



**Université de Sherbrooke**

## **Bases de données dimensionnelles**

*Interrogation*

**UdeS:BDD\_20**

Christina KHNAISSER ([christina.khnaisser@usherbrooke.ca](mailto:christina.khnaisser@usherbrooke.ca))

—

*CoFELI/Scriptorum/BDD\_30-Interrogation, version 1.0.1.b, en date du 2025-11-20*

*— en vigueur —*

## Sommaire

Introduction aux différentes catégories de requêtes pour les bases de données dimensionnelles.

## Mise en garde

Le présent document est en cours d'élaboration ; en conséquence, il est incomplet et peut contenir des erreurs.

## Historique

diffusion	resp.	description
2025-11-09	LL	Correction de coquilles.
2025-04-05	CK	Mise à jour et corrections mineures.
2024-10-06	CK	Ébauche initiale.

## Table des matières

Introduction.....	4
1. Mise en contexte .....	4
2. Catégorie de requêtes .....	4
2.1. Browser .....	4
2.2. Drill up and down .....	5
2.3. Drill across .....	6
2.4. Pivot.....	11
3. Schémas dérivés .....	12
3.1. Fusion des faits ( <i>merged fact</i> ) .....	12
3.2. Pivotement des faits ( <i>Pivoted fact</i> ) .....	13
3.3. Partitionnement des faits ( <i>Sliced fact</i> ) : .....	13
3.4. Aggrégation des faits.....	14
3.5. Et les vues?.....	15
4. Règles de pratique.....	16
A. Exemple de code SQL .....	16
Références .....	16

# Introduction

Le présent document a pour but de présenter les différentes catégories de requêtes pour les bases de données dimensionnelles.

## Contenu des sections

- La section 1 expose les catégories de requêtes.
- La section 2 présente les catégories de requêtes.
- La section 3 présente les schémas dérivés
- La section 4 présente les règles de pratiques pour la mise en oeuvre des requêtes

## Évolution du document

La première version du document a été établie sur la base des travaux publiés par Adamson et plus particulièrement [Adamson2010a], chapitres 4, 9, 14 et 15.

## 1. Mise en contexte

Une base de données dimensionnelles s'intéresse à l'évaluation des processus d'un domaine.

Les besoins analytiques permettent d'identifier des requêtes pour extraire les données nécessaires pour l'évaluation.

## 2. Catégorie de requêtes

- *Browser*
- *Drill down*
- *Drill up*
- *Drill across*
- *Pivot*

La plupart des requêtes définies sur un schéma dimensionnel suivent un patron. Un ou plusieurs faits sont sélectionnés puis agrégés en fonction des attributs des dimensions sélectionnées. Les valeurs des dimensions sont également utilisées pour limiter la portée de la requête, en servant de base à des conditions de restrictions sur les données à extraire et à agréger.

### 2.1. Browser

*Browser* est une catégorie de requête qui consiste à explorer les données d'une dimension.

Les résultats de ces requêtes apparaissent comme des données de référence pour des rapports ou des résultats intermédiaires d'une requête sur la relation factuelle. Ce type de requêtes peut être envisagé lors de la prise de décisions sur la manière de filtrer le résultat ou de regrouper les attributs en dimensions.

### SQL Query

```
SELECT DISTINCT
    product.category
FROM
    product
ORDER By
    product.category
```

### Query Results

```
CATEGORY
=====
.
.
.
Fasteners
Folders
Packaging
Pens
Measurement
Notebooks
Storage
.
.
.
```

### SQL Query

```
SELECT DISTINCT
    product.category,
    product.product
FROM
    product
WHERE
    product.category = "Packaging"
ORDER BY
    product.product
```

### Query Results

CATEGORY	PRODUCT
=====	=====
Packaging	Box - Large
Packaging	Box - Medium
Packaging	Box - Small
Packaging	Clasp Letter
Packaging	Envelope #10
Packaging	Envelope Bubble

**Figure 1-7** Browse queries and their results

*Figure 1. Exploration de la relation Produit [Adamson2010a]*

## 2.2. Drill up and down

- *Drill down* est une catégorie de requête qui consiste à **ajouter** une ou plusieurs dimensions de la même hiérarchie à une requête pour raffiner le résultat.
- *Drill up* est une catégorie de requête qui consiste à **retirer** une ou plusieurs dimensions de la même hiérarchie à une requête pour synthétiser le résultat.

La capacité à rendre compte des faits est principalement limitée par le niveau de détail des mesures qui sont stockées. S'il est possible d'agréger les lignes détaillées de la relation factuelle en fonction de n'importe quel ensemble de dimensions, il n'est pas possible de produire un niveau de détail inférieur.

### Exemple

Si une relation factuelle contient des totaux quotidiens, elle ne peut pas être utilisée pour examiner une commande spécifique.

