

Data Warehouse

Professeur:

Mr ABDELILAH BOUSLAMA

bo.abdelilah@gmail.com

EMSI –Feb 2025

Plan

- **Vue d'ensemble sur les entrepôts de données**
- **Architecture des entrepôts de données**
- **Modélisation dimensionnelle**
- **Option de modélisation**
- **Comparaison entre les différentes approches**
- **Conclusion**
- **Bibliographie**
- **Questions**



Vue d'ensemble

Un **système d'information décisionnel** est un ensemble de données organisées de façon spécifique, appropriées à la prise de décision.

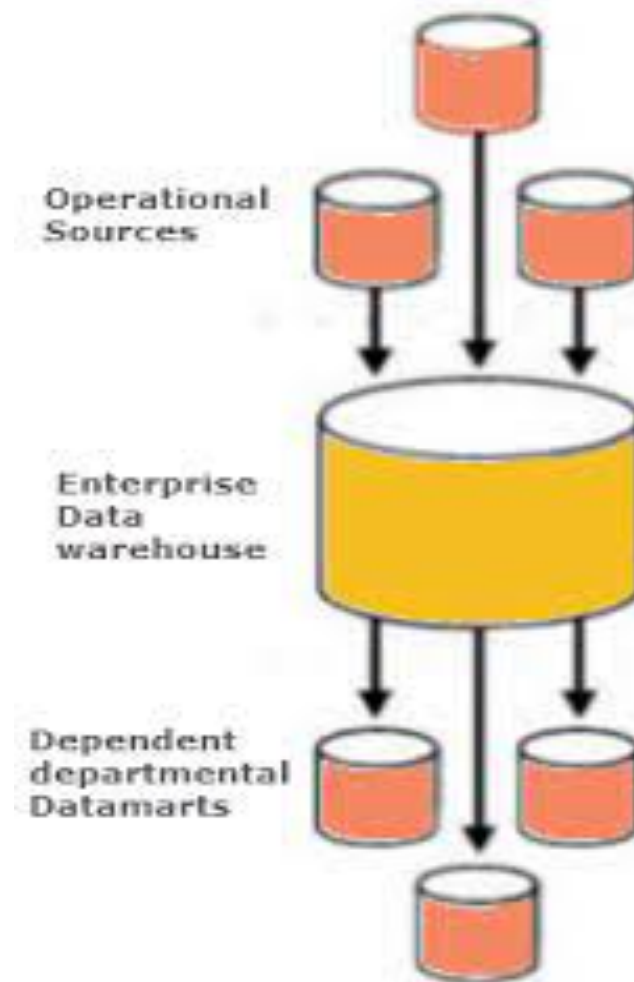
Connaître l'environnement dans lequel on évolue

Finalité d'un système décisionnel : **pilotage de l'entreprise**

– Outils :

☐ ☐ . **Data warehouse**

☐ ☐ . **OLAP**



Vue d'ensemble



- **Entrepôt de données :**

D'après BILL Inmon : “Un ED est une collection de données **thématiques, intégrées, non volatiles** et **historisées**, organisées pour la prise de décision.”

Thématiques : thèmes par activités majeures ;

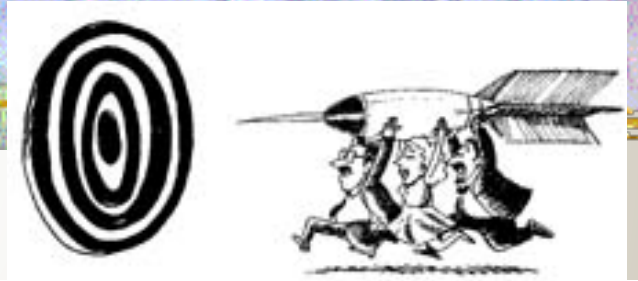
Intégrées : divers sources de données ;

Non volatiles : ne pas supprimer les données du DW ;

Historisées : trace des données, suivre l'évolution des indicateurs.

Pb de volumétrie, de stockage, d'accès.

Objectifs de l'Entrepôt de Données



- [?] Accessibilité des informations**
facile à comprendre donc à utiliser
- [?] Information cohérente**
idempotence avec le temps
incomplétude signalée
- [?] Manipulation des mesures de l'activité**
combinaison et séparation (tranches et dés)
- [?] Ensemble de données et de moyens**
requêtes, analyse, présentation, ...
- [?] Publication de données déjà servies**

Vue d'ensemble



Qu'est ce que l'Entreposage des données ?

- ☐ Conception
- ☐ Construction
- ☐ Administration
- ☐ Restitution

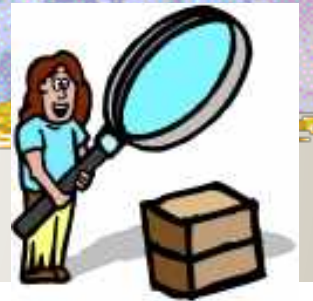


Conception

Il s'agit de définir la finalité du ED :

- ♠ Piloter quelle activité de l'entreprise ;
- ♠ Déterminer et recenser les données à entreposer
- ♠ Définir les aspects techniques de la réalisation ;
- ♠ modèle de données ;
- ♠ démarches d'alimentation ;
- ♠ stratégies d'administration ;
- ♠ définition des espaces d'analyse ;
- ♠ mode de restitution...





Construction (Travail technique.)



- ♠ Extraction des données des différentes BD de production (internes ou externes)
- ♠ Nettoyage des données, règles d'homogénéisation des données sous formes **de métadonnées**.
- ♠ Techniques d'alimentation :
 - ☐ * Chargement des données dans l'ED ;
 - ☐ * Fréquences de rafraîchissement :
 - ☐ par applications d' interfaces entre les sources de données et l'ED ;
 - ☐ par serveurs de réplication du SGBD ou par outils spécialisés.
 - ☐ ☐

Vue d'ensemble

☐ Administration



Elle est constituée de plusieurs tâches pour assurer :

- ♠ la qualité et la pérennité des données aux différents applicatifs ;
- ♠ la maintenance ;
- ♠ la gestion de configuration ;
- ♠ les mises à jour ;
- ♠ l'organisation, l'optimisation du SI ;
- ♠ la mise en sécurité du SI.



