

# Datawarehouse

# Plan

---

- Partie 1 : Introduction
  - ◆ 1. Objectifs
  - ◆ 2. Qu'est ce qu'un datawarehouse?
  - ◆ 3. Pourquoi ne pas réutiliser les BD?
- Partie 2 : Implémentation d'un datawarehouse
  - ◆ ROLAP
  - ◆ MOLAP
  - ◆ HOLAP
- Partie 3 : Cube
  - ◆ Hypercube: Mesures et Dimension
  - ◆ Opérateurs de manipulation
- Partie 4 : Modélisation en étoile et en flocon
- Partie 5 : Conception d'un datawarehouse

# Partie 1: Introduction

# 1- Objectifs

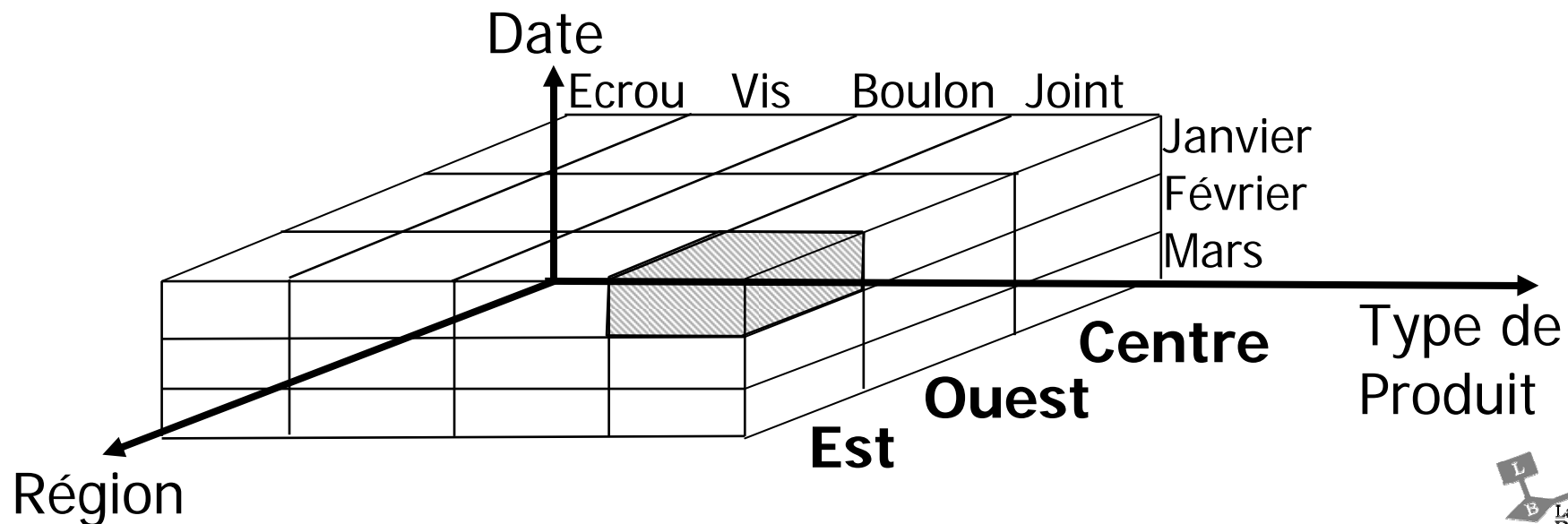
---

- Transformer un système d'information qui avait une vocation de production en un SI décisionnel
  - ◆ = Transformation des données de production en informations stratégiques
- Exemple de requêtes décisionnelles :
  - ◆ Catégorie socioprofessionnelle des meilleurs clients de chaque région
  - ◆ Evolution de la part de marché d'un produit particulier
  - ◆ Nombre d'employé de l'entreprise par classe d'âge, par sexe, par grade
  - ◆ Quel est le profil des employés les plus performants?

# Objectifs (2)

---

- Gestion et visualisation des données doit être rapide et intuitive
- -> visualisation multi-dimensionnelle des données:



# Objectifs (3)

---

- Pour cela, nécessaire de retrouver et d'analyser rapidement les données provenant de diverses sources
- DW offre une vision transversale des données de l'entreprise
  - ◆ intégration de différentes BDs
- Les données doivent être :
  - ◆ Extraites
  - ◆ Groupées ensembles et organisées
  - ◆ Corrélées
  - ◆ Transformées (résumé, agrégation)

## 2- Qu'est ce qu'un data Warehouse?

---

- Ensemble de données
  - ◆ destinées aux "décideurs"
  - ◆ souvent une copie des données de production
  - ◆ avec une valeur ajoutée (agrégation, historique)
  - ◆ intégrées
  - ◆ historisées
- Ensemble d'outils permettant
  - ◆ de regrouper les données
  - ◆ de nettoyer, d'intégrer les données, ...
  - ◆ de faire des requêtes, rapports, analyses
  - ◆ de faire du data mining
  - ◆ faire l'administration du warehouse

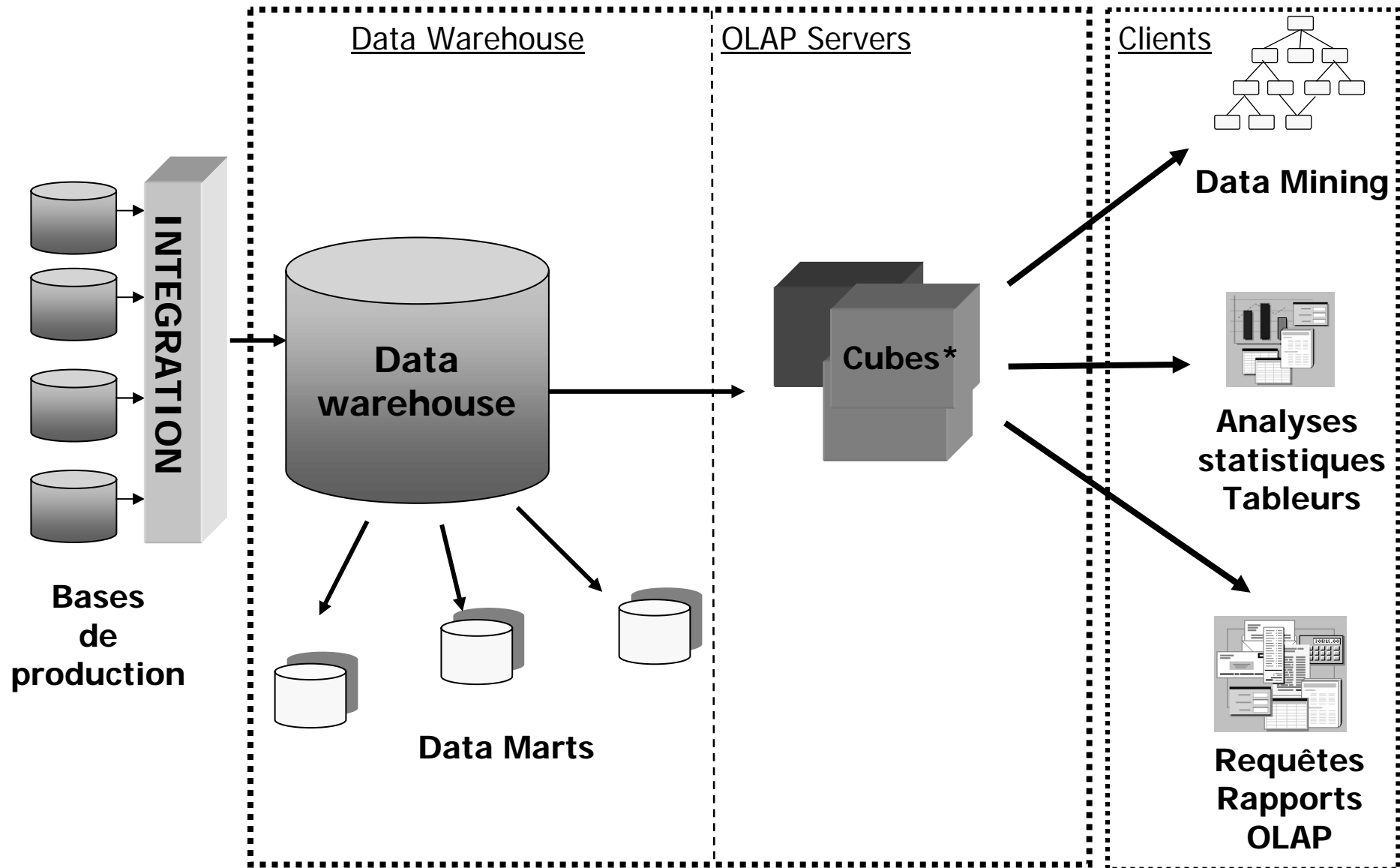
# Fonctions d'un Data Warehouse

---

- Récupérer des données existants dans différentes BD sources
- Stocker les données (historisées)
- Mettre à disposition les données pour :
  - ◆ Interrogation
  - ◆ Visualisation
  - ◆ Analyse



# Architecture



\*MOLAP, ROLAP, HOLAP

# 3- Pourquoi ne pas utiliser un SGBD?

---

- SGBD et DW :

- ◆ ont des objectifs différents et font des traitements différents
- ◆ stockent des données différentes
- ◆ font l'objet de requêtes différentes

-> SGBD et DW ont besoin d'une organisation différente des données

-> SGBD et DW doivent être physiquement séparés.

# SGBD: Objectifs et traitements

---

- Les SGBD sont des systèmes dont le mode de travail est transactionnel (OLTP On-Line Transaction Processing).
- Permet d'insérer, modifier, interroger des informations rapidement, efficacement, en sécurité.
- Deux objectifs principaux :
  - ◆ Sélectionner, ajouter, mettre à jour et supprimer des tuples
  - ◆ Ces opérations doivent pouvoir être effectuées très rapidement, et par de nombreux utilisateurs simultanément.