### Exercices

- Calculer le taux de réussite des personnes étudiantes (drill-up)
  - par programme d'études

- par département
- par facultés
- Calculer le taux d'occupation des lits d'hôpital (drill-down)
  - par hôpital
  - par service
  - par département

*Exemple 1. Calcul de la vente totale [Adamson2010a]*

#### Hiérarchie de Produit

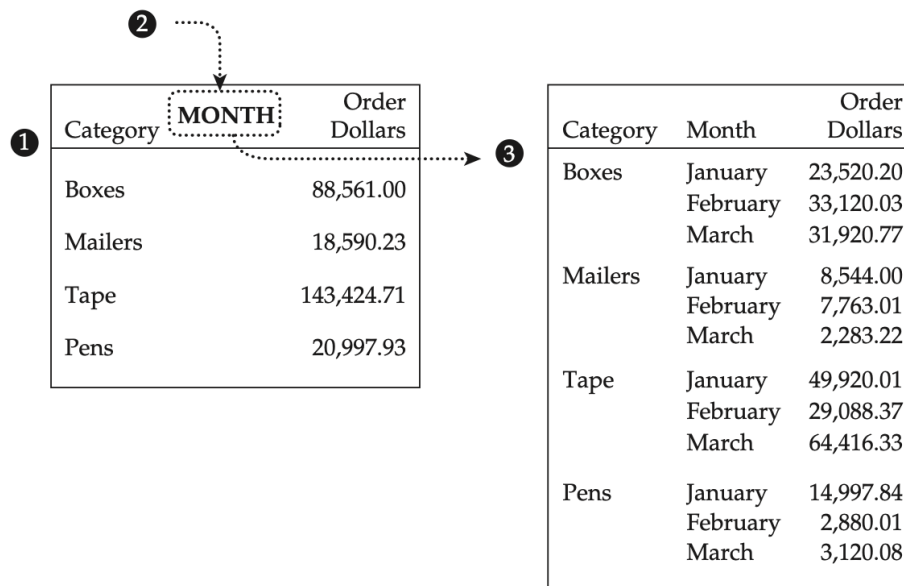
- Calculer la vente totale par produit
- Calculer la vente totale par catégorie

#### Hiérarchie de Moment

- Calculer la vente totale par mois
- Calculer la vente totale par année

#### Plusieurs dimensions

- par catégorie de produit
- par mois par catégorie de produit



**Figure 7-1** Adding dimensional detail

## 2.3. Drill across

*Drill across* est une catégorie de requête qui consiste à fusionner de requêtes provenant de plusieurs étoiles (plusieurs relations factuelles).

L'objectif de ces requêtes est de comparer deux ou plusieurs processus.

Pour pouvoir fusionner des faits, les attributs des dimensions communes les mêmes.

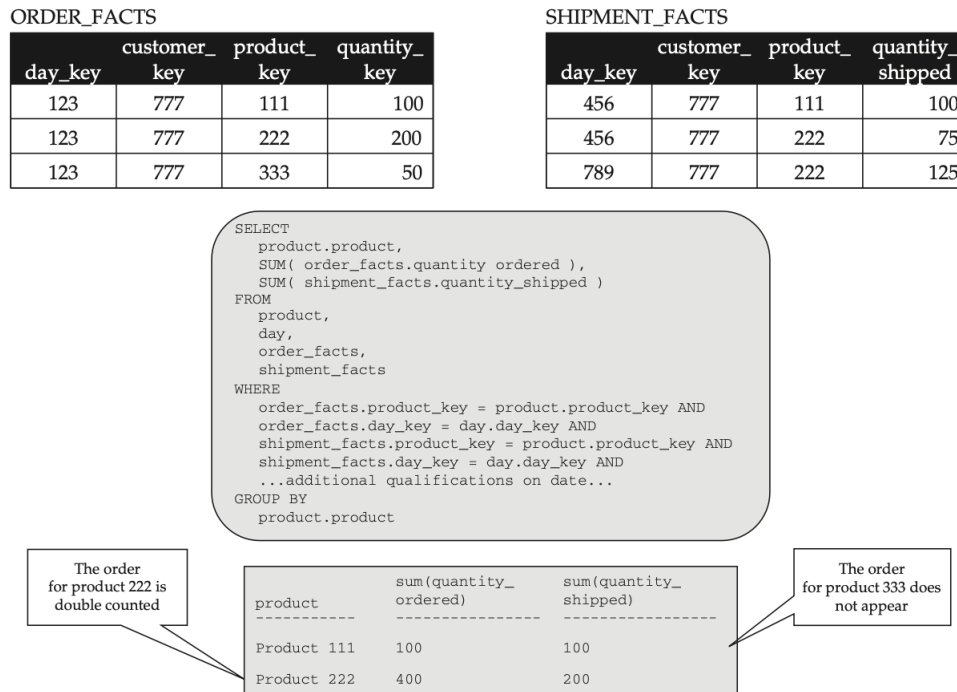
*Exemple*

- Évaluer la qualité des produits par produit et par mois (ventes et retours)
  - nombre de ventes

- nombre de retours
- montant total des ventes
- montant total des retours

### 2.3.1. Multiplicité des tables de faits

Lorsque l'on compare des faits provenant de différentes relations factuelles, il est important de ne pas les rassembler dans la même instruction SQL. En effet, les différences de synchronicité ou de granularité risquent d'entraîner des décomptes partiels ou multiples (typiquement un résultat incomplet ou un double décompte).