Vue d'ensemble



Restitution

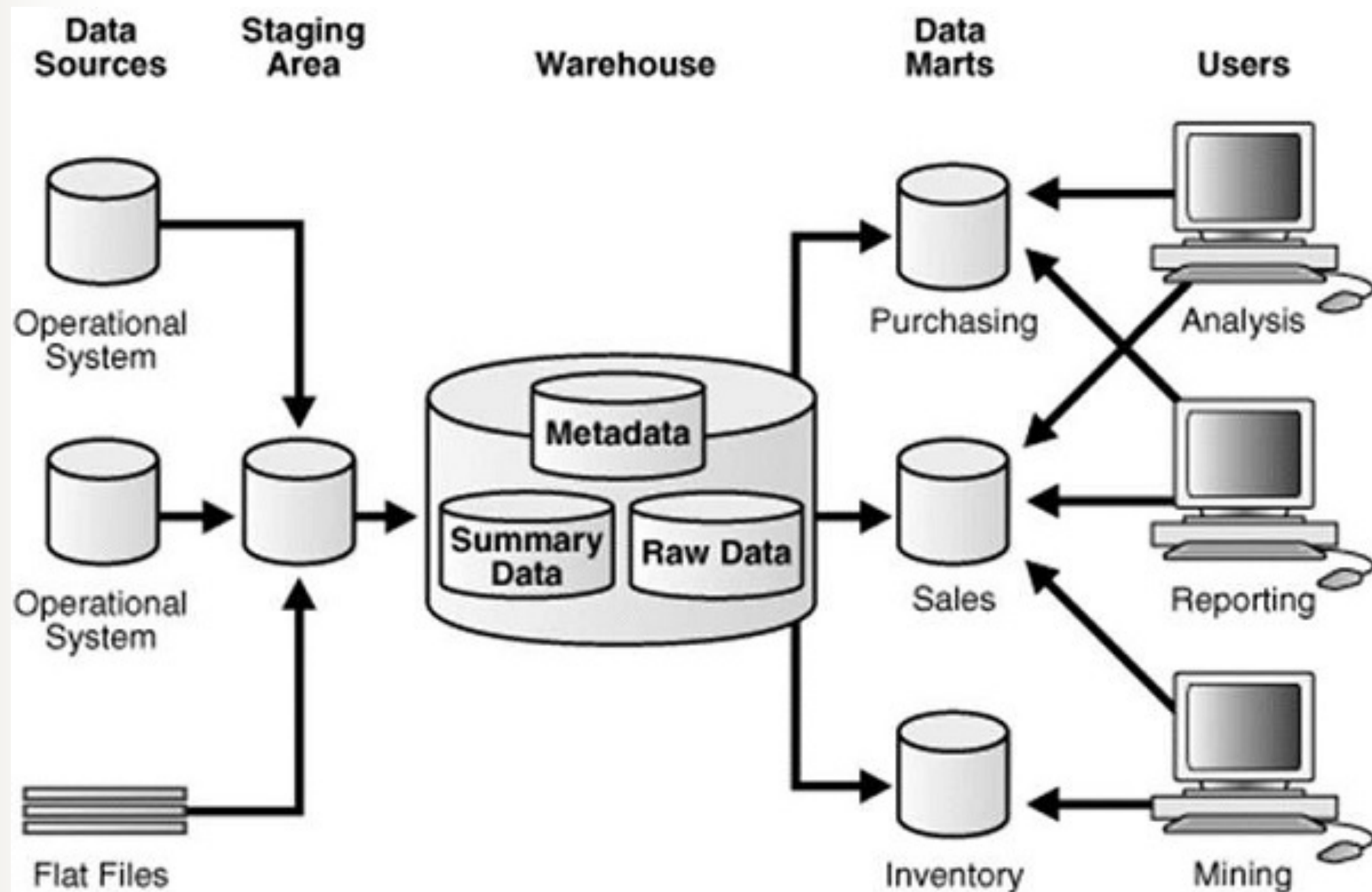
- ♠ C'est le but du processus d'entreposage des données.
- ♠ Elle conditionne le choix de l'architecture de l'ED et de sa construction.
- ♠ Elle doit permettre toutes les analyses nécessaires pour la construction des indicateurs recherchés.



Vue d'ensemble



Les Magasins de données sont des entrepôts de données avec une portée plus étroite et spécifique d'utilisation.



Vue d'ensemble



- **Entrepôt de données Versus base de données relationnelle.**
- **OLTP : "On-Line Transactional Processing "**
 - [?] gérer les importants volumes d'informations contenus dans leurs systèmes opérationnels**
- **OLAP : "On-Line Analytical Processing"**
 - [?] répondent aux besoins spécifiques d'analyse d'informations qui doit se faire de manière interactive et rapide, pour des données quelconques et historisées**

Vue d'ensemble

Entrepôt de données Versus base de données relationnelle.



Critère	OLTP	OLAP
But	Contrôler et exécuter les tâches quotidiennes et fondamentales de l'entreprise	Assister dans la planification, la résolution de problème et la prise de décision
Types de données	Données opérationnelles (transactions)	Données historiques consolidées
Sources de données	BD transactionnelles	Entrepôts de données ou magasins de données
Ce que montre les données	Portrait instantané des processus d'affaires de l'entreprise	Vue multidimensionnelle de plusieurs activités d'affaires de l'entreprise
Insertions et mises-à-jour	Courtes requêtes d'insertion et de mise-à-jour lancées par les usagers finaux	Longs traitements en lot servant à rafraichir les données
Requêtes	Simple requêtes retournant quelques enregistrements (lignes) de la BD	Requêtes complexes impliquant souvent plusieurs tables et faisant l'agrégation de valeurs
Temps de réponses	Instantané	Quelques secondes à 1 minute max.



Modélisation multidimensionnelle

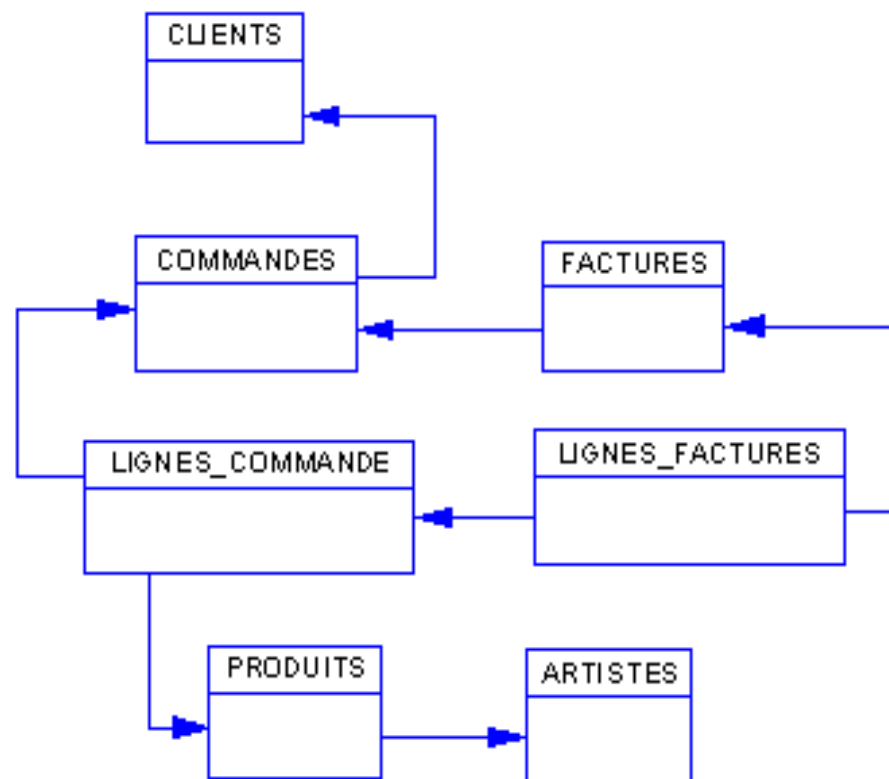
- Considère le sujet analysé comme un point dans un espace à plusieurs dimensions.
- Les données sont organisées de manière à mettre en évidence le sujet analysé et les différentes perspectives de l'analyse
- Objectif :
 - ❑ produire des structures de base de données
 - ❑ formuler des requêtes pour des utilisateurs finaux
 - ❑ maximiser l'efficacité des requêtes



Organisation des données

- ❑ Les applications conçues pour des opérations quotidiennes dans les BDs.
- ❑ applique en général à la création et à la mise à jour individuelle des informations.
- ❑ Les BD vont de quelques milliers de Mo à des Go.

