic Ultra	Vie Claire
Organic Ultra	Vie Claire
Cash Plus	Espace C.
Cash Plus	Espace C.
Cash Plus	Espace C.
Cash Plus	Espace C.
...	...

Products. Products
Pine apples
Carrots
Banana
...
Pine apples
Carrots
Banana
...
...

Transport	
Average distance (km)	Total of weight (kg)
12,6	500
2	1000
55,1	250
...	...
13	2000
10,9	0
23,7	5000
...	...
...	...

# Partitions de la table de faits

“Partitionneur”

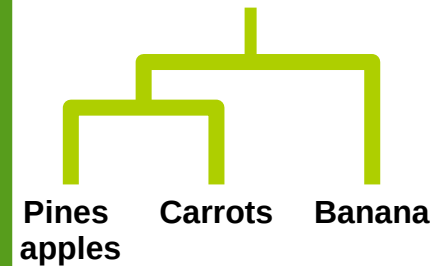
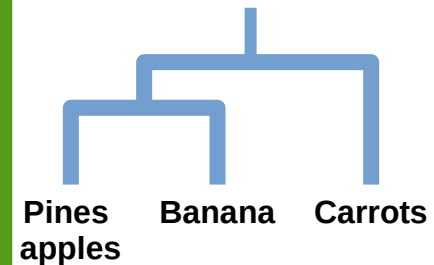
Objets

Descriptions des objets

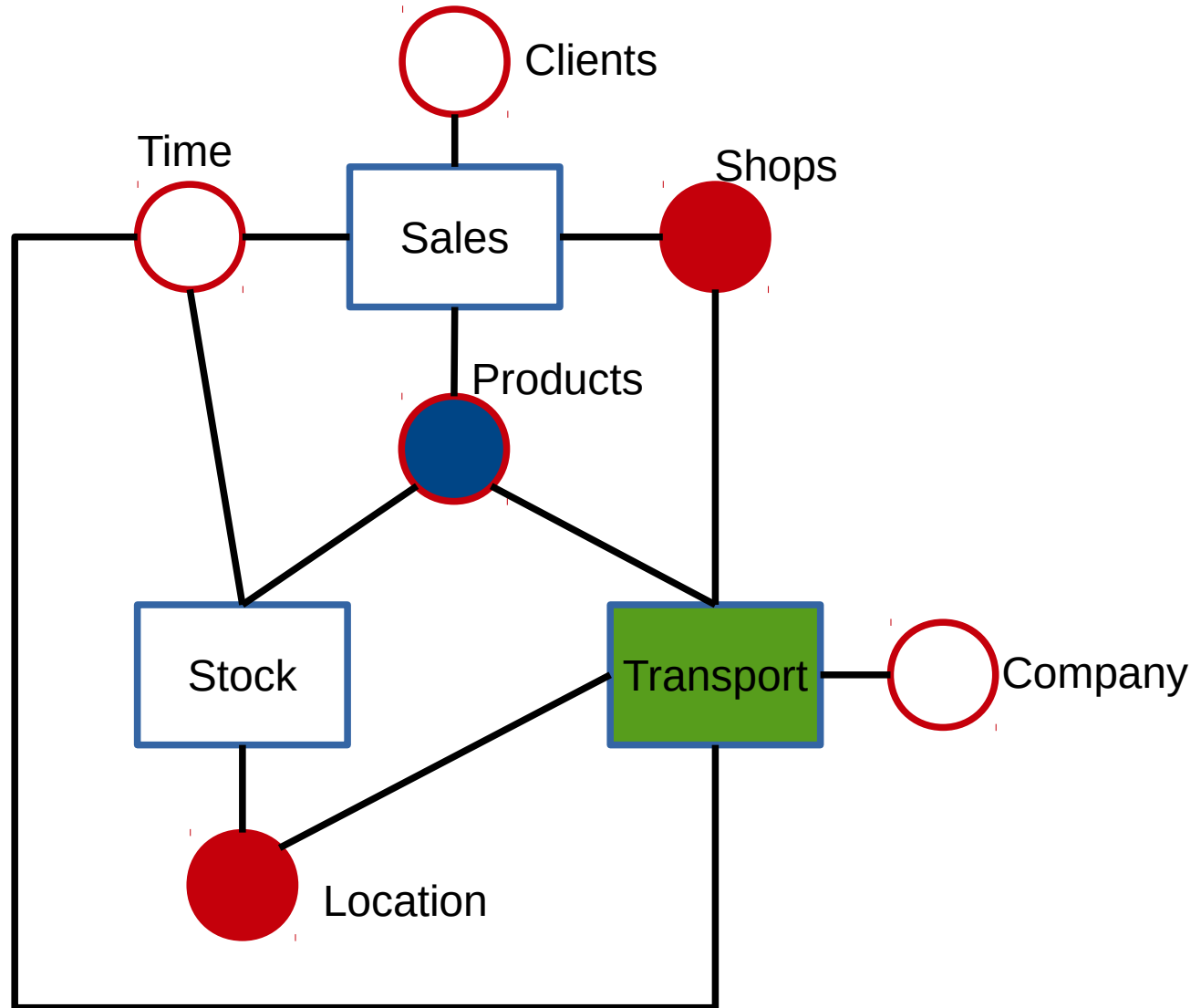
Shops. Franchise	Shops. Shop
Organic Ultra	Vie Claire
Organic Ultra	Vie Claire
Organic Ultra	Vie Claire
Organic Ultra	Vie Claire
Cash Plus	Espace C.
Cash Plus	Espace C.
Cash Plus	Espace C.
Cash Plus	Espace C.
...	...

Products. Products
Pine apples
Carrots
Banana
...
Pine apples
Carrots
Banana
...
...

Transport	
Average distance (km)	Total of weight (kg)
12,6	500
2	1000
55,1	250
...	...
13	2000
10,9	0
23,7	5000
...	...
...	...



# Schéma en constellation



# Partitions de la table de faits

“Partitionneur”