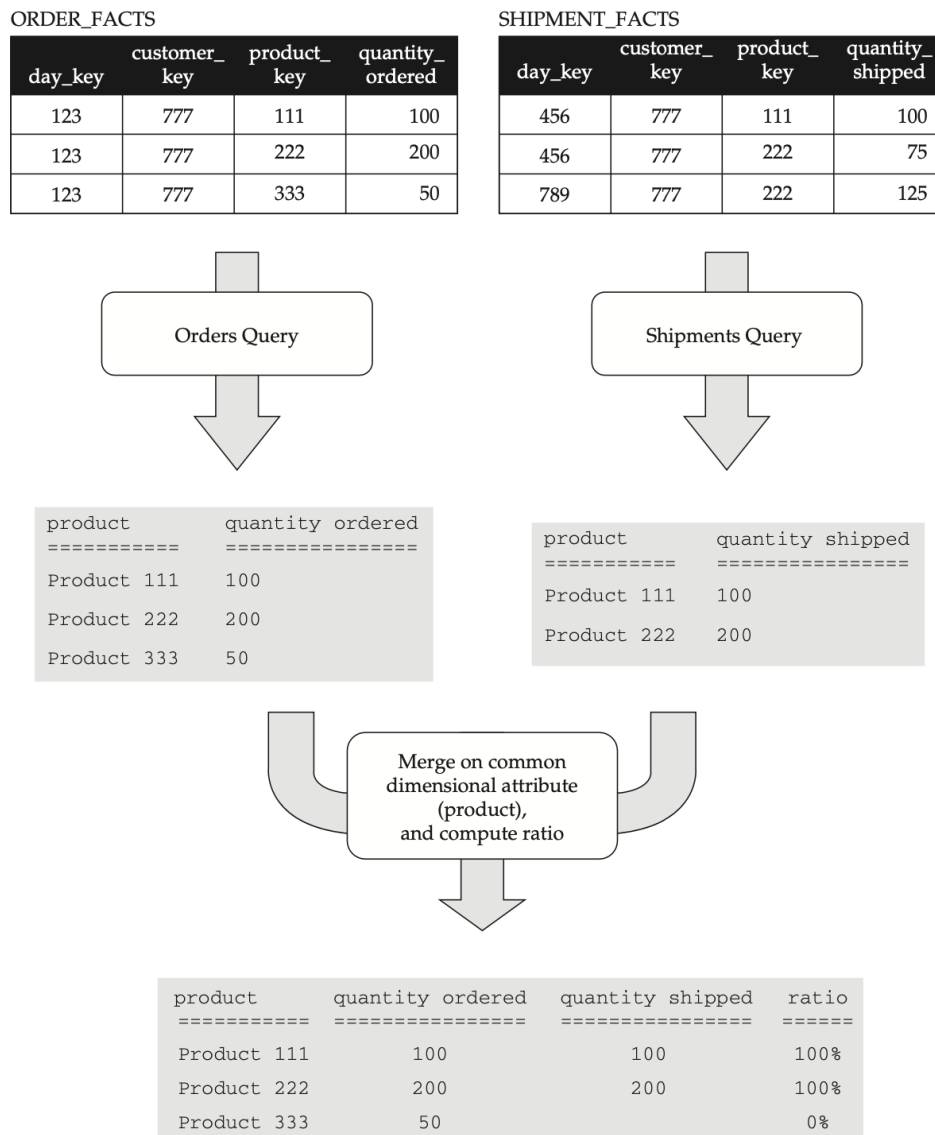


**Figure 4-10** Joining two fact tables leads to trouble

*Figure 2. Problèmes de jointure de deux relations factuelles [Adamson2010a]*

Les données doivent être recueillies en trois étapes qui consistent à passer d'une étoile à une autre :

1. calculer le résultat d'un processus à la fois selon la granularité choisie
2. joindre les résultats par paires en utilisant une jointure complète (*full outer join*),
3. ajouter des attributs calculés aux besoins.