OLTP (On-Line Transactional Processing)

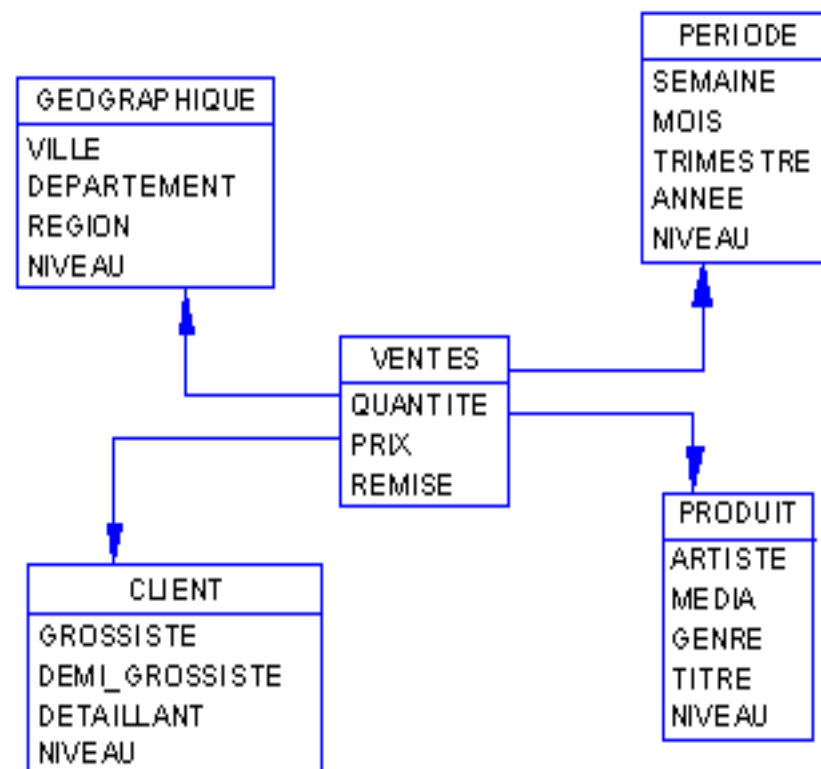




Organisation des données

- ❑ Les données sont historiées, résumées, consolidées.
- ❑ Les EDs contiennent des données sur une longue période de temps.
- ❑ Les EDs vont de centaines de Go à des To (Téra octets).

OLAP (On-Line Analytical Processing)

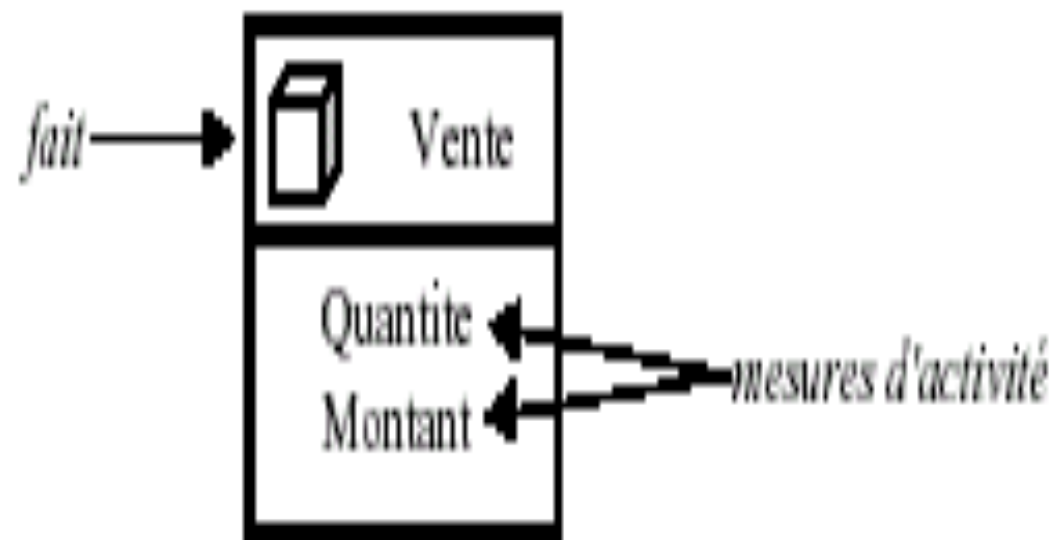


Concepts de fait et de dimension



Le fait:

- *Modélise le sujet de l'analyse.*
- *Il est formé de mesures correspondant aux informations de l'activité analysée.*



Concepts de fait et de dimension



Dimension:

- *Modélise une perspective de l'analyse.*
- *Elle se compose de paramètres correspondant aux informations faisant varier les mesures de l'activité..*