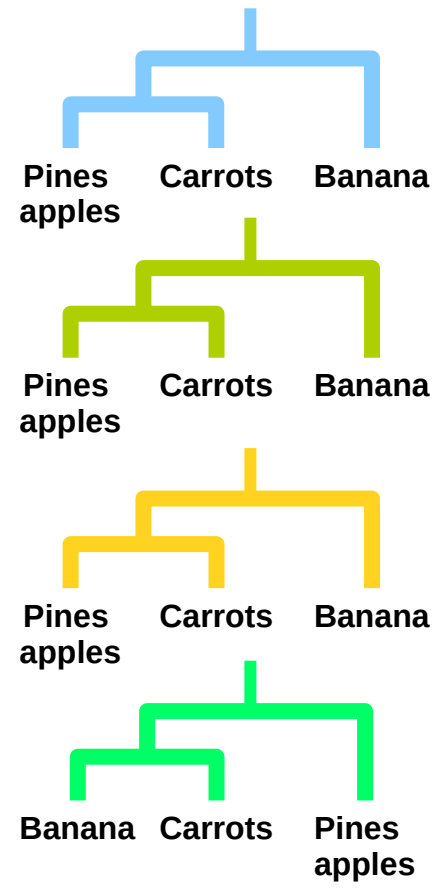
Objets

Descriptions des objets

Location. Farm	Shops. Franchise	Shops. Shop
Jasper's Farm	Organic Ultra	
Jasper's Farm	Organic Ultra	
Jasper's Farm	Organic Ultra	
Jasper's Farm	Organic Ultra	
Jasper's Farm	Organic Ultra	Vie Claire
Jasper's Farm	Organic Ultra	Vie Claire
Jasper's Farm	Organic Ultra	Vie Claire
Jasper's Farm	Organic Ultra	Vie Claire
Jasper's Farm	Organic Ultra	Organik Market
Jasper's Farm	Organic Ultra	Organik Market
Jasper's Farm	Organic Ultra	Organik Market
Jasper's Farm	Organic Ultra	Organik Market
The Grouse	Organic Ultra	
The Grouse	Organic Ultra	
The Grouse	Organic Ultra	
The Grouse	Organic Ultra	

Products. Product
Pine apples
Carrots
Banana
...
Pine apples
Carrots
Banana
...
Pine apples
Carrots
Banana
...
Pine apples
Carrots
Banana
...

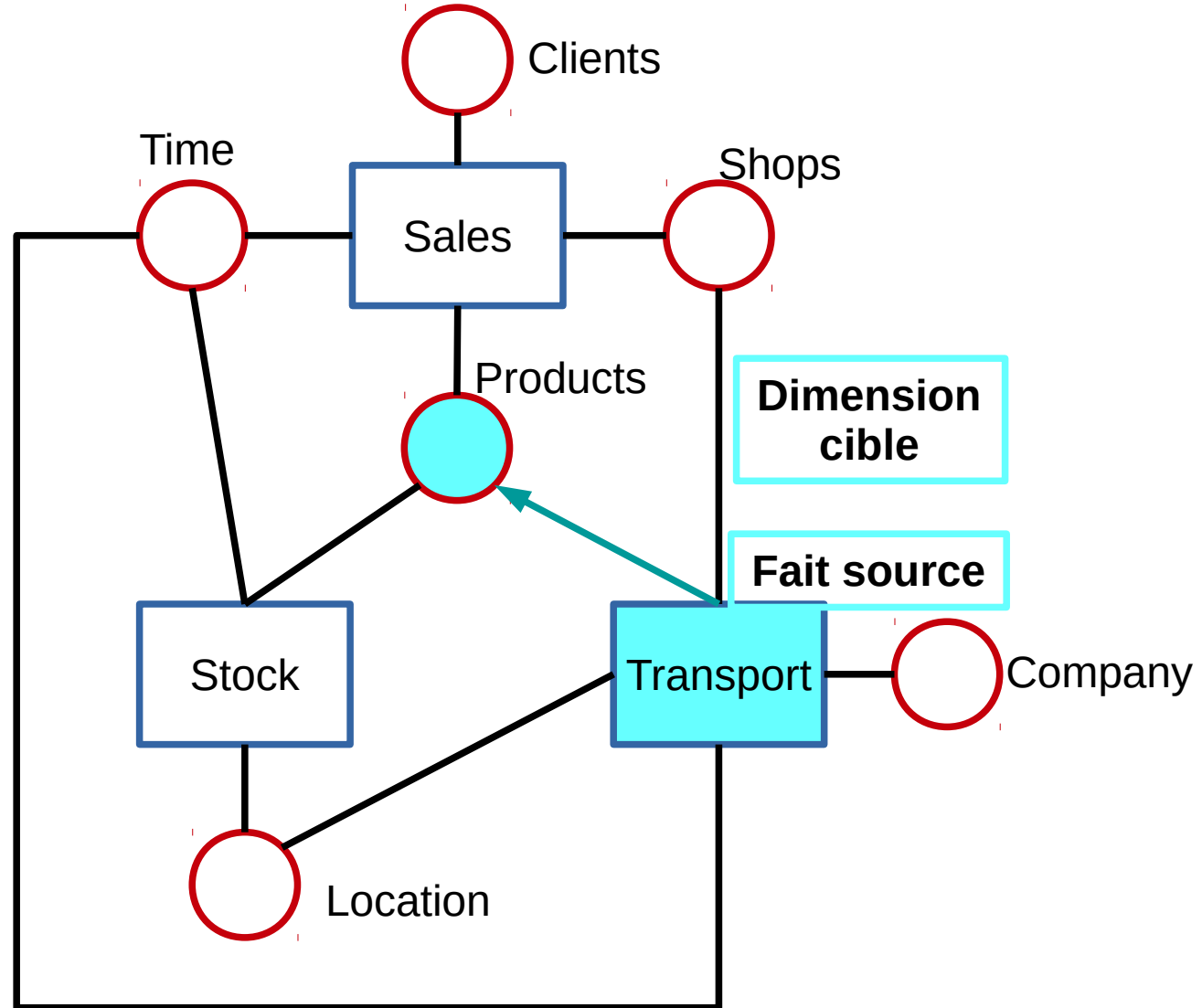
Transport	
Average istance (km)	Total of weight (kg)
12,9	350
5,4	800
39,5	50
...	...
14	250
2	800
55,1	0
...	...
9	100
10,9	0
23,7	50
...	...
16,2	50
6	100
17,4	200
...	...