**Figure 4-11** Drilling across orders and shipments

*Figure 3. Requête Drill across avec 2 relations factuelles [Adamson2010a]*



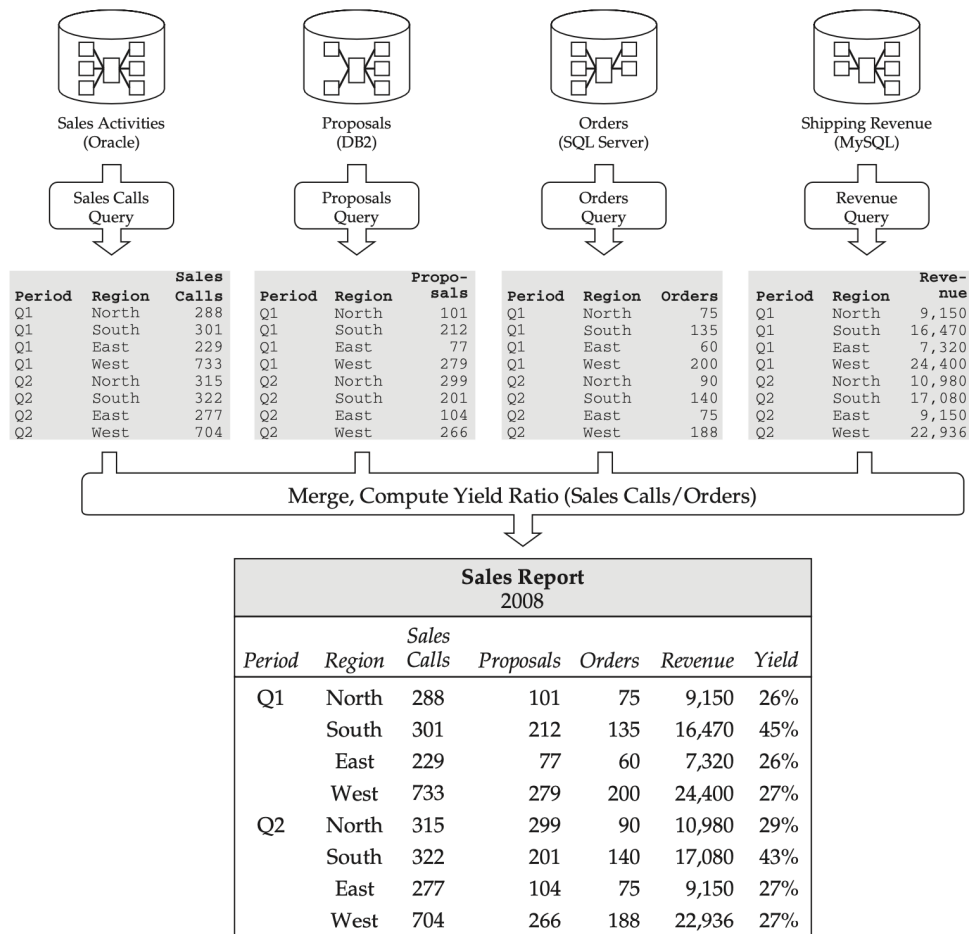
Notez les jointures décomposées en produits cartésiens (FROM) et restrictions (WHERE). Plusieurs membres de la communauté de pratique ont une telle phobie (injustifiée) des jointures qu'ils les simulent ainsi (pensant sans doute les éviter), complexifiant en fait le travail du SGBD (qui doit recomposer les jointures) afin de pouvoir les optimiser adéquatement. De même pour le réviseur humain qui doit les vérifier. On remarque par ailleurs que c'est une source fréquente d'erreur (restriction erronée ou absente).

En conclusion, l'écriture explicite des jointures est adéquate, en particulier du point de vue de la validité, de l'efficacité et de l'efficience.

### 2.3.2. Multiplicité des sources

Notez que les schémas dimensionnels peuvent être dans différents SGBD et sur différents sites. Cependant, il faut s'assurer que les dimensions d'intérêt ont la structure dans tous les schémas.