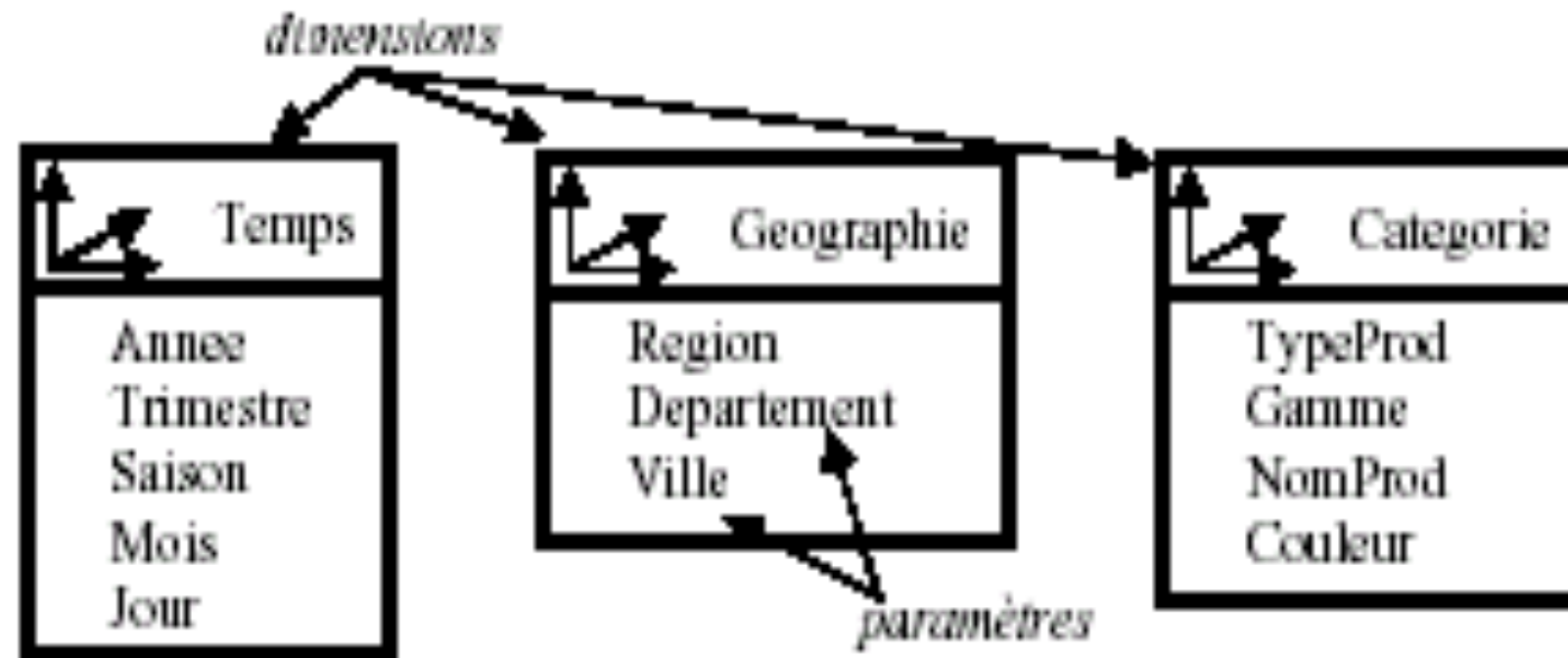
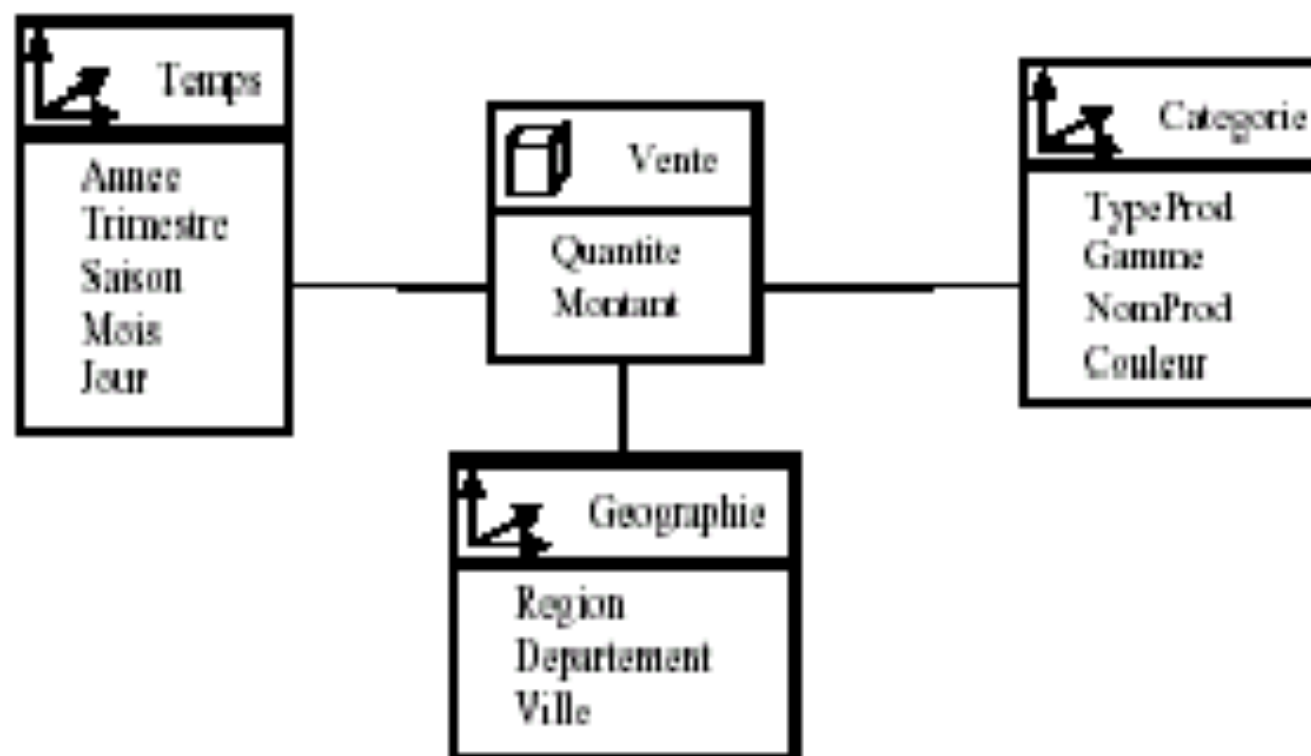