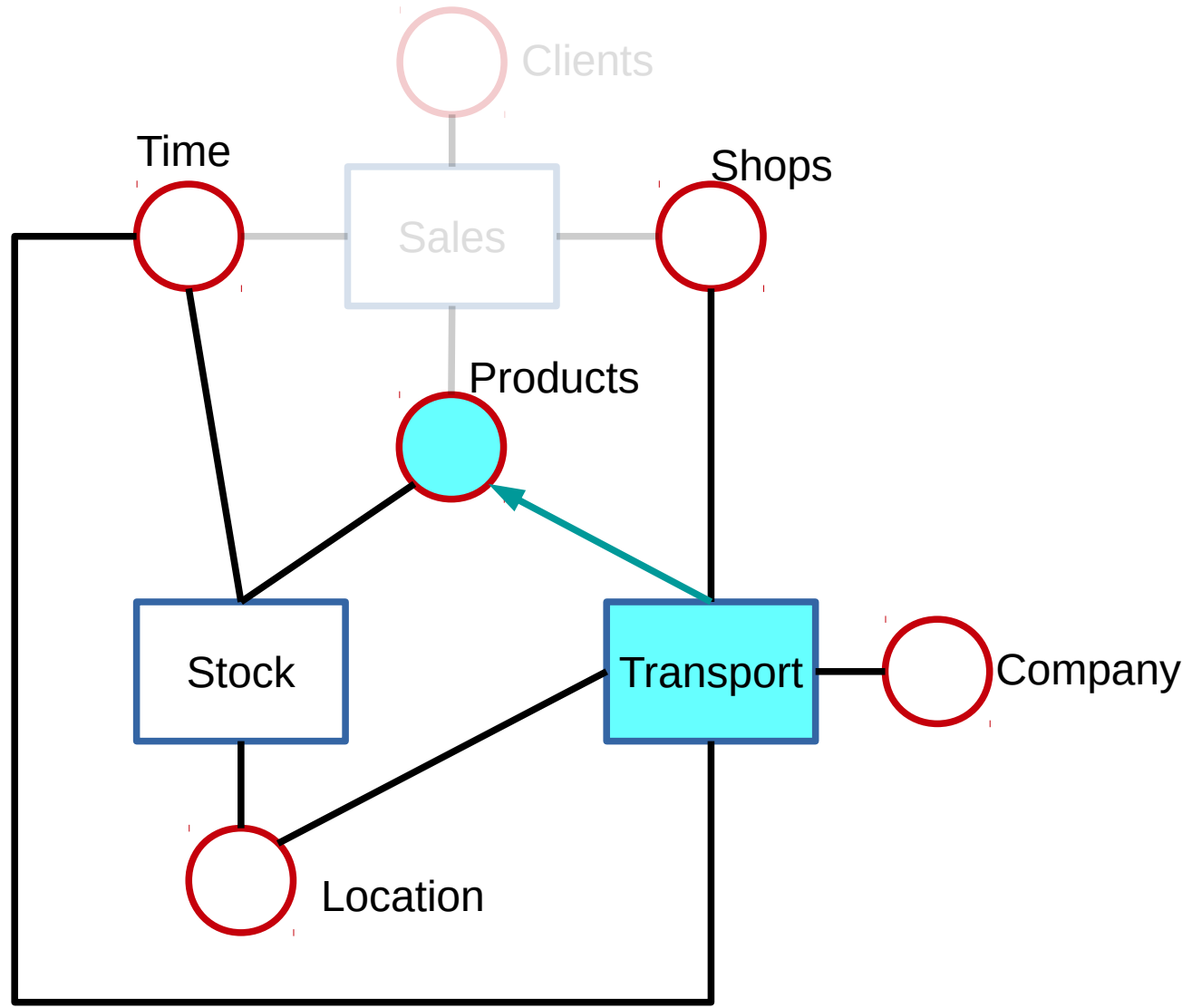
# Contribution

- Vue générale de la méthodologie
- Calcul de nouvelles hiérarchies
- Transformation du schéma multidimensionnel
- Réduction du nombre de hiérarchies
- Expérimentation

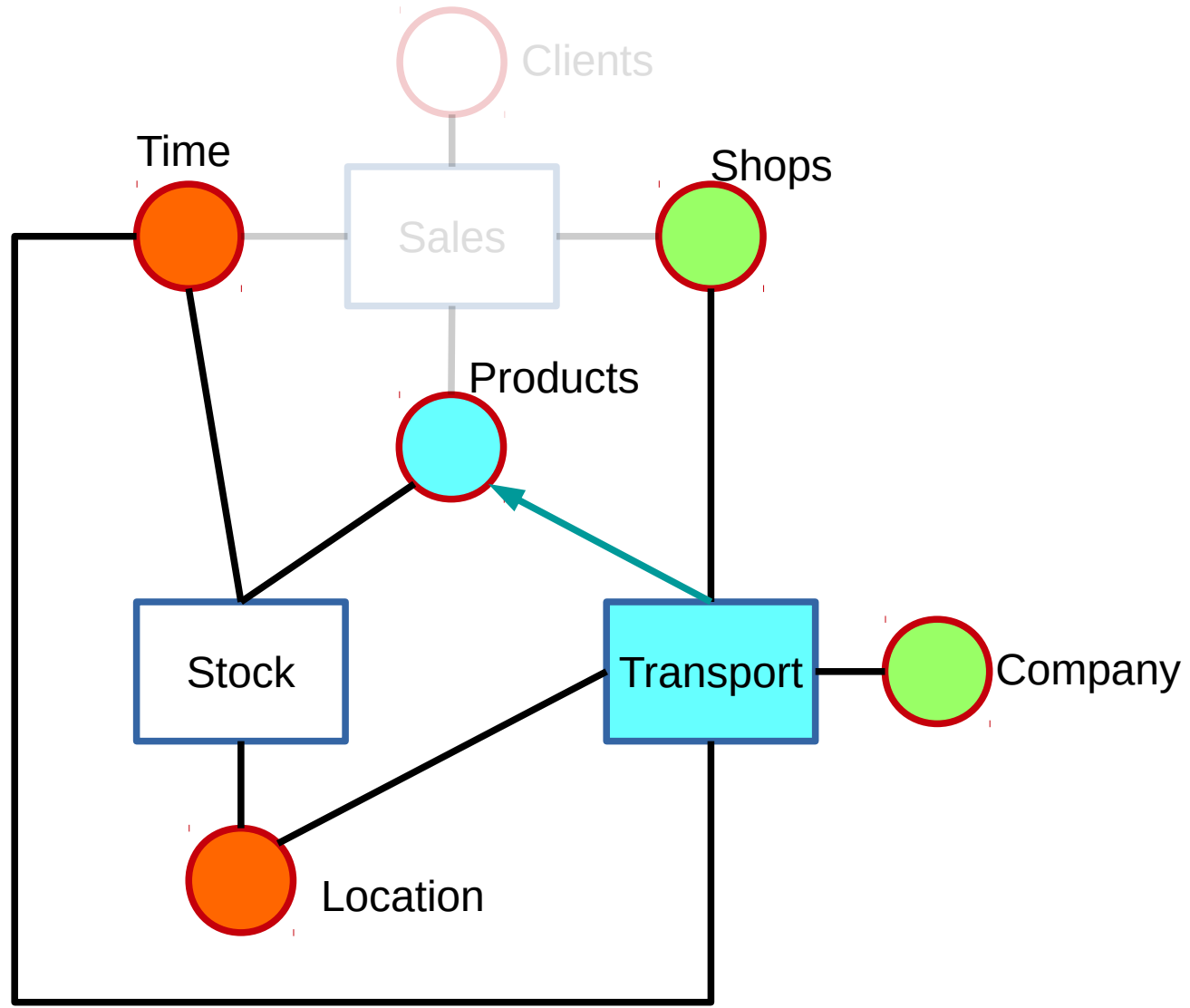
# Transformation du schéma multidimensionnel