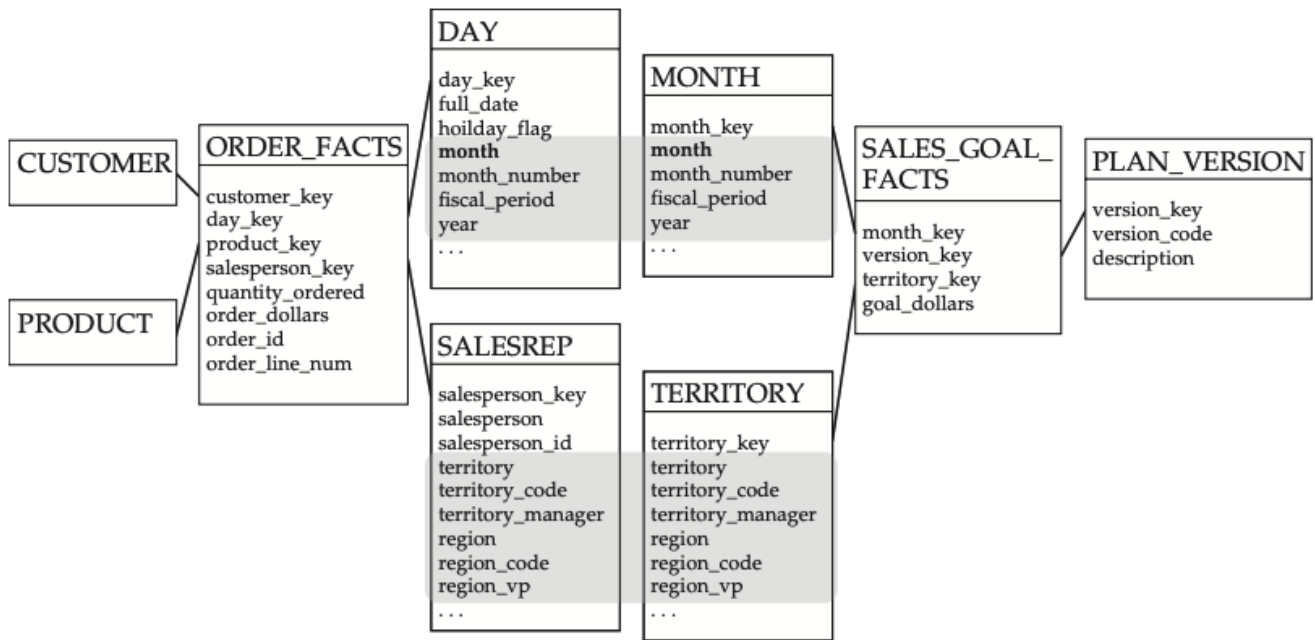




**Figure 4-13** Drilling across four fact tables

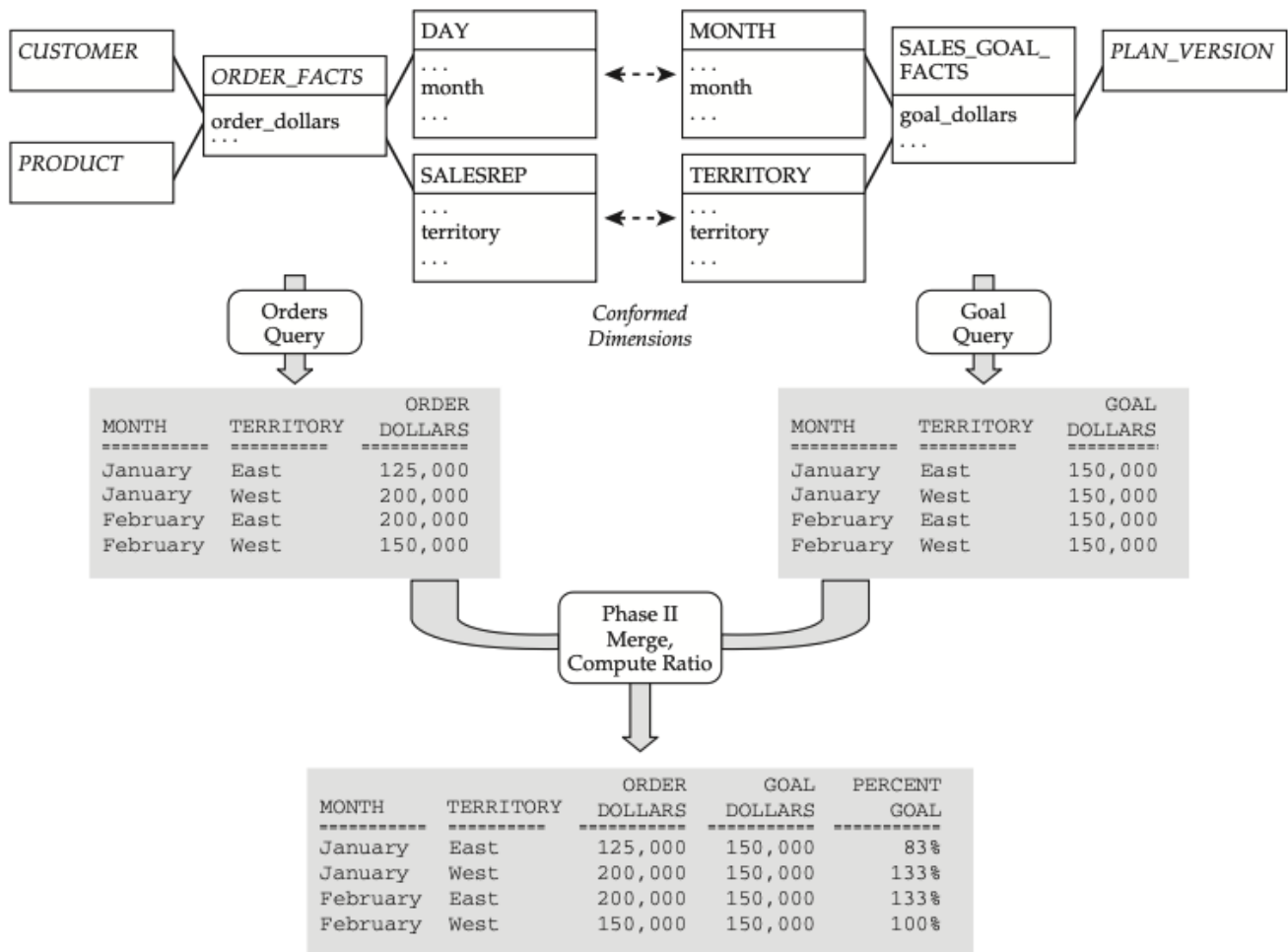
*Figure 4. Requête Drill across avec 4 relations factuelles [Adamson2010a]*