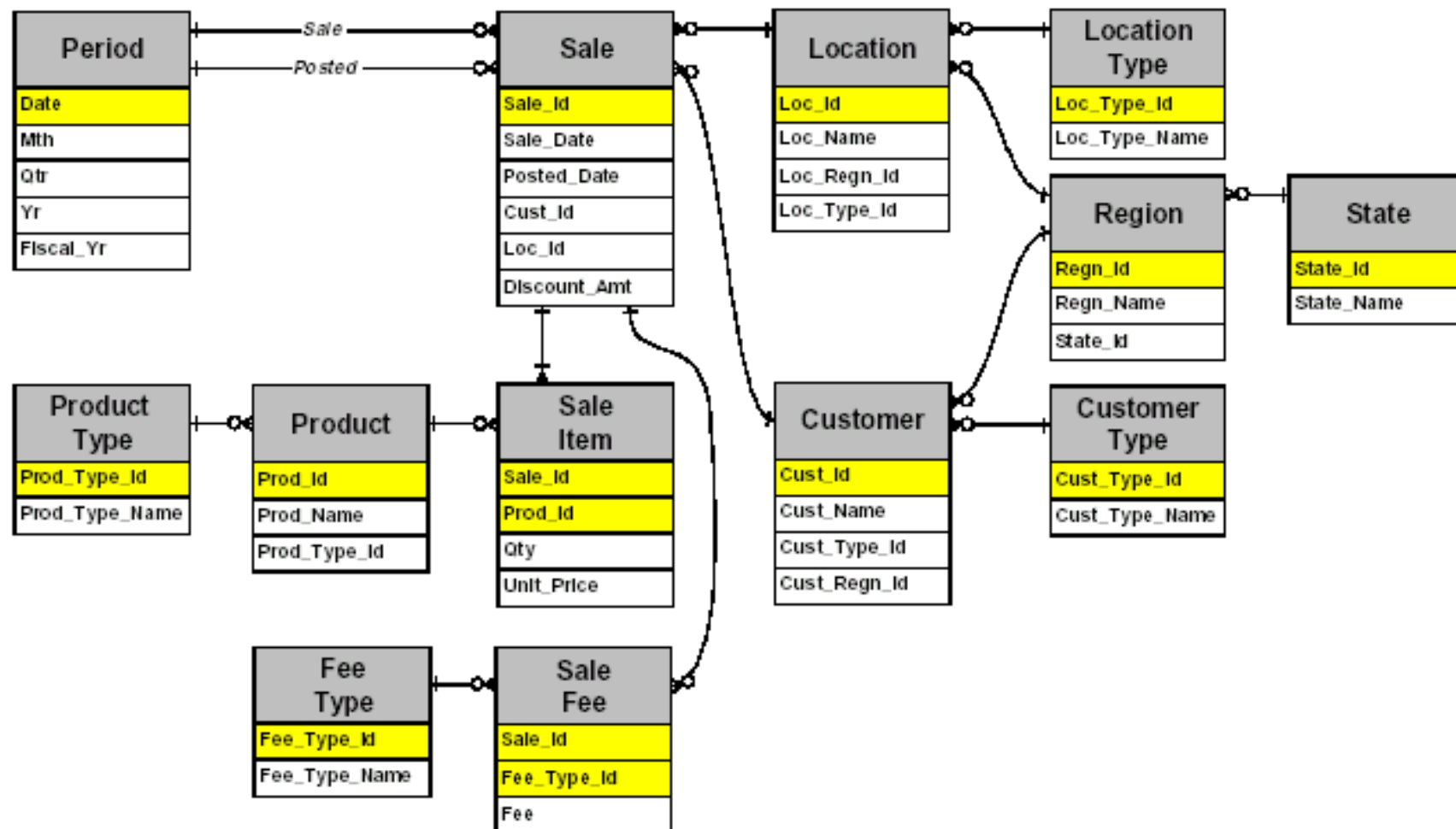
Schéma en étoile



Du modèle entité relation aux modèle dimensionnel



- Conception des magasins de données basés sur un modèle de données d'entreprise



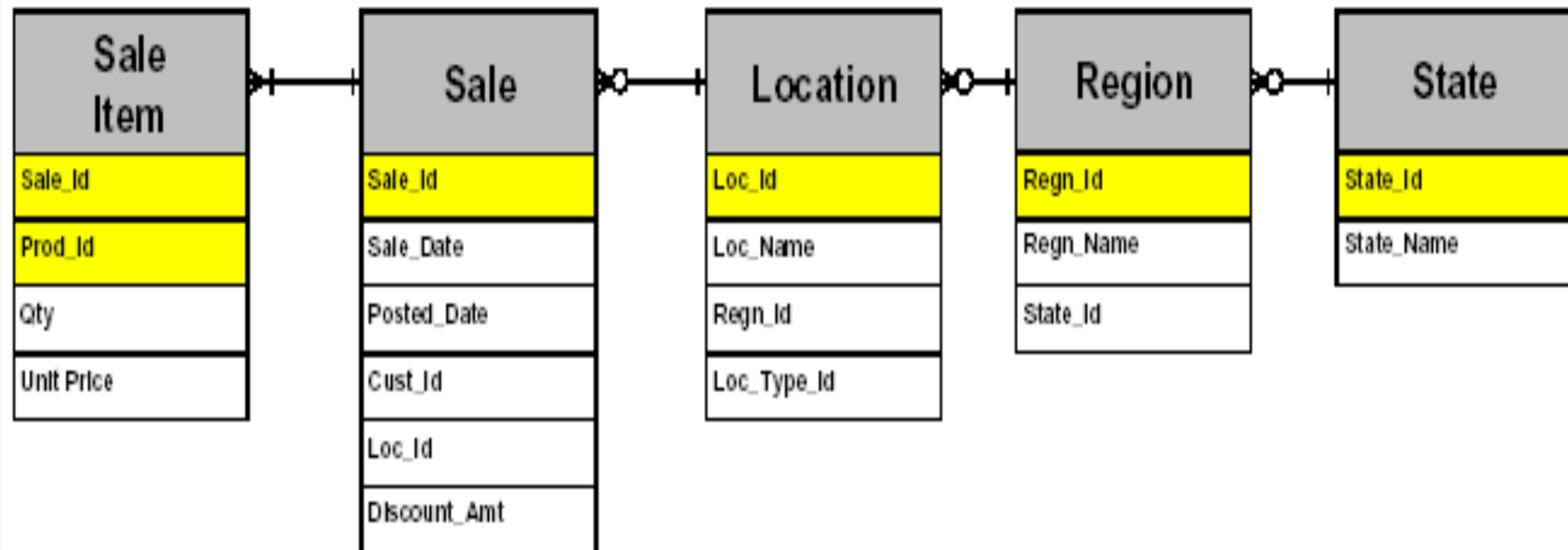
Du modèle entité relation aux modèle dimensionnel

Étapes de conception



Étape 2: Identifications hiérarchiques

Une hiérarchie organise les paramètres d'une dimension selon une relation "est_plus_fin" conformément à leur niveau de détail .



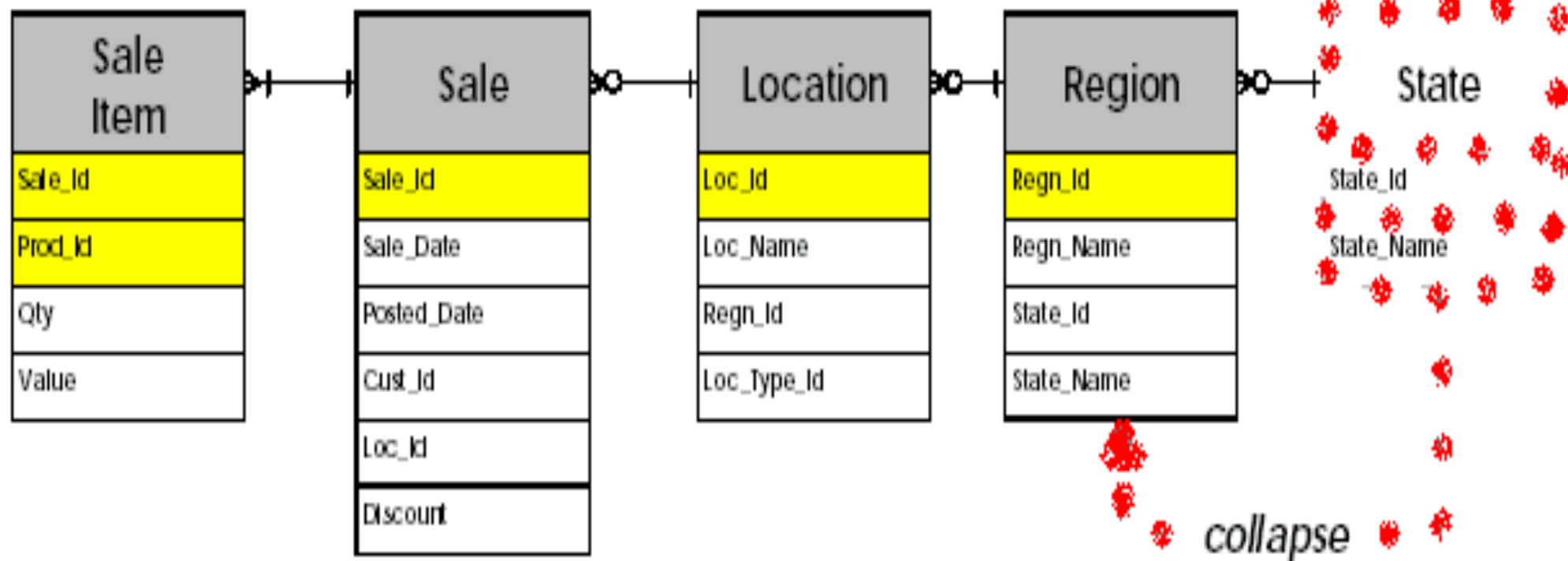
Du modèle entité relation aux modèle dimensionnel