# Du point de vue de “stock”



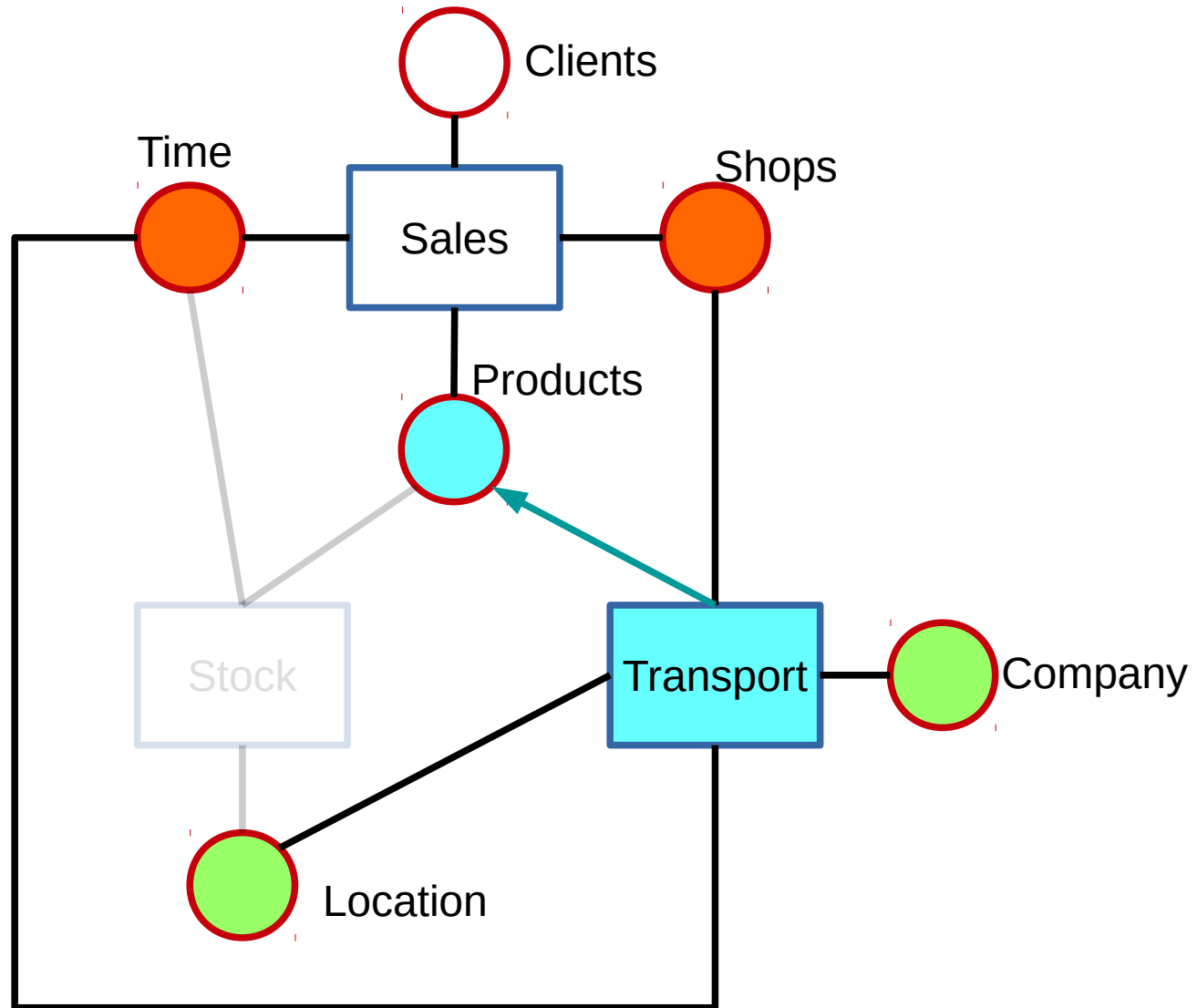
# Du point de vue de “stock”