Enfin, une requête *drill across* peut être appliquée à une seule étoile pour produire des comparaisons sur le même processus. Par exemple, pour comparer les ventes par région de cette année par rapport à l'année précédente avec un ratio indiquant le pourcentage d'augmentation. Cette requête peut être créée en interrogeant les commandes par région pour l'année en cours, puis pour l'année précédente. Les deux séries de résultats peuvent ensuite être jointes sur les valeurs de la région pour calculer le ratio (année en cours/année précédente).



**Figure 5-3** These stars do not share common dimension tables but do share common dimension attributes

*Figure 5. Dimensions conformes [Adamson2010a]*



**Figure 5-4** Drilling across order\_facts and sales\_goal\_facts

*Figure 6. Fusion des dimensions conformes [Adamson2010a]*

Pour comparer des faits de deux processus, les dimensions doivent avoir la même structure et le même contenu (elles doivent être conformes).



On constate aisément qu'il est préférable de traiter la multiplicité des sources au moment de l'alimentation.

On s'assure ainsi que le travail est fait, bien fait et fait une seule fois, simplifiant ainsi l'expression de toutes les requêtes nécessitant lesdites sources.

Voici quelques (mauvaises) raisons pour lesquelles ce traitement n'est pas réalisé au bon moment (alimentation) ou au bon endroit (*staging region*):


- Qui va payer pour ce travail préparatoire? Que ceux qui en ont besoin le fassent!
  - Pour quelle raison fait-on un entrepôt de données déjà?
- La conciliation des sources dont j'ai besoin est différente de celle réalisée lors de l'alimentation
  - Ce ne peut être si on respecte la sémantique et les prédicats des attributs.

## 2.4. Pivot

- *Pivot* est une catégorie de requête qui consiste *reformater* le résultat d'une requête en transformant les lignes en colonnes ou vice-versa. Le résultat de ces requêtes permet de faciliter la présentation des

données selon un format de rapport ou pour les algorithmes d'intelligence artificielle.

ACCOUNT	TRANSACTION_TYPE	SUM (AMOUNT)
=====	=====	=====
01-2211	Credits	20,301.00
01-2211	Debits	- 17,691.30
07-4499	Credits	1,221.23
07-4499	Debits	- 2,220.01
.		
.		
.		



ACCOUNT	DEBITS	CREDITS
=====	=====	=====
01-2211	17,691.30	20,301.00
07-4499	2,220.01	1,221.23
.		
.		
.		

Figure 7. Requête pivoting [Adamson2010a]

### 3. Schémas dérivés

La base de données analytique sert à répondre à diverses questions pour évaluer l'état d'une organisation afin de permettre aux gestionnaires de prendre des décisions éclairées.

L'objectif principal qui a amené à la création du modèle dimensionnel est la facilité d'écrire des requêtes et l'optimisation des performances.

- Restructurer les données
- Synthétiser les données

Les schémas dérivés offrent une structure adaptée à un besoin d'analyse spécifique.

#### 3.1. Fusion des faits (*merged fact*)

La comparaison d'évènements provenant de plusieurs relations factuelles. (élimine plusieurs étapes d'une requête drill-across).

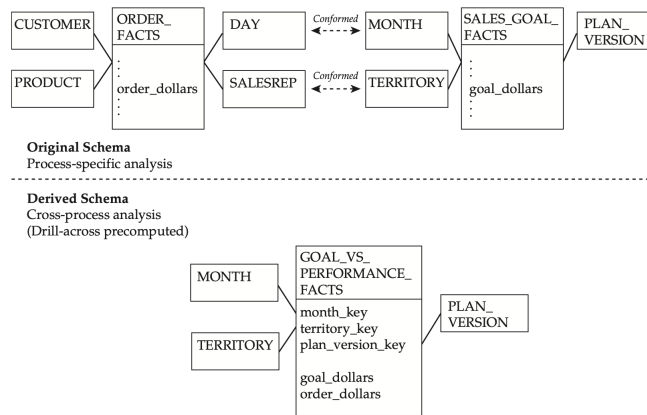


Figure 14-1 A merged fact table compares plans to actuals

Figure 8. Fusion des faits [Adamson2010a]

### 3.2. Pivotement des faits (*Pivoted fact*)

La transposition des lignes en colonnes pour certains types de rapports.