# DW: Objectifs et traitements

---

- Les datawarehouse sont des systèmes conçus pour l'aide à la prise de décision. (Mode de travail: OLAP On-Line Analytical Processing)
- La plupart du temps sont utilisés en lecture (utilisateurs)
- Les objectifs principaux sont
  - ◆ regrouper, organiser des informations provenant de sources diverses,
  - ◆ les intégrer et les stocker pour donner à l'utilisateur une vue orientée métier,
  - ◆ retrouver et analyser l'information facilement et rapidement.

# Données différentes

---

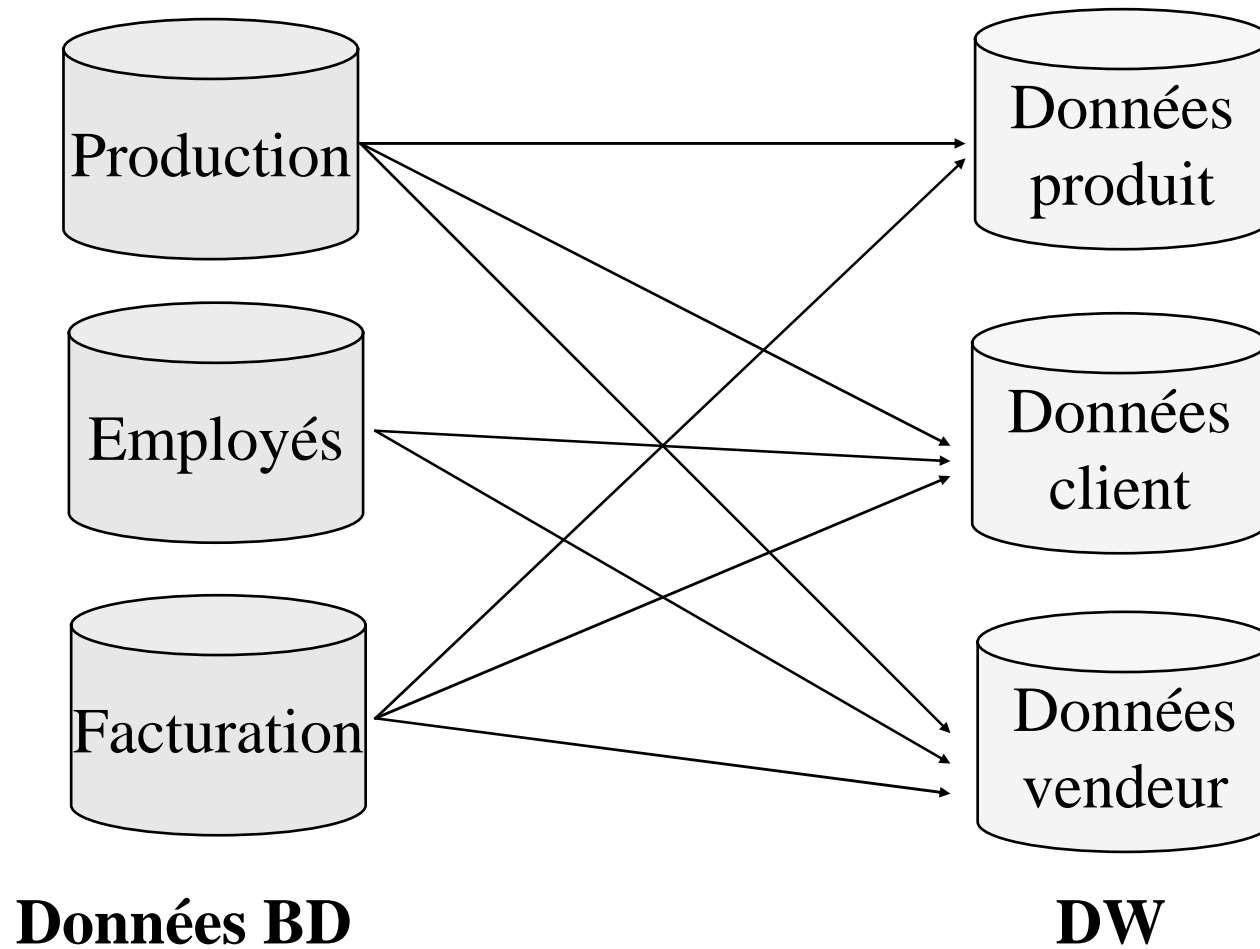
- D'après BILL Inmon :

“Un DW est une collection de données orientées sujet, intégrées, non volatiles, historisées, organisées pour la prise de décision.”

- ◆ Orientées sujet: 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.