Étapes de conception



Étape 3: Production du model dimensionnel

Opérateur 1 : La Hiérarchie d'annihilation (*Collapsing*). (Possibilité d'itérer)

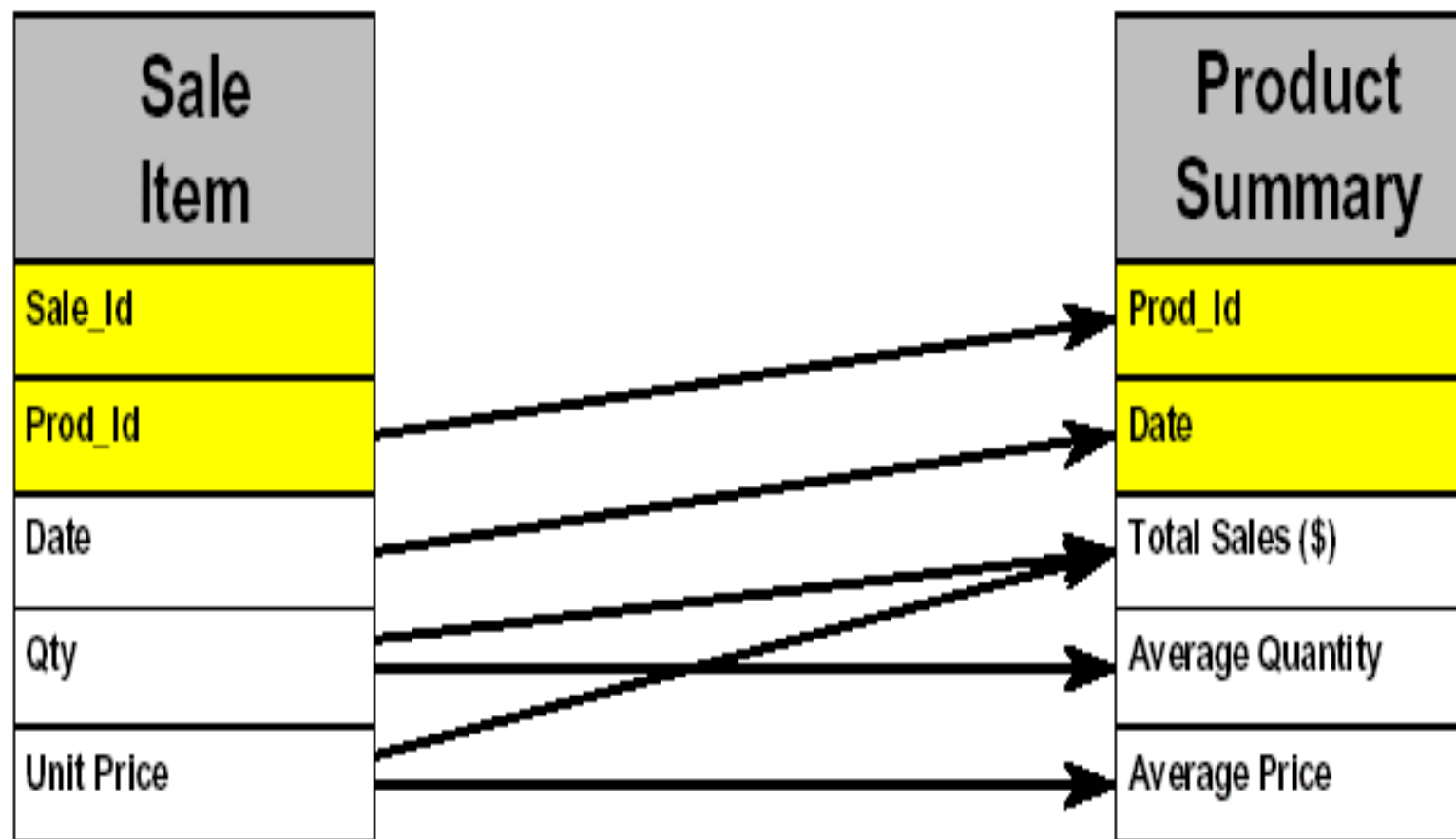


Du modèle entité relation aux modèle dimensionnel

Étapes de conception



Opérateur 2 : l'agrégation



Du modèle entité relation aux modèle dimensionnel

Étapes de conception



Étape 4: Évaluation et raffinement

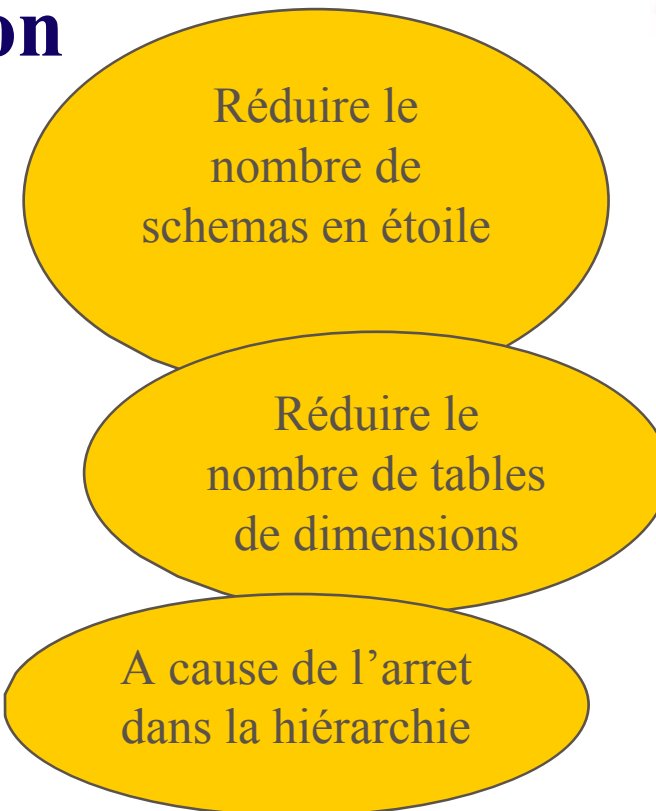
- ✓ *Union de tables de Fait*
- ✓ *Union de Tables de Dimension*
- ✓ *Relation plusieurs à beaucoup.*
- ✓ *Sous-types de Traitement*

- **Relation plusieurs à beaucoup.**



Solution:

- Ignorer l'entité d'intersection
- Créer une relation primaire



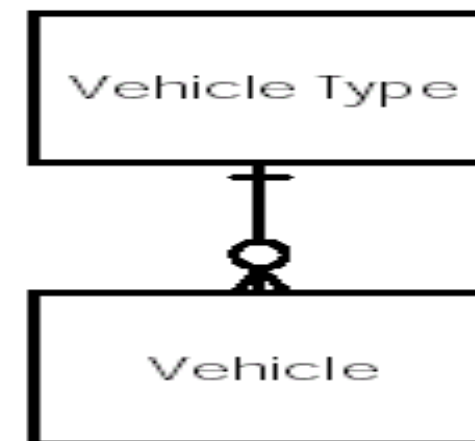
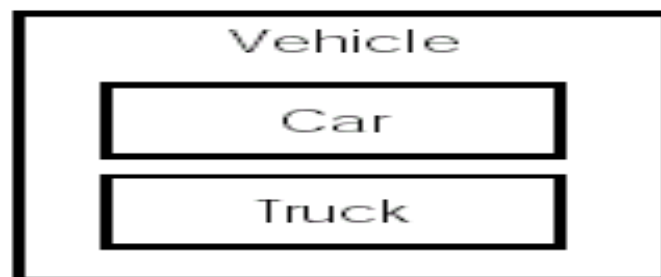
Du modèle entité relation aux modèle dimensionnel

Étapes de conception



Sous-types de Traitement

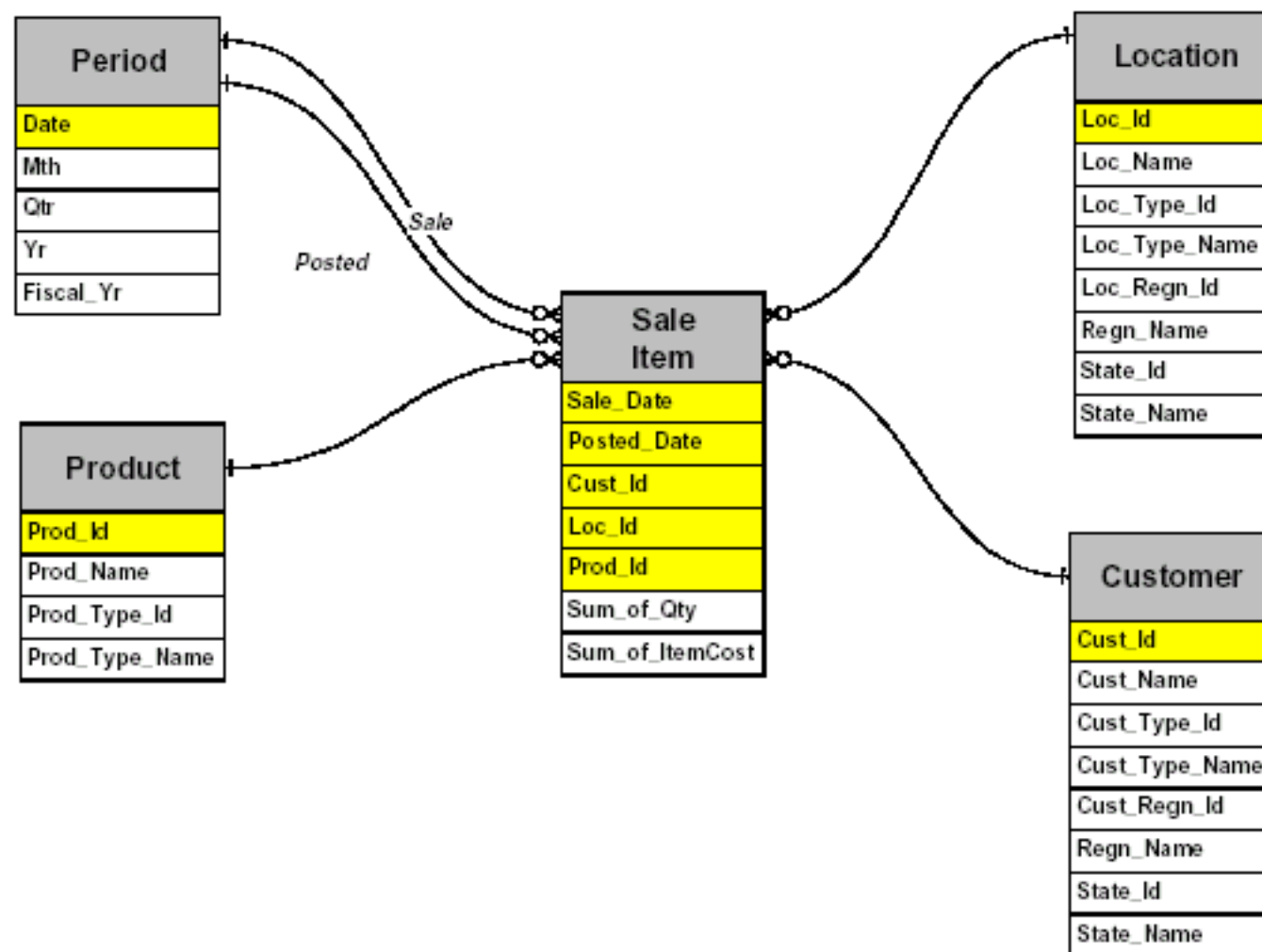
- Relation super type/super type: Conversion directe