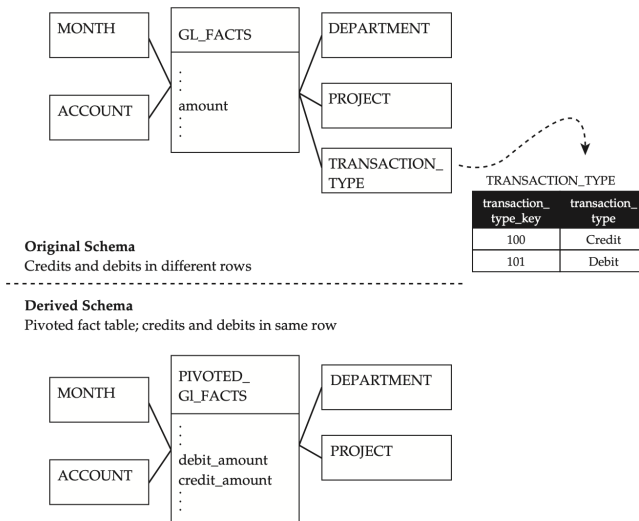


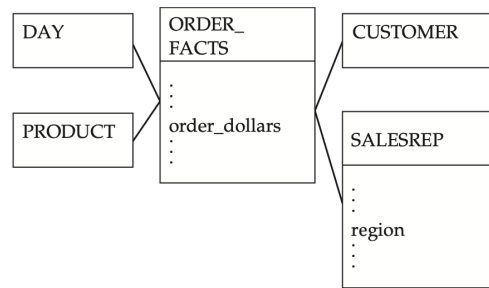
Figure 14-2 A pivoted fact table

Figure 9. Pivoter des faits [Adamson2010a]

### 3.3. Partitionnement des faits (*Sliced fact*) :

La sélection d'un sous-ensemble des faits selon une ou plusieurs dimensions.

Exemple, partitionné par magasin, région, par département.

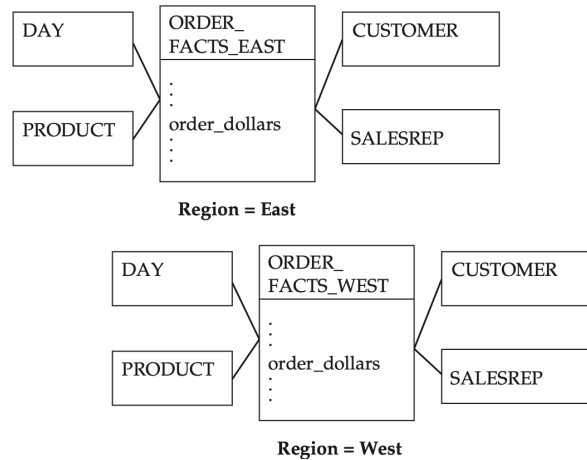


#### Original Schema

All regions

#### Derived Schema

Region-specific stars



**Figure 14-3** Sliced fact tables

*Figure 10. Partitionner les faits [Adamson2010a]*

### 3.4. Aggrégation des faits

Synthétiser des mesures selon une ou plusieurs dimensions.

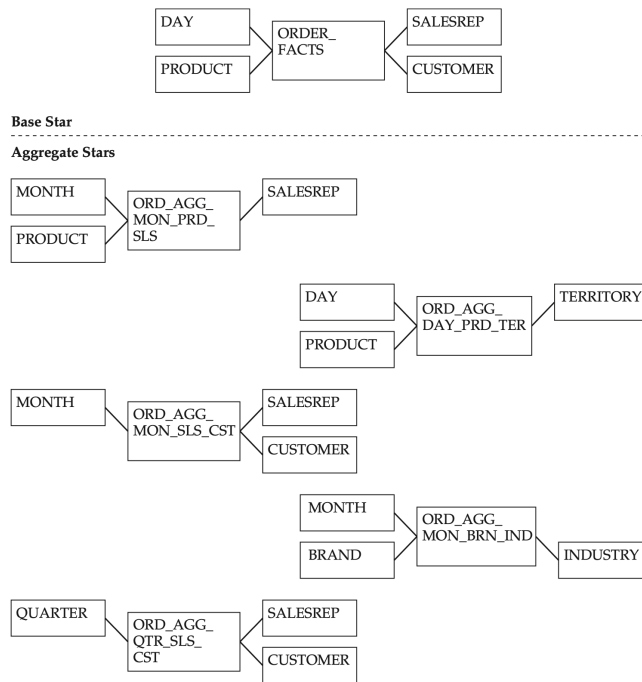


Figure 15-2 A base star and several aggregates

Figure 11. Aggréger les faits [Adamson2010a]