$$C(\text{stock}) = \{\text{location, time}\}$$

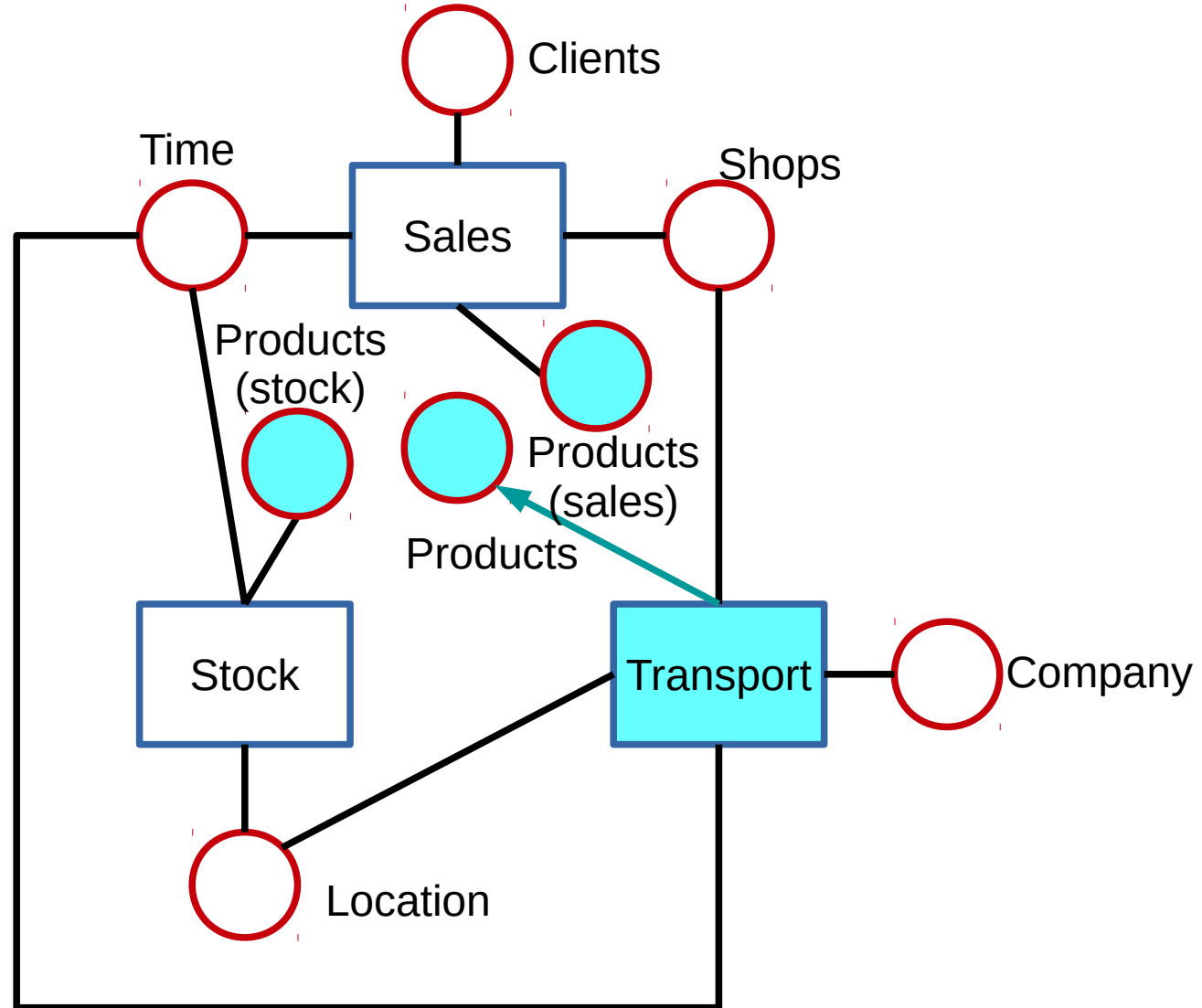


# Du point de vue de “sales”

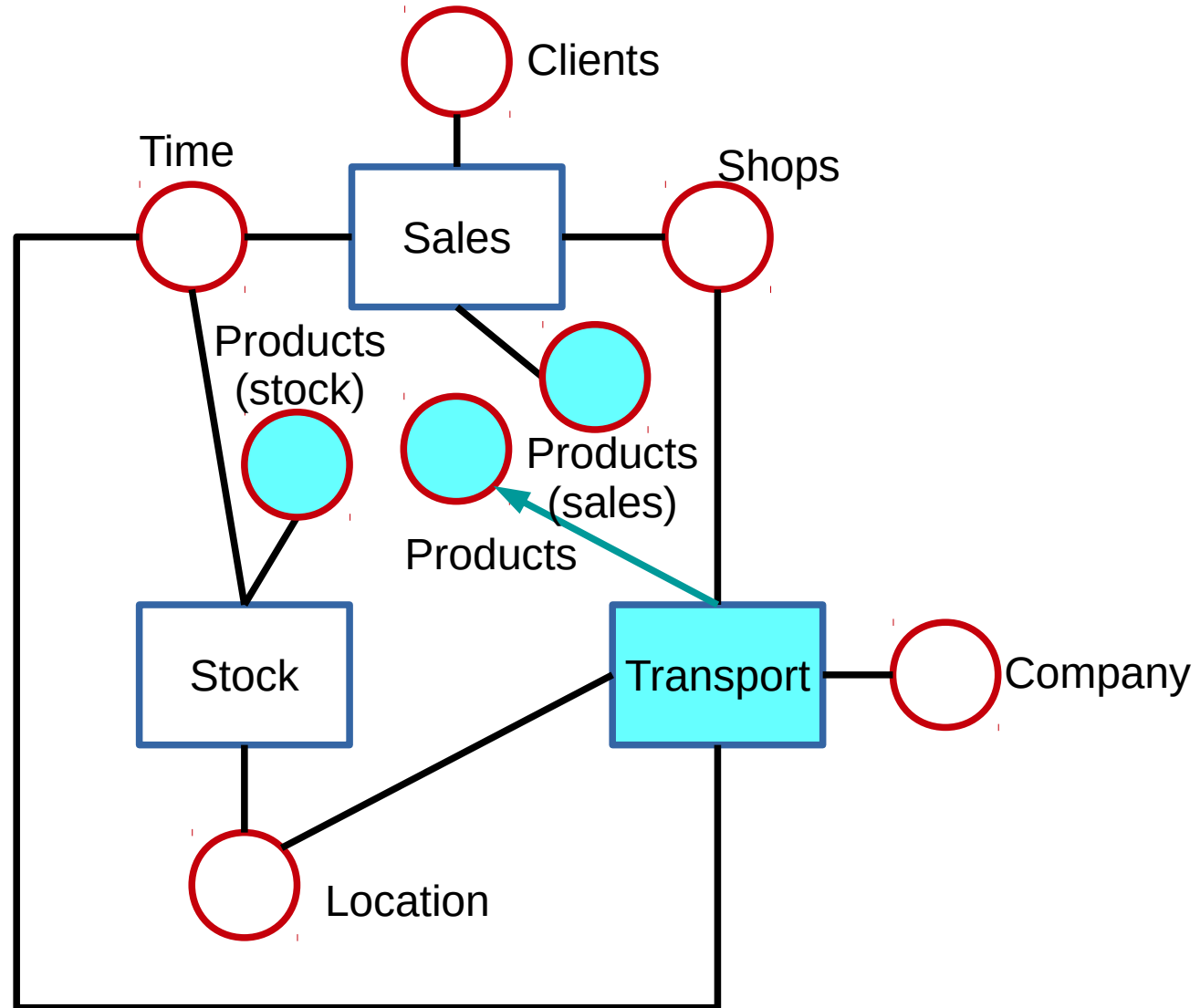


$C(\text{stock}) = \{\text{location}, \text{time}\}$   
 $C(\text{sales}) = \{\text{shops}, \text{time}\}$

# Transformation du schéma multidimensionnel



# Transformation du schéma multidimensionnel



# Partitions de la table de faits

“Partitionneur”

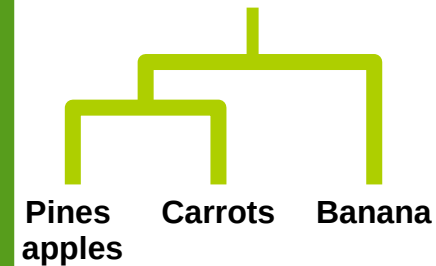
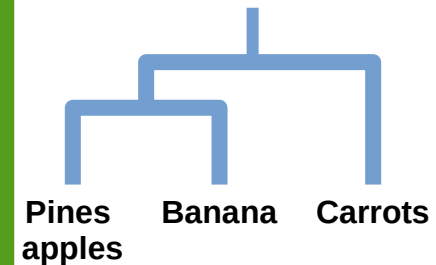
Objets