Option de la modélisation



- Option 3: le Schéma en étoile



Option de la modélisation



Schéma de Constellation

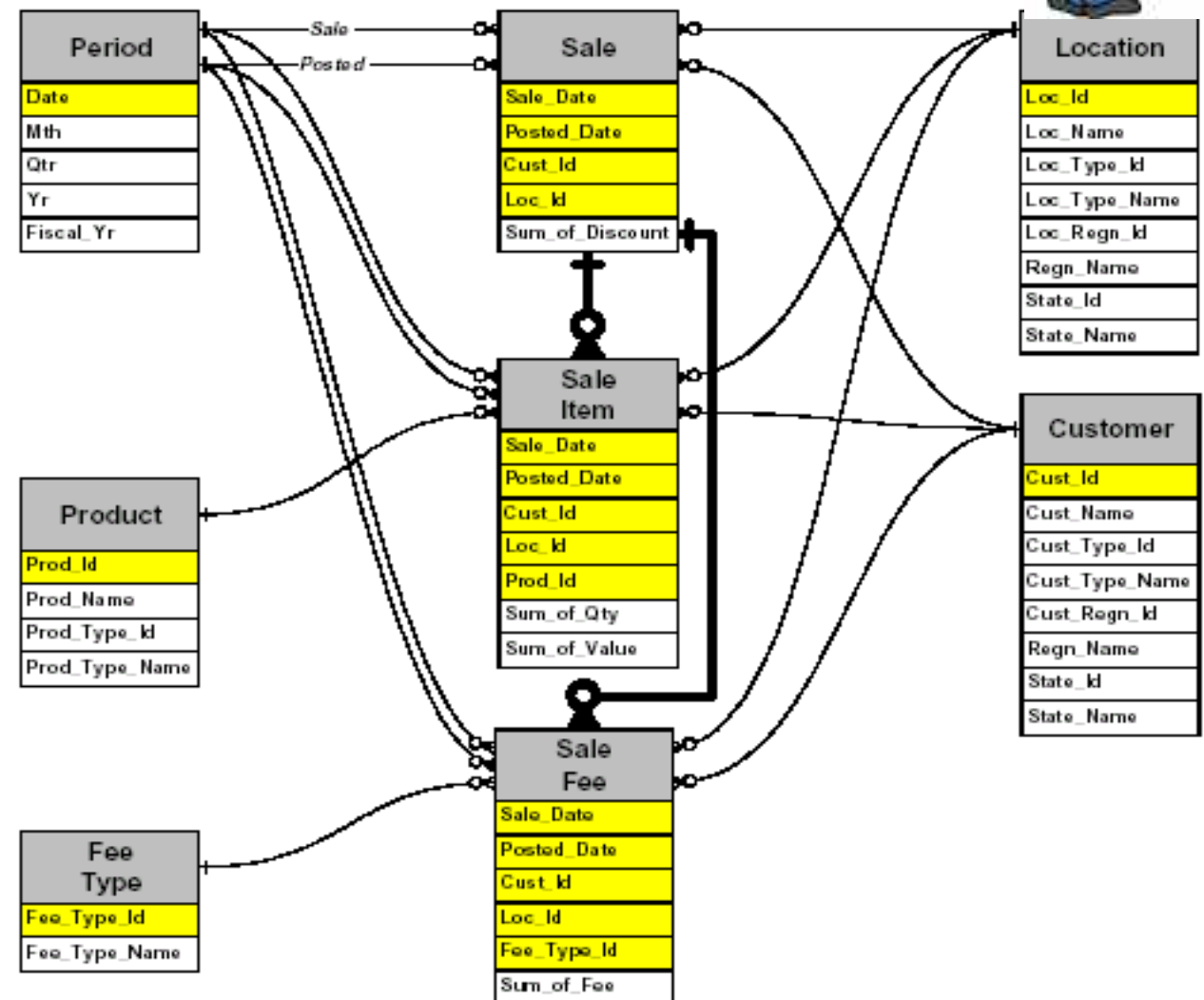


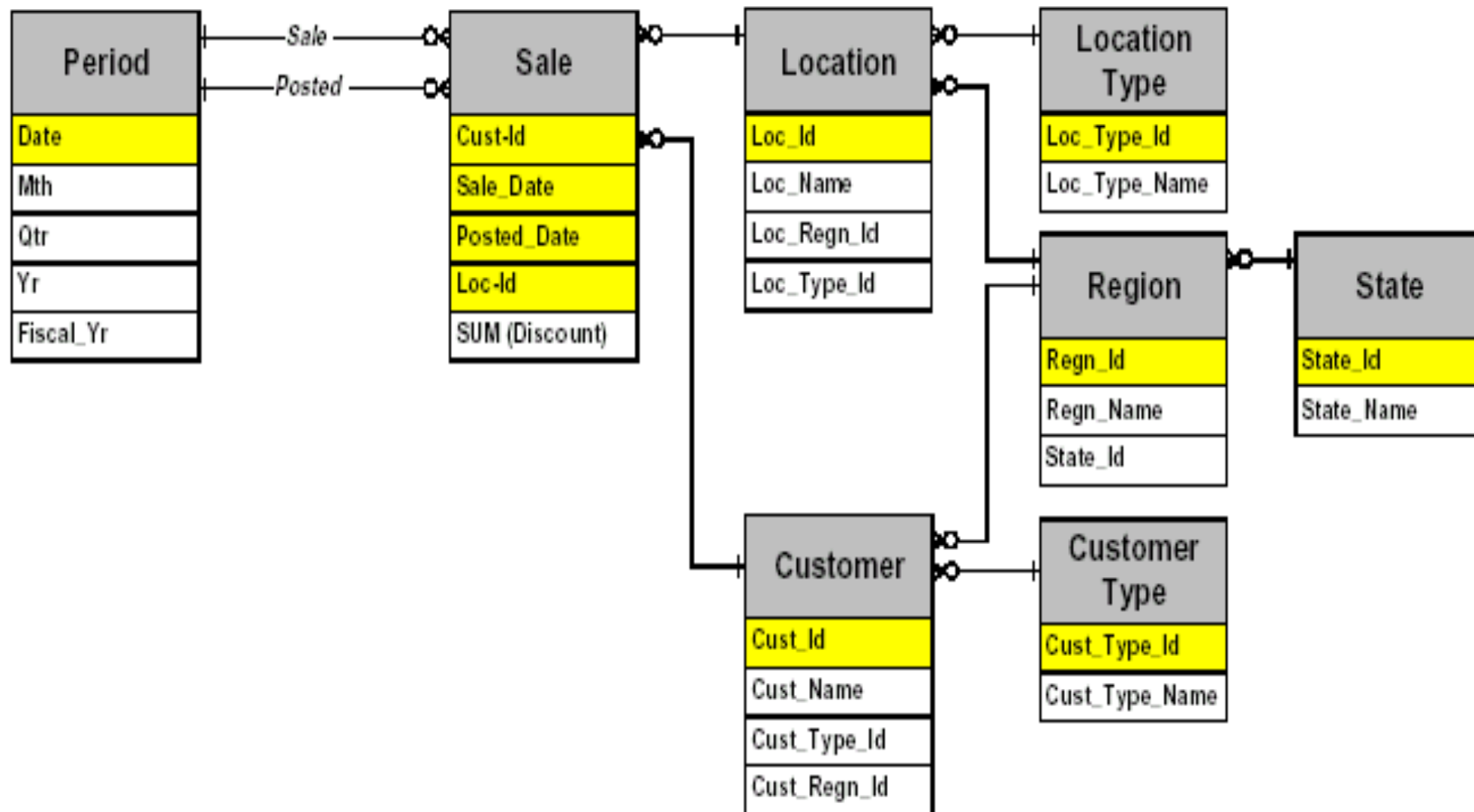
Schéma de Galaxie

Plus généralement, un jeu de schémas en étoile ou des constellations peuvent être combiné ensemble pour former une galaxie

Option de la modélisation



- **Option 3: le Schéma de Flocon de neige**



Comparaison entre les diverses méthodes



Modélisation dimensionnelle vs modélisation Entité-relation:

- ✓ *Présentation des données à l'utilisateur final d'une façon simple et intuitive*
- ✓ *considération particulière pour le support des requêtes de haut degrés de performance sur des tables de fait très grandes.*

Schéma en étoile:

- ✓ *L'utilisateur ne manipule pas des structures de dimension complexe (ER).*
- ✓ *Il peut utiliser le modèle dimensionnel sans avoir à apprendre comment interpréter correctement les schémas ER.*

Modélisation temporelle de données :

Problèmes de redondance car une même information est conservée plusieurs fois; pour résoudre ce problème, le concept de stockage du différentiel (delta) a été proposé



Conclusion



L'entrepôt de données est donc bien différent des bases de données de production car les besoins pour lesquels on veut le construire sont différents.

Il contient des informations historisées, globalement cohérentes, organisées selon les métiers de l'entreprise pour le processus de décision.

L'entrepôt n'est pas un produit ou un logiciel mais un environnement.

Il se bâtit et ne s'achète pas.

Les données sont puisées dans les bases de production, nettoyées, normalisées, puis intégrées.

Des métadonnées décrivent les informations dans cette nouvelle base pour lever toute ambiguïté quant à leur origine et leur signification.

Chaque approche de modélisation a sa place dans le contexte approprié

Le succès de telle ou telle approche de modélisation d'un entrepôt de données exige que les techniques soient appliquées dans le bon contexte

Questions ?



Merci !