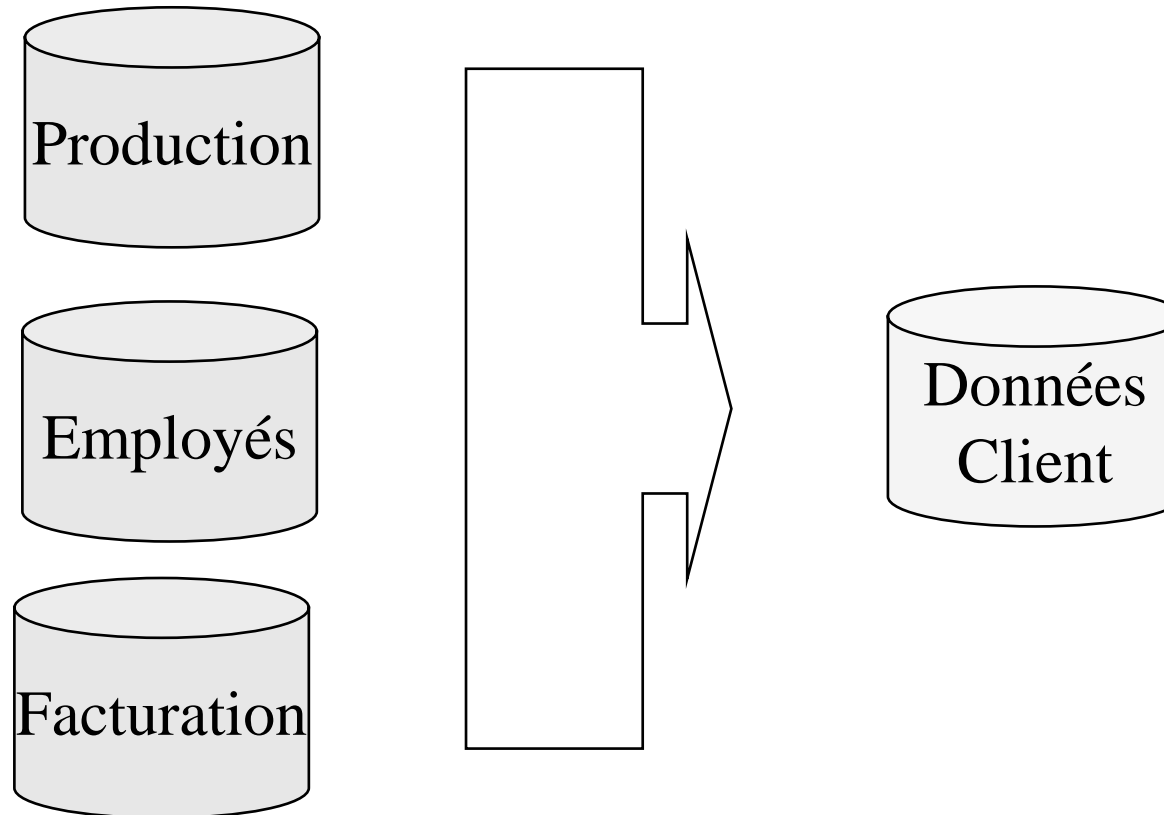
# Orientées sujet

---



# Données intégrées

---

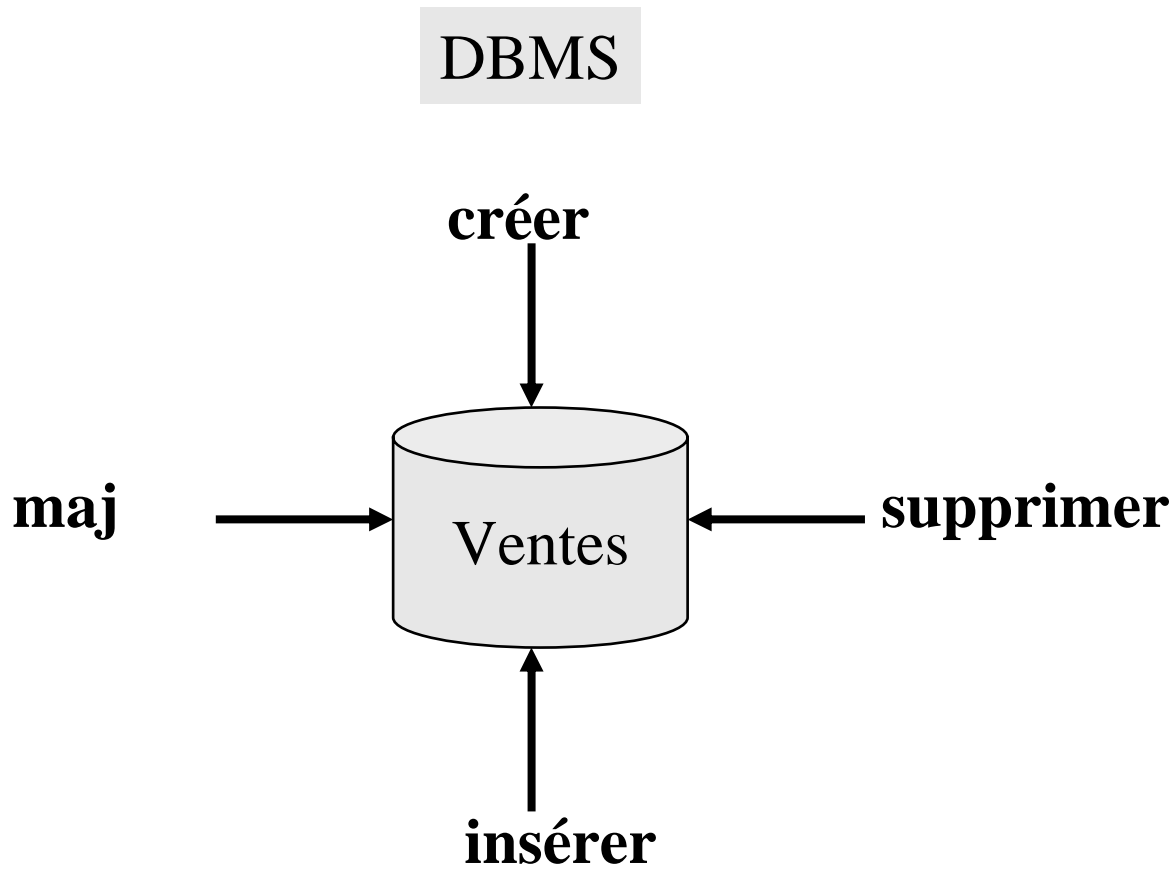


Techniques d'intégration des données

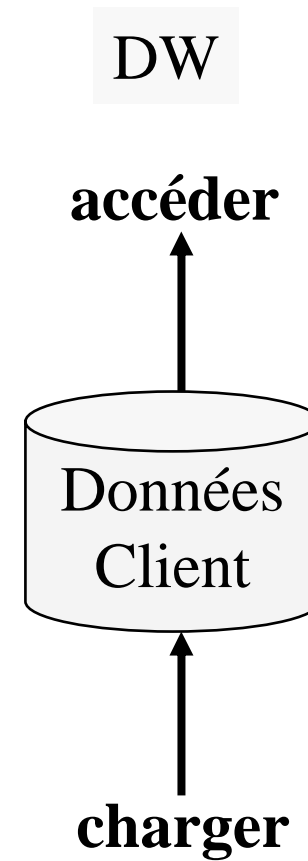
Techniques de nettoyage: Cohérence entre les différentes sources des noms, unités de mesures, etc

# Données non volatiles

---



Pas d'historique



Données historisées



# Les données

---

- Volume très important
- Données dispersées, souvent difficiles d'accès
- Peu ou mal intégrées
- Complexes
- Non structurées pour les applications décisionnelles

# Requêtes (1)

---

- BD-OLTP représentent les données sous forme aplatie: relation, données normalisées

produit	région	vente	date	vendeur
écrou	Est	50	01012004	X
écrou	Ouest	60	12122003	X
écrou	Centre	110	01112003	Y
vis	Est	70	01042004	Y
vis	Ouest	80	10022004	Z
vis	Centre	90	29032004	Y
boulon	Est	120	05052004	X
boulon	Ouest	10	24042004	Z
boulon	Centre	20	11022004	Y
joint	Est	50	01032004	X
joint	Ouest	40	01102003	Y
joint	Centre	70	01012003	Z

produit	prix	fournisseur
écrou	44	CC
vis	2	DD
boulon	3	VV
joint	1	BB

fournisseur	ville
...	...

# Requêtes (2)

---

- OLTP: Requêtes simples "qui, quoi"
  - ◆ par ex. les ventes de X.
  - ◆ jointures: les ventes de X à quel prix de quel fournisseur,
- OLAP: besoin de données agrégées, synthétisées
  - ◆ nombre de ventes par vendeur, par région, par mois,
  - ◆ nombre de ventes par vendeur, par fournisseur, par mois,
  - ◆ ...
- SQL: Possibilité d'agréger les données (group by) mais très coûteux (parcourir toutes les tables) et il faut recalculer à chaque utilisation
- Sur plusieurs tables (ex : somme des ventes par fournisseur), nécessité de faire des jointures souvent coûteuses

# Différences BD – DW

---

<b>Caractéristiques</b>	<b>OLTP</b>	<b>OLAP</b>
Applications	production	aide à la décision
Utilisateurs	un département professionnel IT	transversal (entreprise) décideur non IT
Données	normalisées, non agrégées	dénormalisées, agrégées
Requêtes	simples, nombreuses, régulières, prévisibles, répétitives	complexes, peu nombreuses, irrégulières, non prévisibles
Nb tuples invoqués par requête (moyenne)	dizaines	millions
Taille données	100 MB à 1 GB	1 GB à 1 TB
Ancienneté des données	récente, mises à jour	historique

# Nécessité d'une structure multi-dimensionnelle

---

- Les BD relationnelles ne sont pas adaptées à l'OLAP car :
  - ◆ Pas les mêmes objectifs
  - ◆ Pas les mêmes données:
    - Les données nécessaires à l'OLAP sont multi-dimensionnelles (i.e. ventes par vendeur, par date, par ville...). Les tables en représentent une vue aplatie.
  - ◆ Pas les mêmes traitements et requêtes:
    - Non seulement perte de performances mais aussi nécessité pour les utilisateurs de savoir comment trouver les liens entre les tables pour recréer la vue multi-dimensionnelle.
- Il est donc nécessaire de disposer d'une structure de stockage adaptée à l'OLAP, i.e. permettant de
  - ◆ représenter les données dans plusieurs dimensions,
  - ◆ manipuler les données facilement et efficacement.

# Séparation BD et DW

---

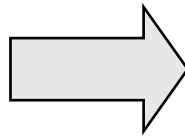
- Les DW vont être physiquement séparés des BD, pour des raisons de:
  - ◆ Performance : systèmes de production ne sont pas organisés pour pouvoir répondre efficacement aux requêtes des systèmes d'aide à la décision. Même les requêtes simples peuvent dégrader sérieusement les performances.
  - ◆ Données différentes:
    - Données historisées : aide à la décision nécessite des données sur une longue durée, non conservée dans les BD
    - Données agrégées
    - Qualité des données : sources différentes qui utilisent souvent des noms, formats, codes et mesures différents devant être uniformisés

# DW-OLAP

---

- DW-OLAP :
  - ◆ représentation des données sous forme multidimensionnelle : 'Cube'

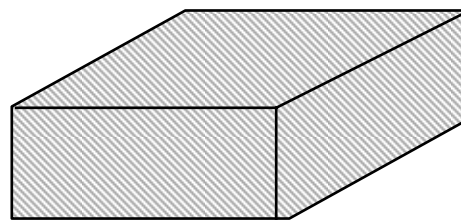
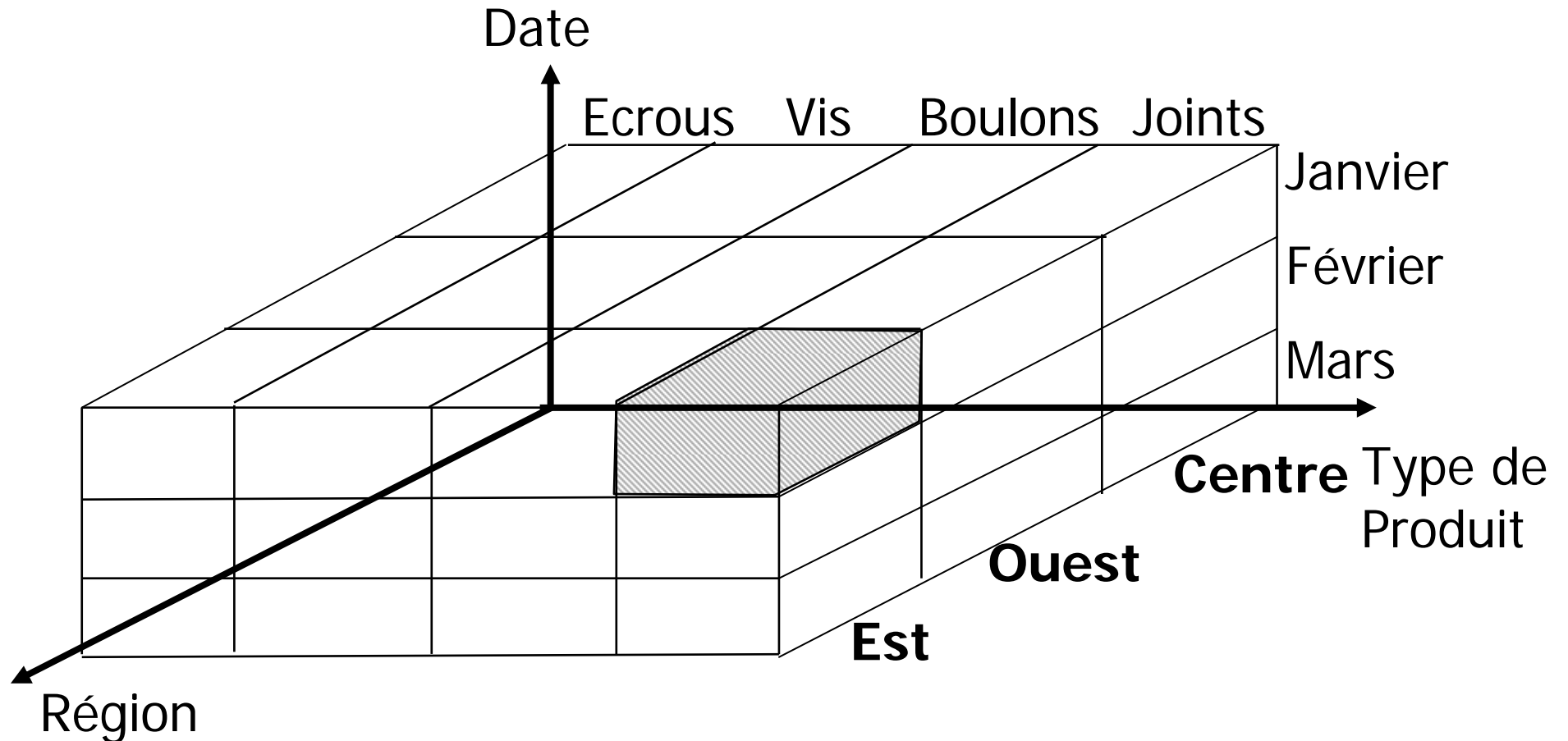
produit	région	vente
écrou	Est	50
écrou	Ouest	60
écrou	Centre	110
vis	Est	70
vis	Ouest	80
vis	Centre	90
boulon	Est	120
boulon	Ouest	10
boulon	Centre	20
joint	Est	50
joint	Ouest	40
joint	Centre	70