Descriptions des objets

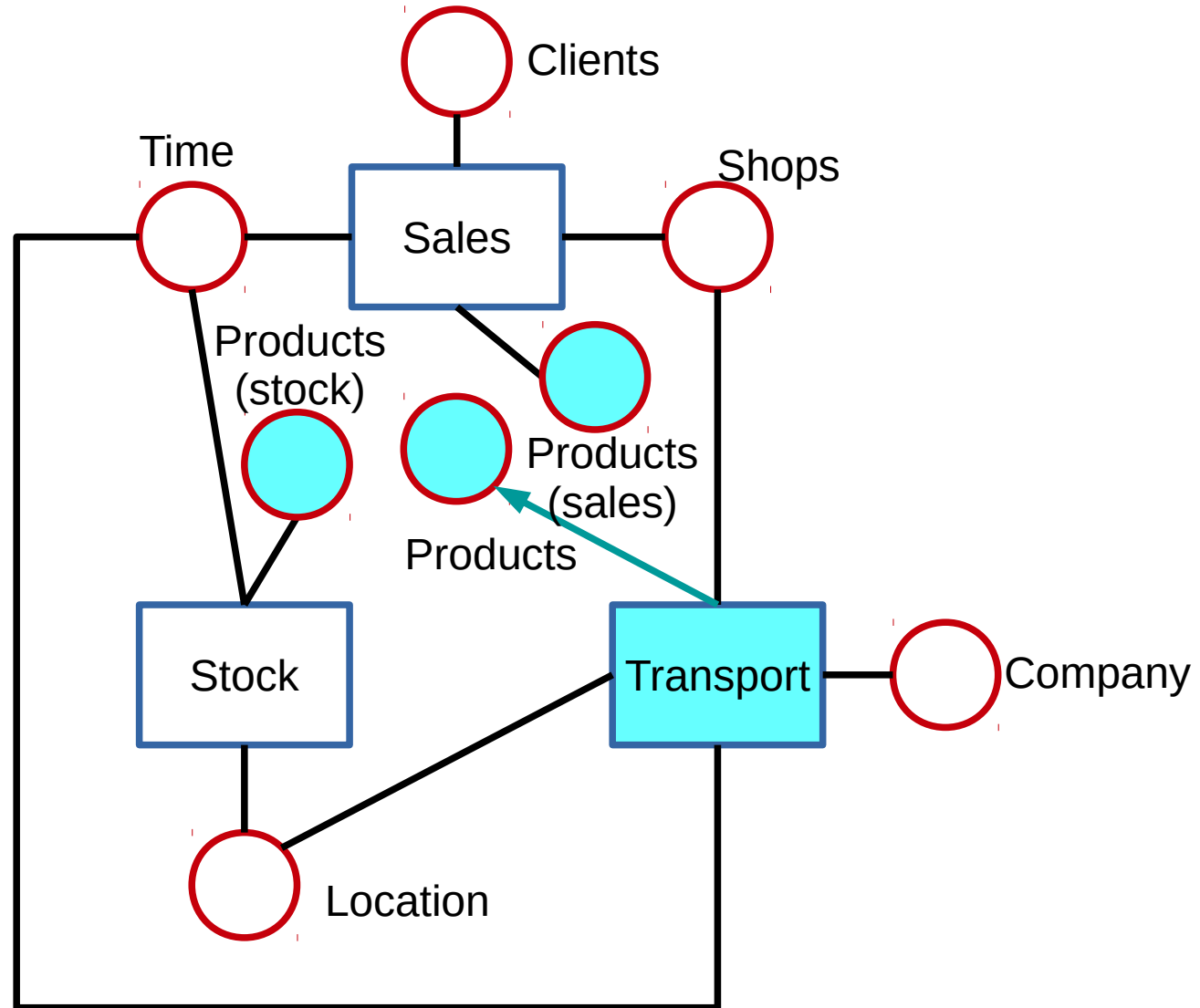
Shops. Franchise	Shops. Shop
Organic Ultra	Vie Claire
Organic Ultra	Vie Claire
Organic Ultra	Vie Claire
Organic Ultra	Vie Claire
Cash Plus	Espace C.
Cash Plus	Espace C.
Cash Plus	Espace C.
Cash Plus	Espace C.
...	...

Products. Products
Pine apples
Carrots
Banana
...
Pine apples
Carrots
Banana
...
...

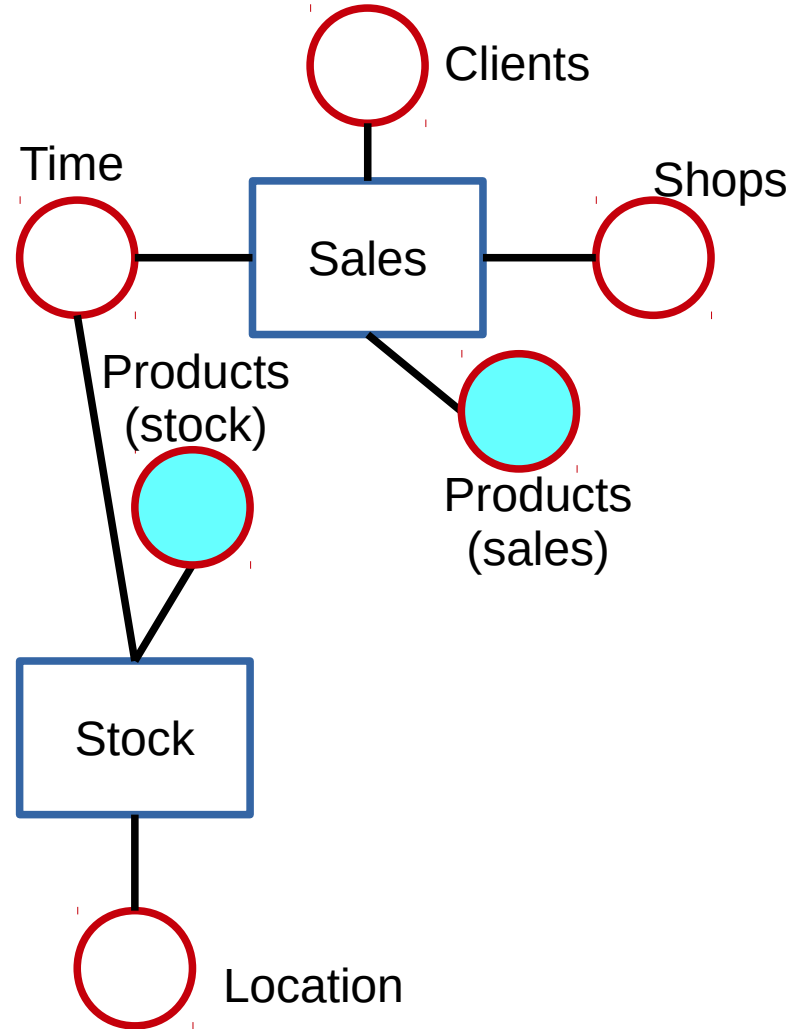
Transport	
Average distance (km)	Total of weight (kg)
12,6	500
2	1000
55,1	250
...	...
13	2000
10,9	0
23,7	5000
...	...
...	...