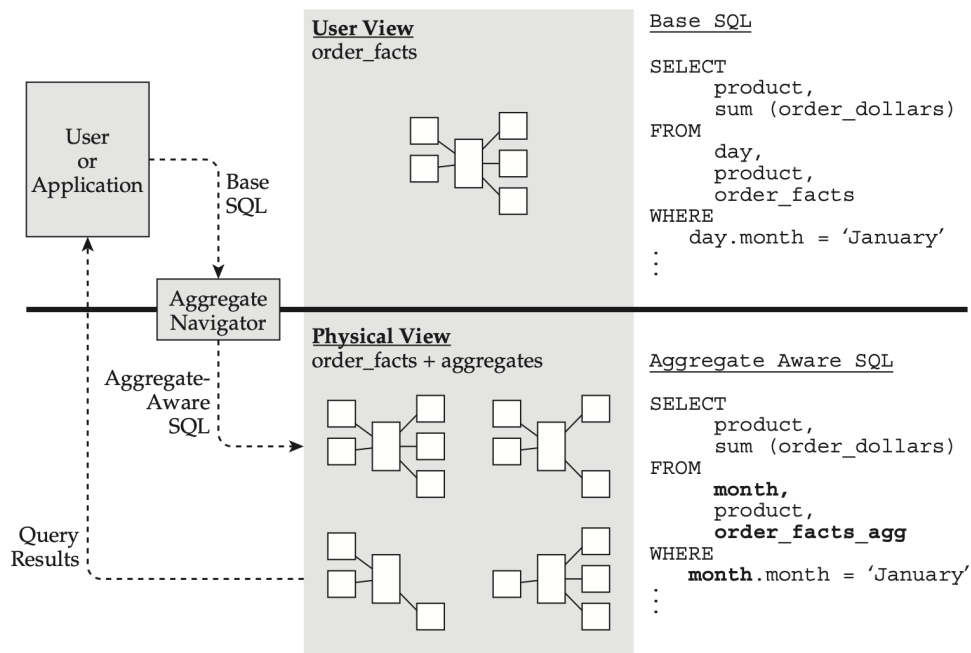


Figure 15-4 Aggregate navigator

Figure 12. SGBD - Navigation des étoiles agrégées [Adamson2010a]

### 3.5. Et les vues ?

Pourquoi n'utilise-t-on pas des vues pour représenter les tables des schémas dérivés ?

La raison invoquée est l'inefficacité (supposée) des vues lorsqu'elles sont utilisées à plusieurs reprises (sans changements aux données des tables sur lesquelles elles sont construites).

Cet argument ne prend pas en compte la possibilité de « matérialiser » les vues, de les « rematérialiser » au

besoin et même d'automatiser la « rematérialisation » au besoin.

Il est vrai que ces possibilités

- sont apparues tardivement (au cours des années 1990) et qu'elles ne sont pas connues de tous ;
- ne sont pas disponibles dans certains jouets utilisés en lieu et place de SGBD.

:-)

## 4. Règles de pratique

- Décrire rigoureusement le besoin d'analyse pour pouvoir choisir la catégorie de requêtes qui convient le mieux.
- N'essayez jamais de joindre deux relations factuelles directement ou par l'intermédiaire d'une dimension commune. Cela peut produire des résultats inexacts.
- Si vous utilisez des logiciels de construction de requêtes, la définition des catégories de requêtes peu variée. Lisez attentivement la documentation avant de construire votre requête.

## A. Exemple de code SQL

*Code SQL pour la requête Drill across*

```
SELECT
  COALESCE (orders_query.product, shipments_query.product),
  orders_query.quantity_ordered,
  shipments_query.quantity_shipped,
  orders_query.quantity_ordered / shipments_query.quantity_shipped as "Ratio"
FROM (
  SELECT
    product.product,
    SUM (order_facts.quantity_ordered) as quantity_ordered
  FROM day, product, order_facts
  WHERE
    ... joins and constraints on date ...
  ) orders_query
FULL OUTER JOIN
  ( SELECT
    product.product,
    SUM (shipment_facts.quantity_shipped) as quantity_shipped
  FROM
    day,
    product,
    shipment_facts
  WHERE
    ... joins and constraints on date ...
  ) shipments_query
ON
  orders_query.product = shipments_query.product
```

## Références

**[Adamson2010a]**

Christopher ADAMSON;

*The complete reference star schema*;

McGraw-Hill, New York (NY, US), 2010;

ISBN 978-0-07-174432-4.



Produit le 2025-11-21 12:12:15 UTC



**Université de Sherbrooke**