	Est	Ouest	Centre
écrous	50	60	110
vis	70	80	90
boulons	120	10	20
joints	50	40	70

# Cube: représentation des données sous forme multidimensionnelle

---



Vente de joints en janvier pour la région est



# Partie 2 : Implémentation d'un data warehouse

# Approches pour créer un DW

---

- 3 possibilités:
- (1) Relational OLAP (ROLAP)
  - ◆ Données sont stockées dans un SGBD relationnel
  - ◆ Un moteur OLAP permet de simuler le comportement d'un SGBD multi-dimensionnel
- (2) Multidimensional OLAP (MOLAP)
  - ◆ Structure de stockage en cube
  - ◆ Accès direct aux données dans le cube
- (3) Hybrid OLAP (HOLAP)
  - ◆ Données stockées dans SGBD relationnel (données de base)
  - ◆ + structure de stockage en cube (données agrégées)

# ROLAP

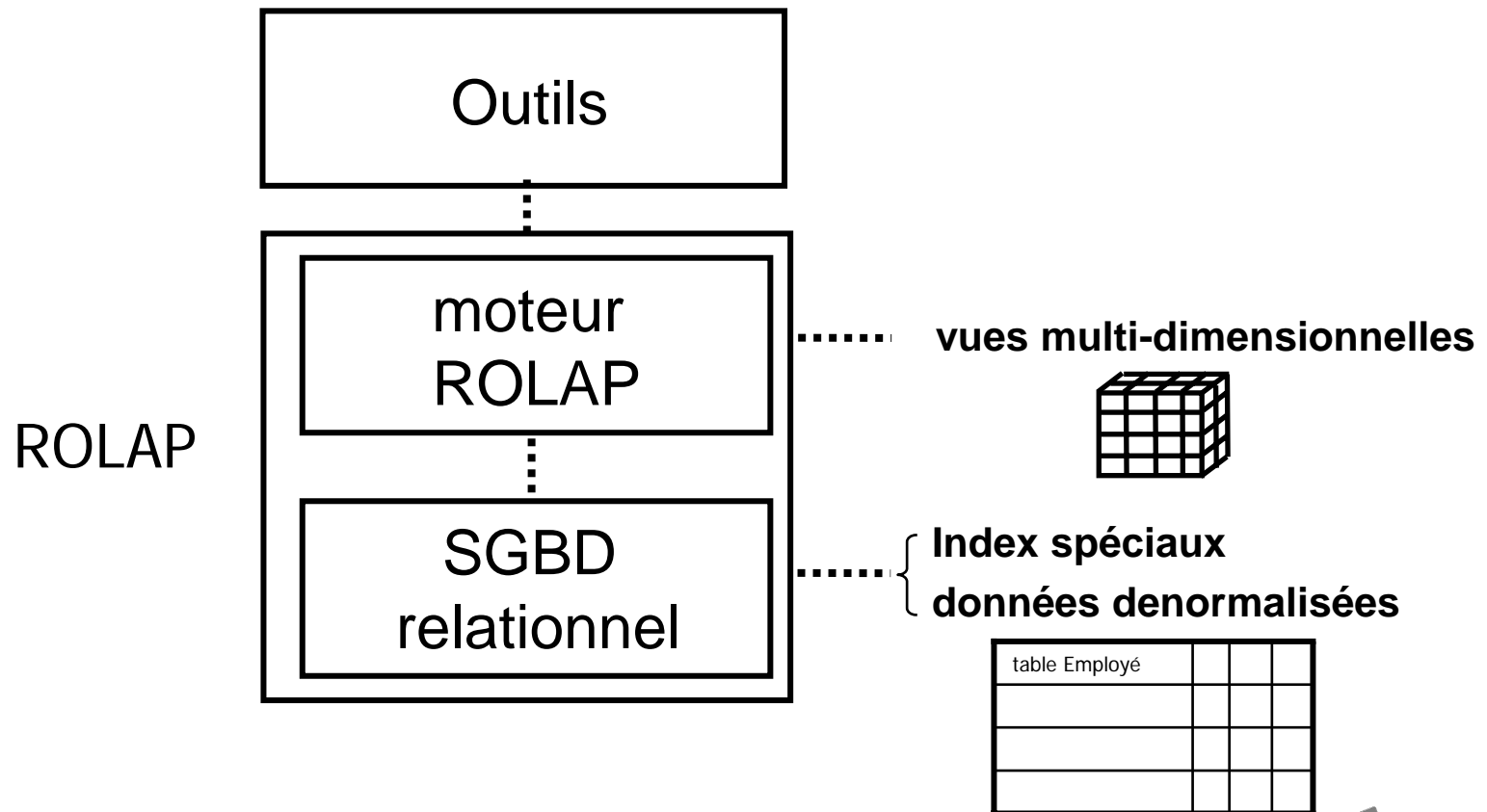
---

- Idée:
  - ◆ Données stockées en relationnel.
  - ◆ La conception du schéma est particulière: schéma en étoile, schéma en flocon
  - ◆ Des vues (matérialisées) sont utilisées pour la représentation multidimensionnelle
  - ◆ Les requêtes OLAP (slice, rollup...) sont traduites en SQL.
  - ◆ Utilisation d'index spéciaux: bitmap
  - ◆ Administration (tuning) particulier de la base
- Avantages/inconvénients
  - ◆ Souplesse, évolution facile, permet de stocker de gros volumes.
  - ◆ Mais peu efficace pour les calculs complexes

# ROLAP

---

- Relational OLAP



# MOLAP

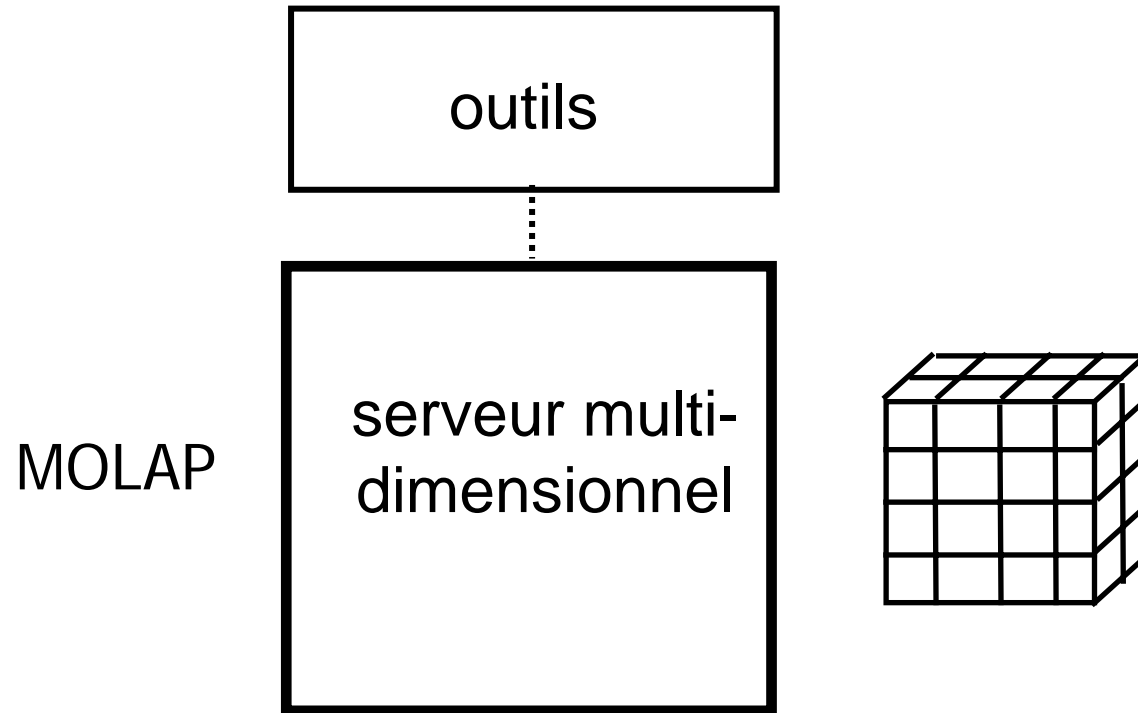
---

- Idée:
  - ◆ Modélisation directe du cube
  - ◆ Ces cubes sont implémentés comme des matrices à plusieurs dimensions
    - CUBE [1:m, 1:n, 1:p...] (mesure)
  - ◆ Le cube est indexé sur ses dimensions
- Avantages/inconvénients:
  - ◆ rapide
  - ◆ formats propriétaires
  - ◆ ne supporte pas de très gros volumes de données