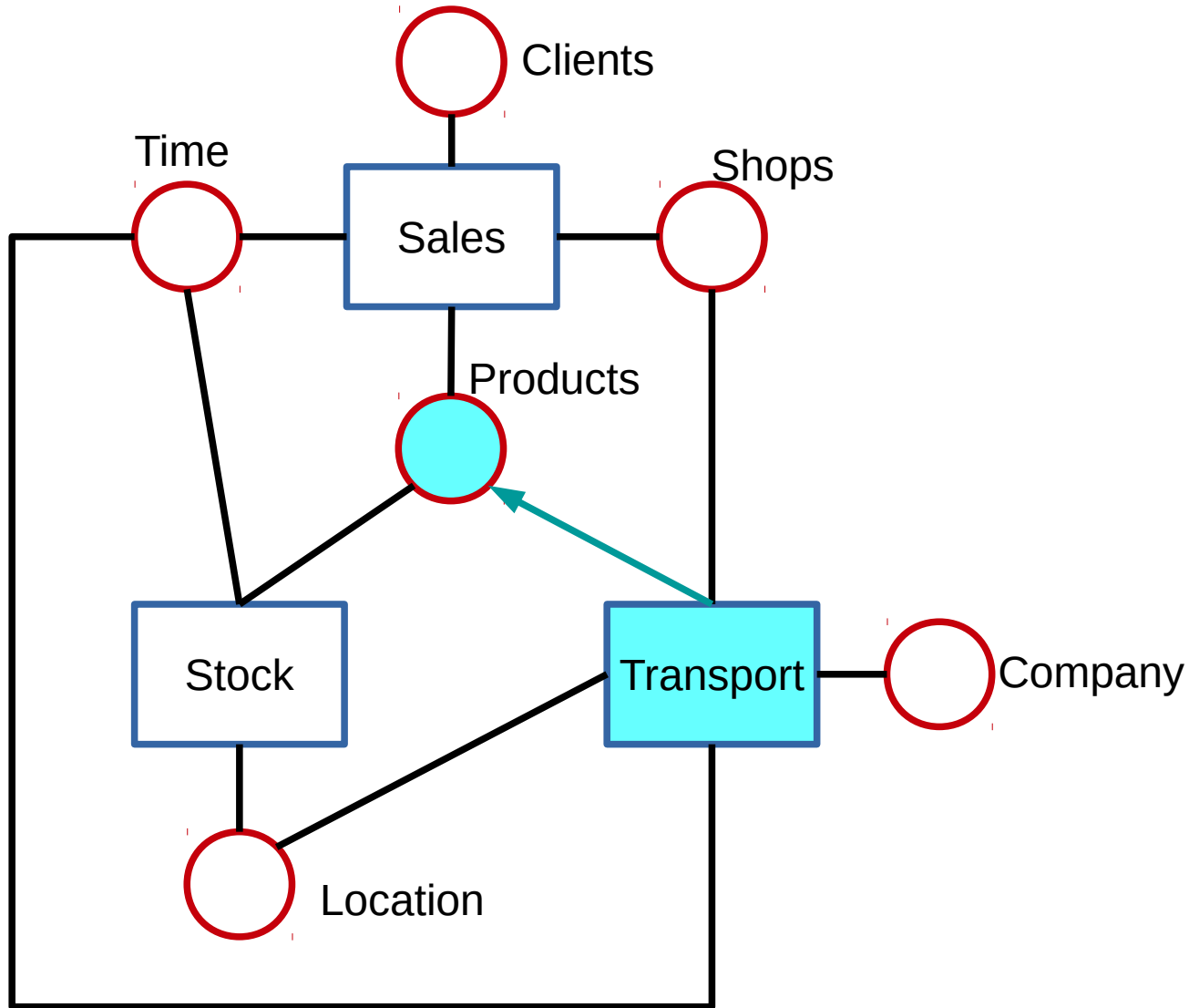
# Transformation du schéma multidimensionnel



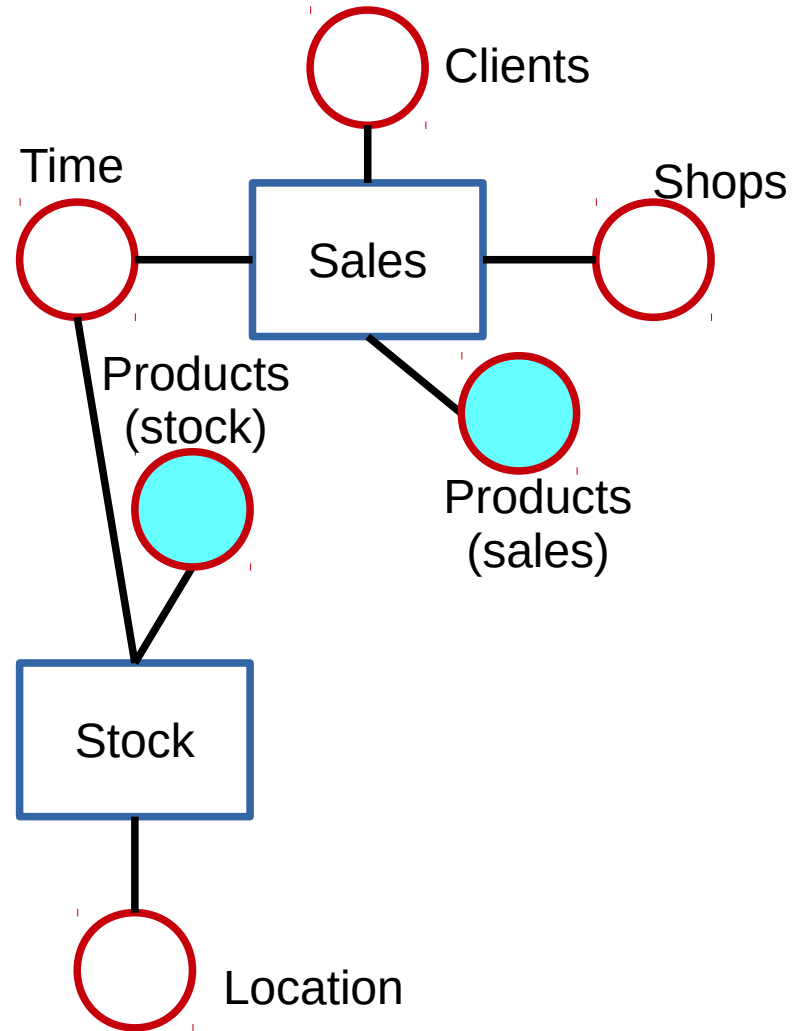
# Transformation du schéma multidimensionnel