# MOLAP

---

- Multi-Dimensional OLAP



# HOLAP

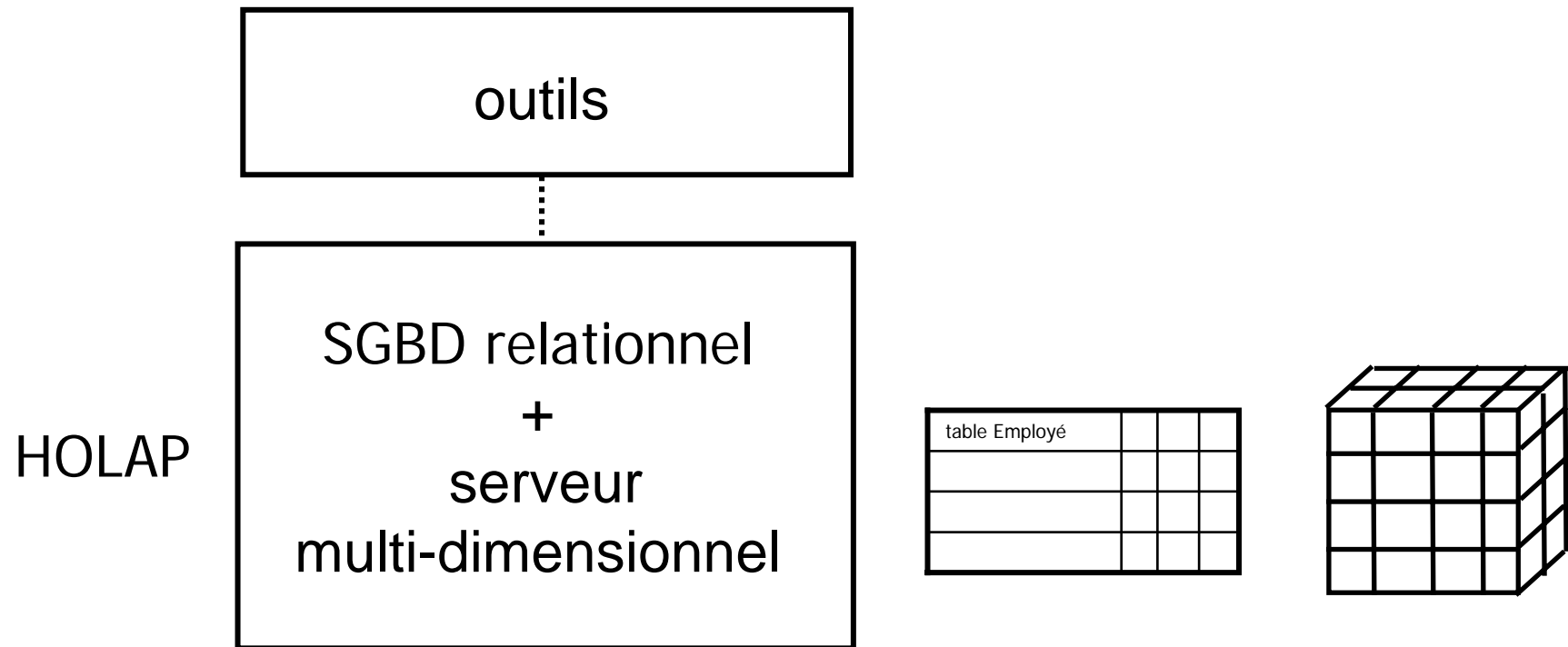
---

- Idée:
  - ◆ MOLAP + ROLAP
  - ◆ Données stockées dans des tables relationnelles
  - ◆ Données agrégées stockées dans des cubes.
  - ◆ Les requêtes vont chercher les données dans les tables et les cubes

# HOLAP

---

- Hybrid OLAP





# Data warehouse

---

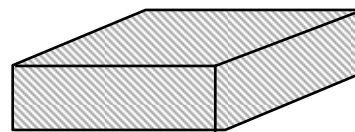
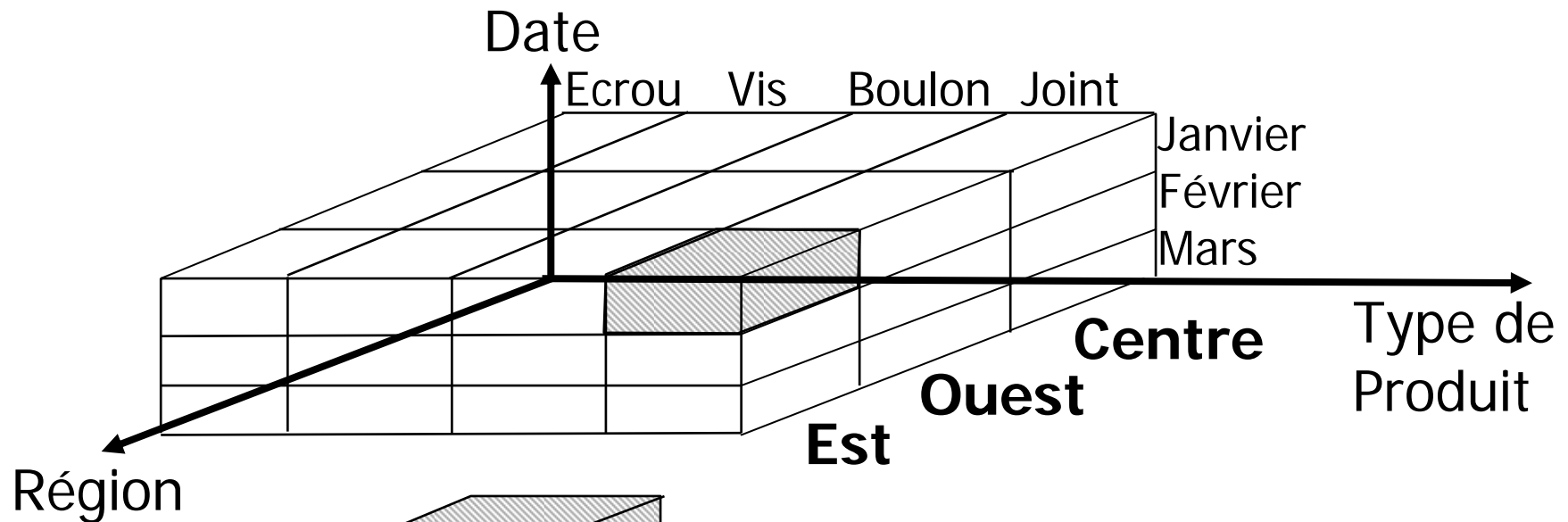
- Implémentation:
  - ◆ ROLAP
  - ◆ MOLAP
  - ◆ HOLAP
- Modèles de représentation:
  - ◆ Cubes (partie 3)
  - ◆ Étoile & flocon (partie 4)

# Partie 3: Cube

# Hyper Cube

---

- Hypercube : BD multidimensionnelle
  - ◆ Axes: dimensions (date, type de produits, région),
  - ◆ Chaque cellule de l'hypercube contient une mesure calculée (vente de produit).

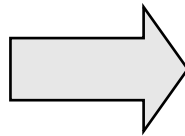


Vente de joints en janvier  
pour la région est

# Rappel

---

produit	région	vente
écrou	Est	50
écrou	Ouest	60
écrou	Centre	110
vis	Est	70
vis	Ouest	80
vis	Centre	90
boulon	Est	120
boulon	Ouest	10
boulon	Centre	20
joint	Est	50
joint	Ouest	40
joint	Centre	70



	Est	Ouest	Centre
<b>écrous</b>	50	60	110
<b>vis</b>	70	80	90
<b>boulons</b>	120	10	20
<b>joints</b>	50	40	70

# Définitions

---

- Principe de base : ce sont les analyses des indicateurs qui intéressent l'utilisateur
- Le modèle multidimensionnel contient 2 types d'attributs : les *dimensions* et les *mesures*
- Les mesures sont les valeurs numériques que l'on compare (ex : montant\_ventes, qte\_vendue)
  - ◆ Ces valeurs sont le résultat d'une opération d'agrégation des données
- Les dimensions sont les points de vues depuis lesquels les mesures peuvent être observées :
  - ◆ Ex : date, région, type de produit, etc.

# Dimension (1)

---

- Dimension = liste d'éléments
- Dimension contient des membres organisés en hiérarchie :
  - ◆ Chacun des membres appartient à un niveau hiérarchique (ou niveau de granularité) particulier
- Granularité d'une dimension : nombre de niveaux hiérarchiques

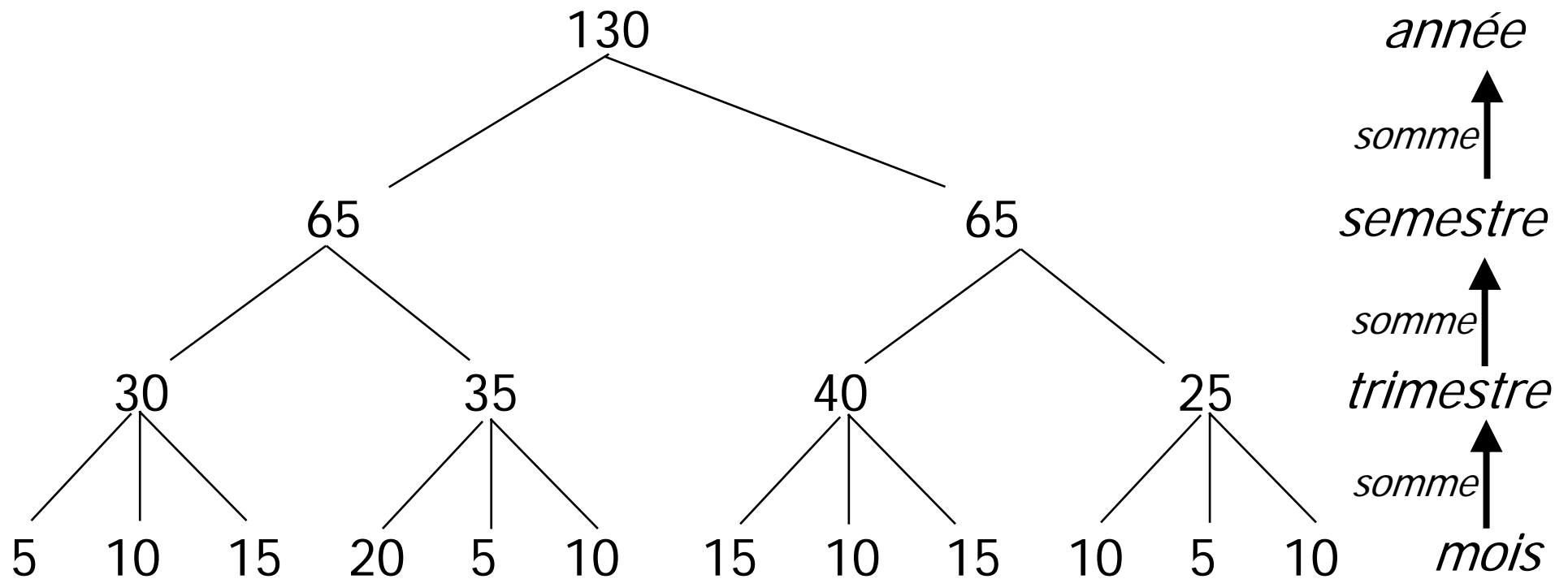
**Temps :**

**année – semestre – trimestre - mois**

# Dimension (2)

---

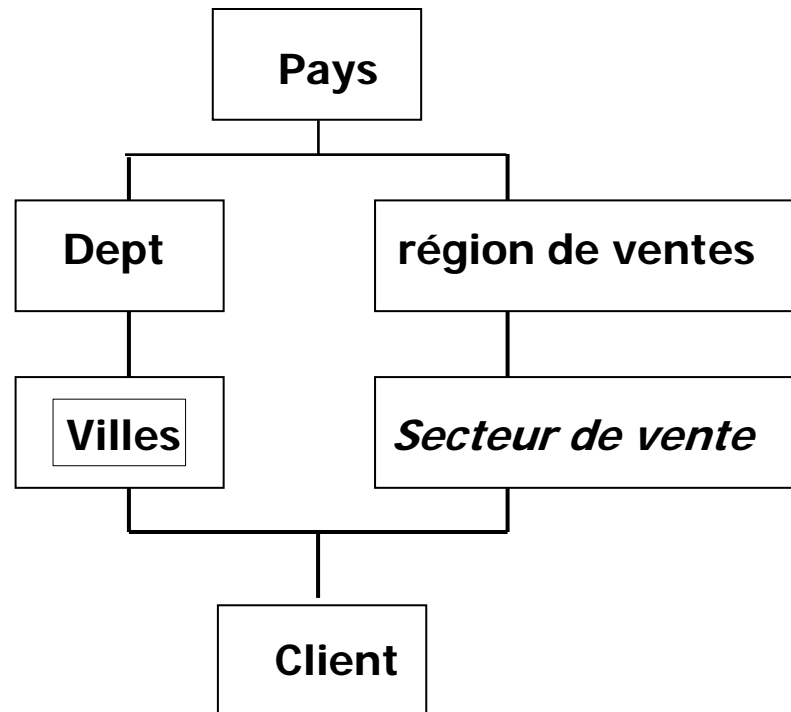
- Les axes de dimension doivent fournir des règles de calcul d'agrégat pour chaque mesure:



# Hiérarchies multiples de dimensions

---

- Hiérarchies multiples dans une dimension :





# Manipulation d'un cube

---

- Opérateurs appliqués sur le cube sont algébriques (le résultat est un autre cube) et peuvent être combinés
- Opérateurs sont:
  - ◆ Slicing & Dicing (extraction) *Opérateur sur le cube*
  - ◆ Changement de la granularité d'une dimension
    - Roll up (agrégation d'une dimension => résumé)
    - Drill down (plus détaillées)

*Opérateurs sur les dimensions*

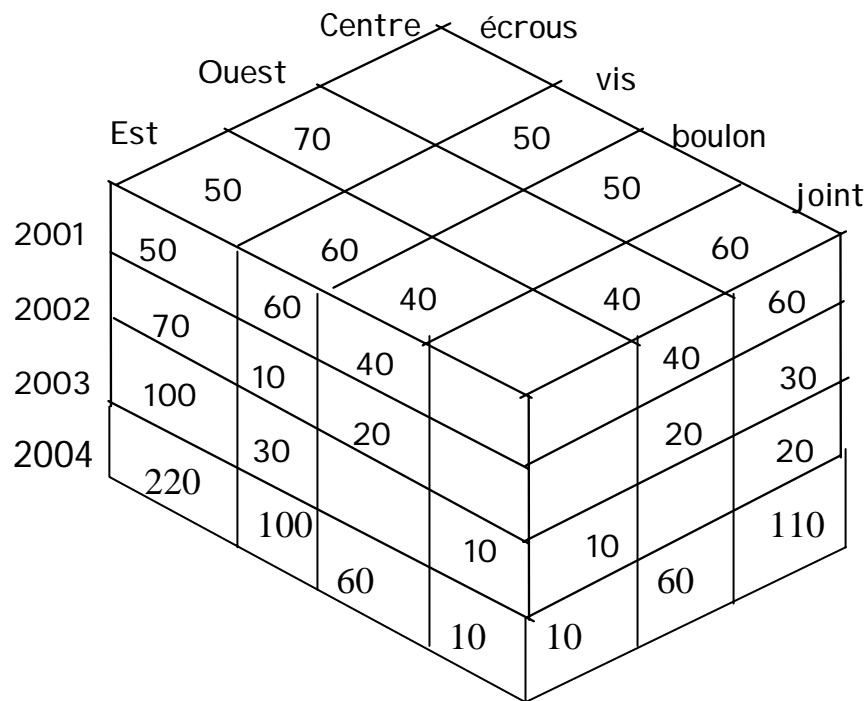
# Slicing et dicing

---

- Slicing: Sélection de tranches du cube par des prédicats selon une dimension
  - ◆ filtrer une dimension selon une valeur
  - ◆ Exemple: Slice (2004) : on ne retient que la partie du cube qui correspond à cette date
- Dicing: extraction d'un sous-cube

# Exemple: slicing

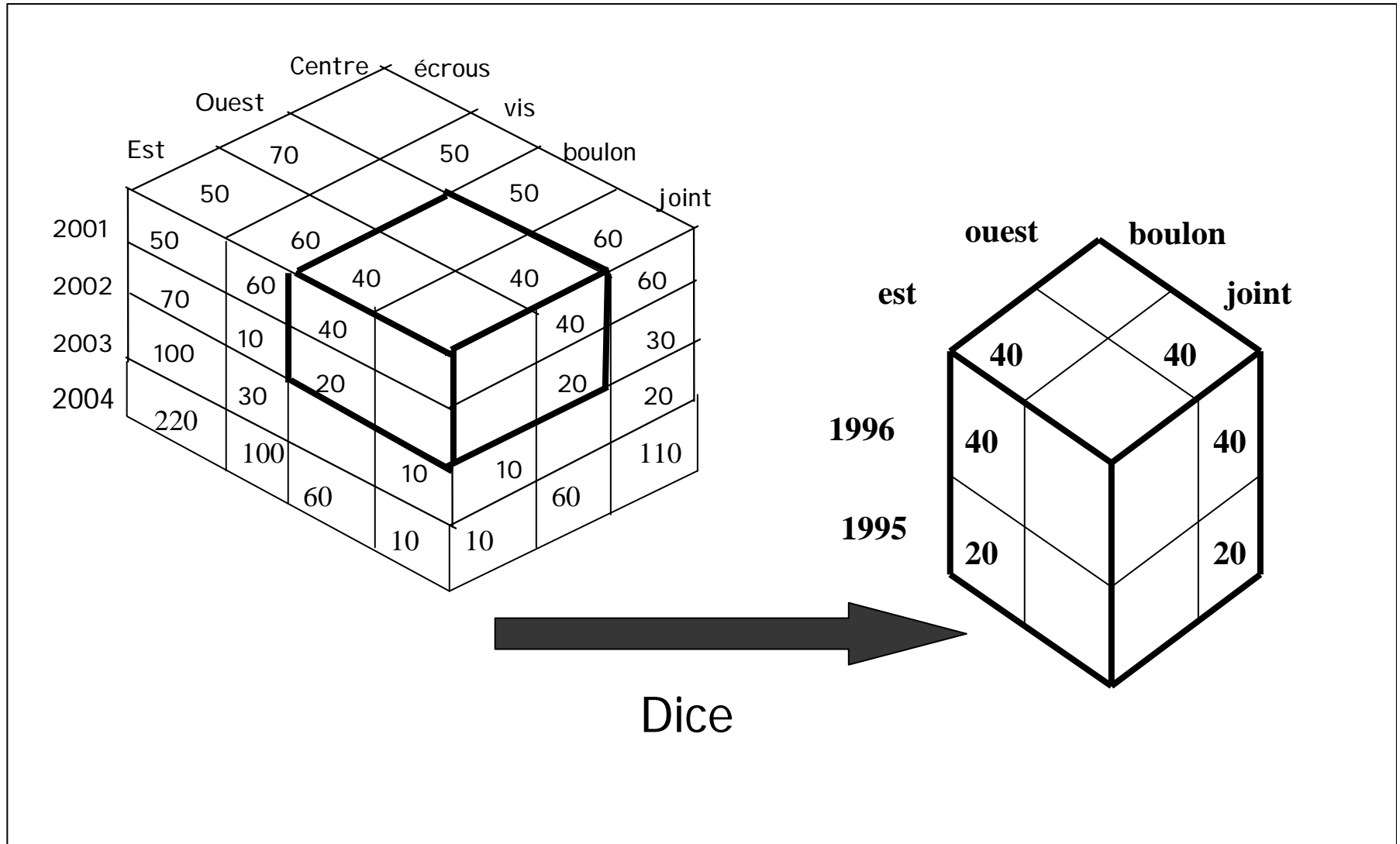
Slice



Slice (2004)