# Avant

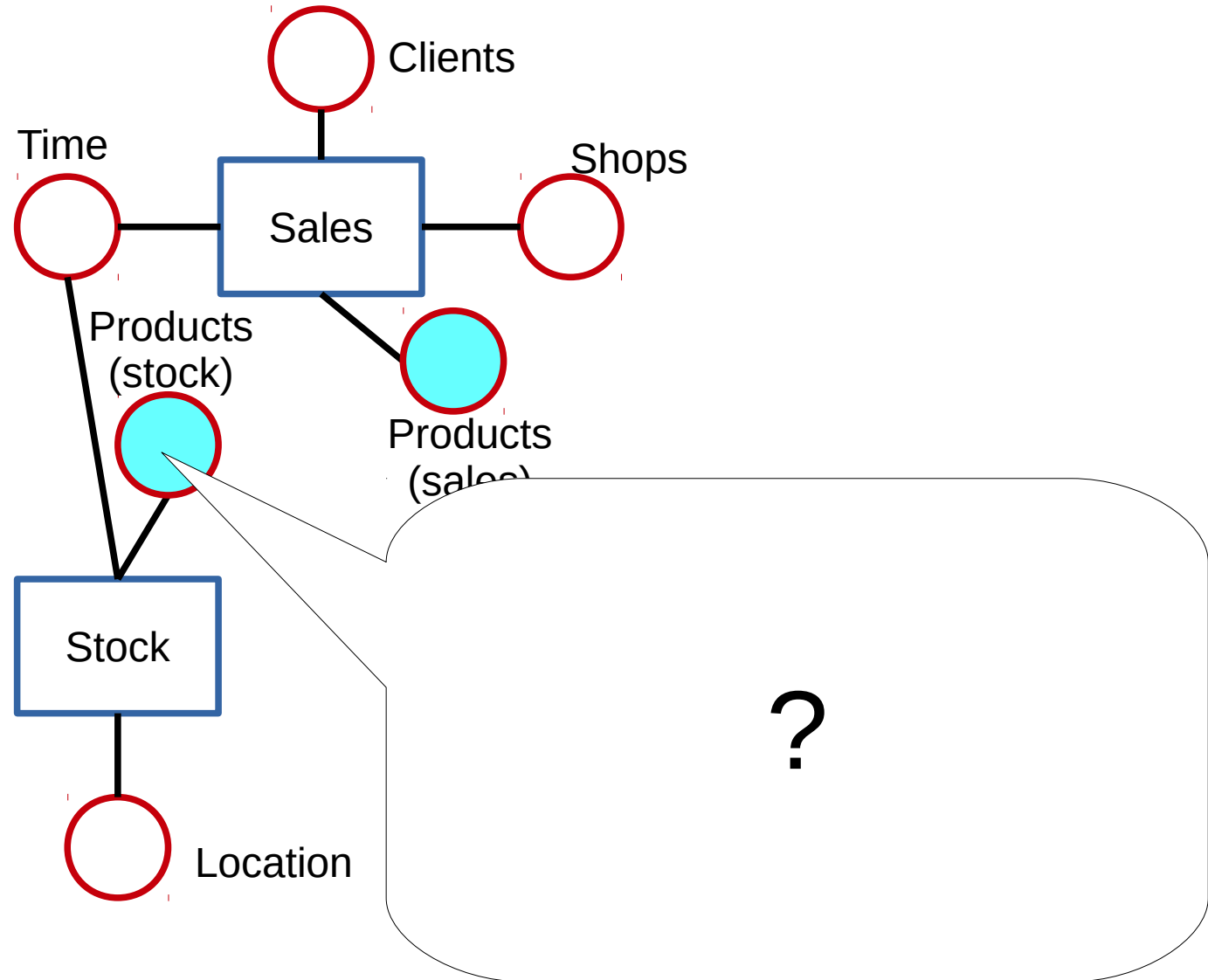


# Après



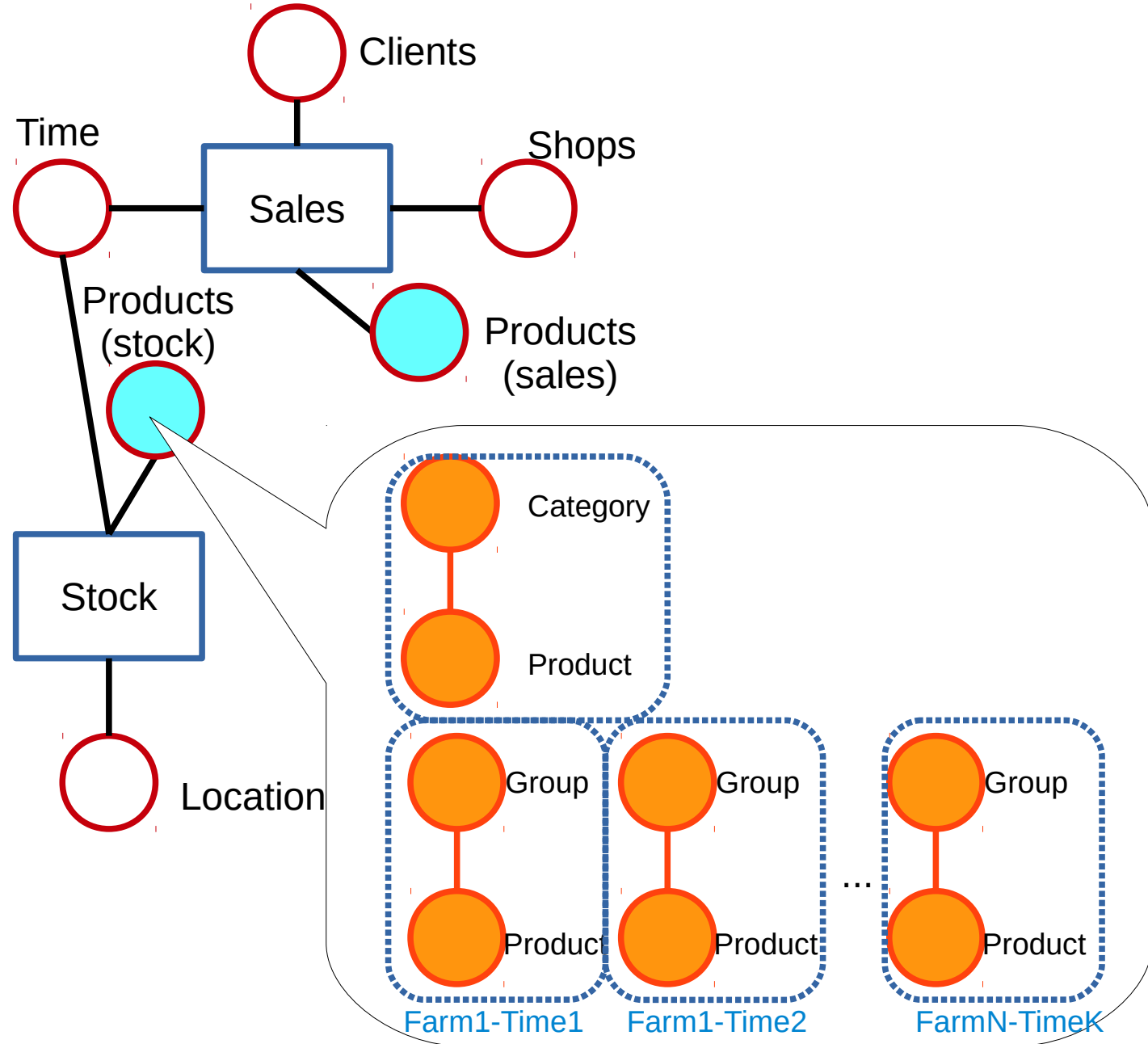


# Les hiérarchies contextuelles calculées



$$C(\text{stock}) = \{\text{location}, \text{time}\}$$

# Les hiérarchies contextuelles calculées



$$C(\text{stock}) = \{\text{location}, \text{time}\}$$

# Réduction du nombre de hiérarchies

Réduction	Réduction appliquée ...	Critère de réduction	Inconvénients / avantages
Roll-Up des dimensions contextuelles	Avant le calcul des hiérarchies	Conceptuel	<p><b>La granularité des faits peut changer</b></p> <p>Temps de calcul ?</p>
Clustering des faits sources	Avant le calcul des hiérarchies	Structurel	<p><b>Les approximations sont tolérées</b></p> <p>Pas de changement de la granularité des faits</p> <p>Temps de calcul ?</p>
Clustering des hiérarchies	Après le calcul des hiérarchies	Structurel	<p>Pas d'approximations</p> <p>Pas de changement de la granularité des faits</p> <p>Temps de calcul ?</p>