<b>Ventes 2004</b>	<b>écrou</b>	<b>vis</b>	<b>boulon</b>	<b>joint</b>
<b>est</b>	220	100	60	10
<b>ouest</b>	160	50	10	60
<b>centre</b>	20	150	170	110

# Exemple: Dicing

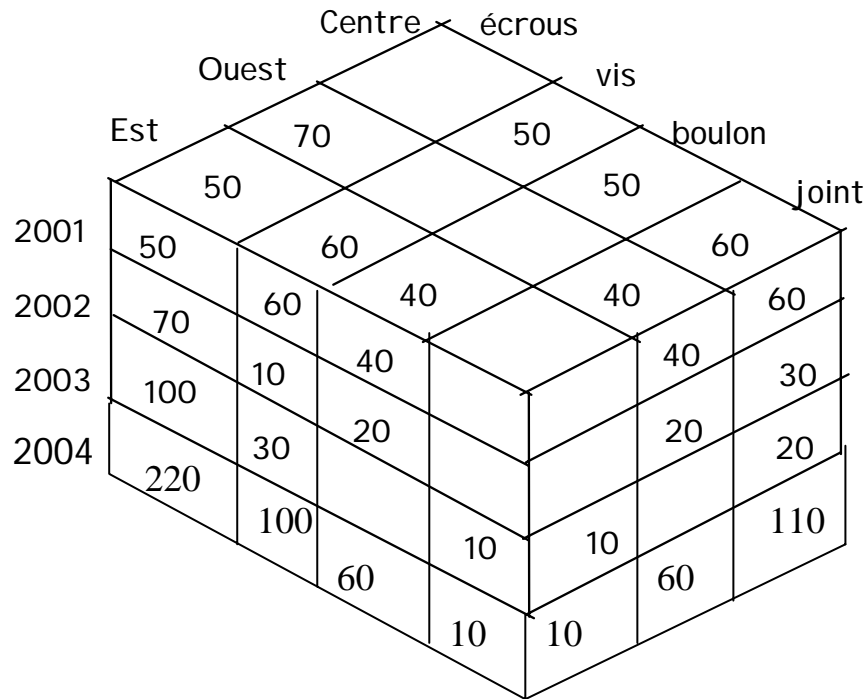


# Opérations liées à la granularité

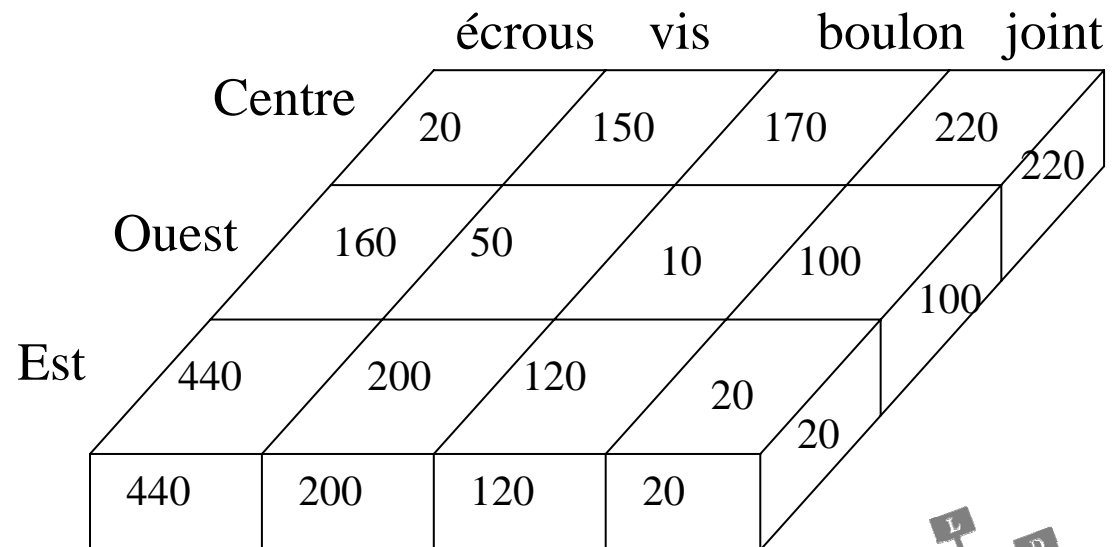
---

- Les opérations agissant sur la granularité d'observation des données caractérisent la hiérarchie de navigation entre les différents niveaux.
- Roll-up ou forage vers le haut: consiste à représenter les données du cube à un niveau de granularité supérieur conformément à la hiérarchie définie sur la dimension.
  - ◆ Utilisation de la fonction d'agrégation (somme, moyenne, etc) spécifiée pour la mesure et la dimension
- Drill-down ou forage vers le bas : consiste à représenter les données du cube à un niveau de granularité de niveau inférieur, donc sous une forme plus détaillée.

# Rollup Exemple



Roll up sur année



# Drill-Down exemple

**Drill-down ~ opération inverse de Roll-up**

**Drill-down du niveau des régions au niveau villes**

		joint	bordeaux	dijon	grenoble	lille	lyon	marseille	montpellier	nantes	paris	poitiers
	écrous	30	20	30	40	20	10	10	10	10	70	10
	vis	10	30	30	30	40	20	30	20	50	10	10
	boulon	20			10		20	30	20	50	10	10
		30	30	30	30	30	30	30	30	30	30	30
2001		30			30					10	10	10
2002		30	20	20	30	30	20	50	10			10
2003		10	10	20	60	20	40	20	30	10	20	
2004		10	20	10	10	70	40	10		10	10	

# Partie 4: Modèles en étoile et en flocon



# Modélisation en étoile ou en flocons

---

- Modélisation conceptuelle BD : entité et relation
- Modélisation de DW : dimension et mesure
- Les mesures sont les valeurs numériques que l'on compare (ex : montant\_ventes, qte\_vendue)
  - ◆ Ces valeurs sont le résultat d'une opération d'agrégation des données
- Les dimensions sont les points de vues depuis lesquels les mesures peuvent être observées :
  - ◆ Ex : date, localisation, produit, etc.
  - ◆ Elles sont stockées dans les tables de dimensions

# Les dimensions

---

- Une dimension peut être définie comme :
  - ◆ un thème, ou un axe (attributs), selon lequel les données seront analysées
  - ◆ Ex : Temps, Découpage administratif, Produits
- Une dimension contient des membres organisés en hiérarchie :
  - ◆ Chacun des membres appartient à un niveau hiérarchique (ou niveau de granularité) particulier
  - ◆ Ex : pour la dimension Temps: année – semestre – mois – jour

# Les mesures

---

- Une mesure est un élément de donnée sur lequel portent les analyses, en fonction des différentes dimensions
  - ◆ Ex : coût des travaux, nombre d'accidents, ventes

# Les faits

---

- Un fait représente la valeur d'une mesure, mesurée ou calculée, selon un membre de chacune des dimensions
  - ◆ Exemple : «250 000 euros » est un fait qui exprime la valeur de la mesure « coût des travaux » pour le membre « 2002 » du niveau année de la dimension « temps » et le membre « Versailles » du niveau « ville » de la dimension « découpage administratif »

# La table de faits

---

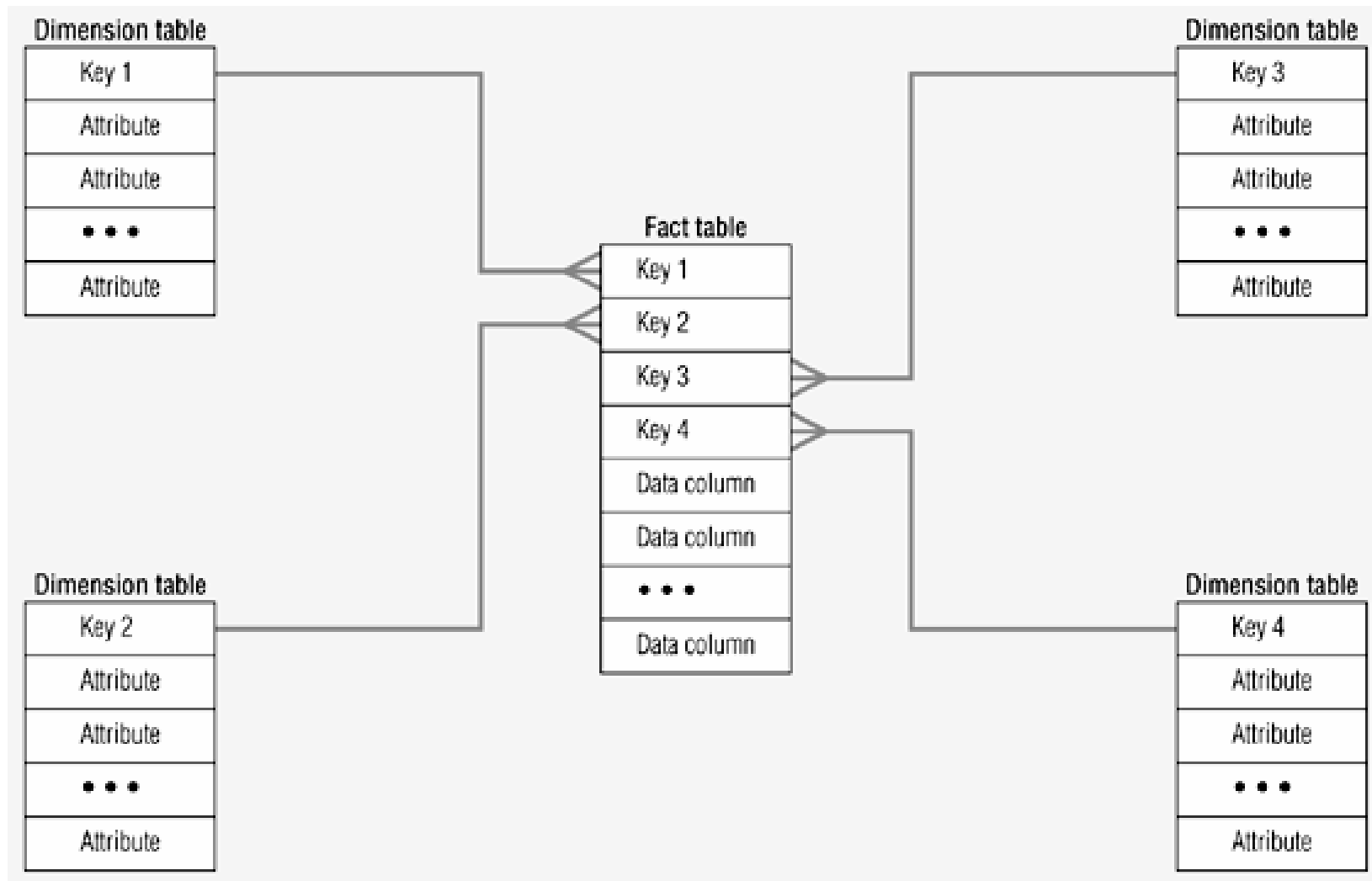
- Les mesures sont stockées dans les tables de faits
  - ◆ Table de fait contient les valeurs des mesures et les clés vers les tables de dimensions

# Le modèle en étoile

---

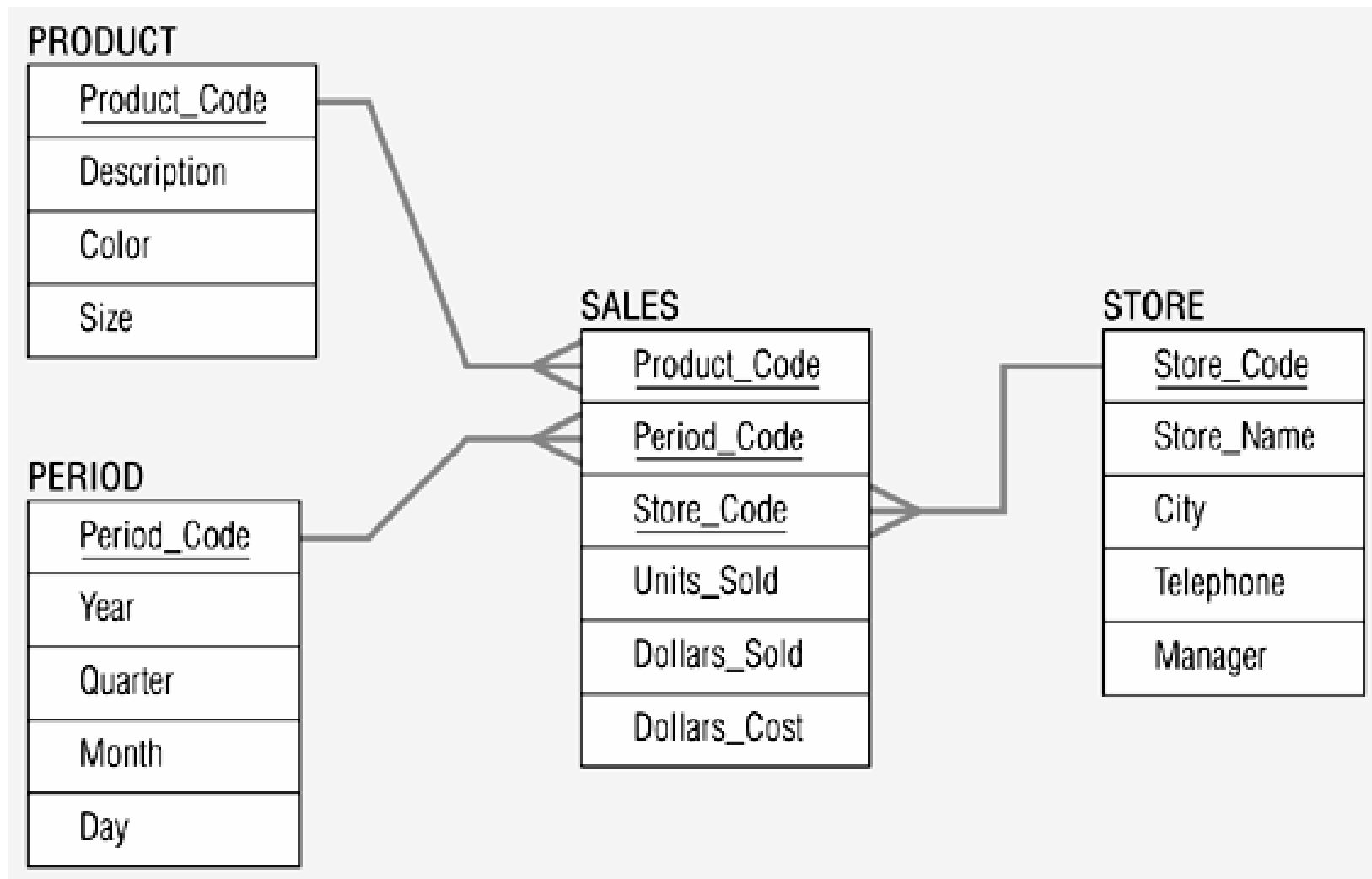
- Une (ou plusieurs) table(s) de faits comprenant une ou plusieurs mesures.
- Plusieurs tables de dimension dénormalisées: descripteurs des dimensions.
- Les tables de dimension n'ont pas de lien entre elles
- Avantages :
  - ◆ Facilité de navigation
  - ◆ Performances : nombre de jointures limité ; gestion des données creuses.
  - ◆ Gestion des agrégats
- Inconvénients :
  - ◆ Toutes les dimensions ne concernent pas les mesures
  - ◆ Redondances dans les dimensions
  - ◆ Alimentation complexe.

# Modèle en étoile



# Exemple

---





Product

<u>Product _Code</u>	Description	Color	Size
100	Sweater	Blue	40
110	Shoes	Brown	10 1/2
125	Gloves	Tan	M
• • •			

Period

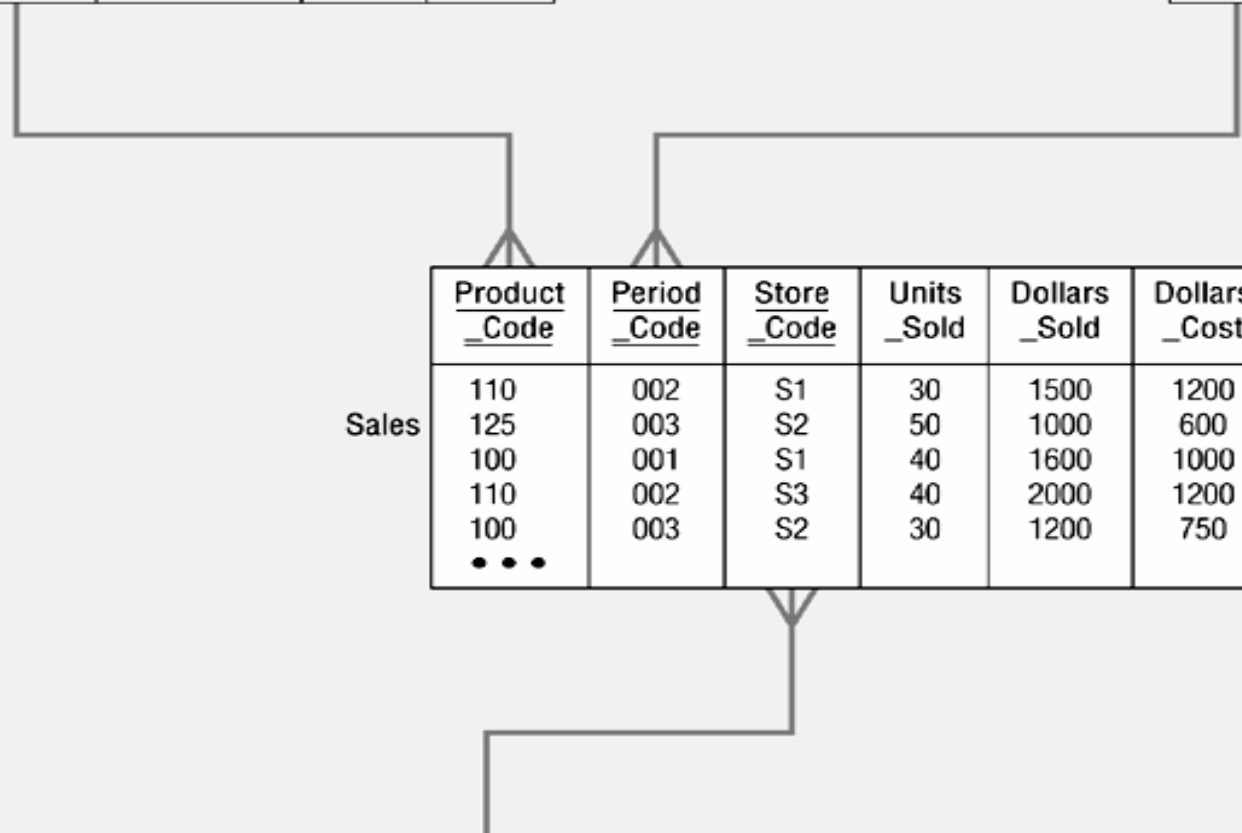
<u>Period _Code</u>	Year	Quarter	Month
001	1999	1	4
002	1999	1	5
003	1999	1	6
• • •			

Sales

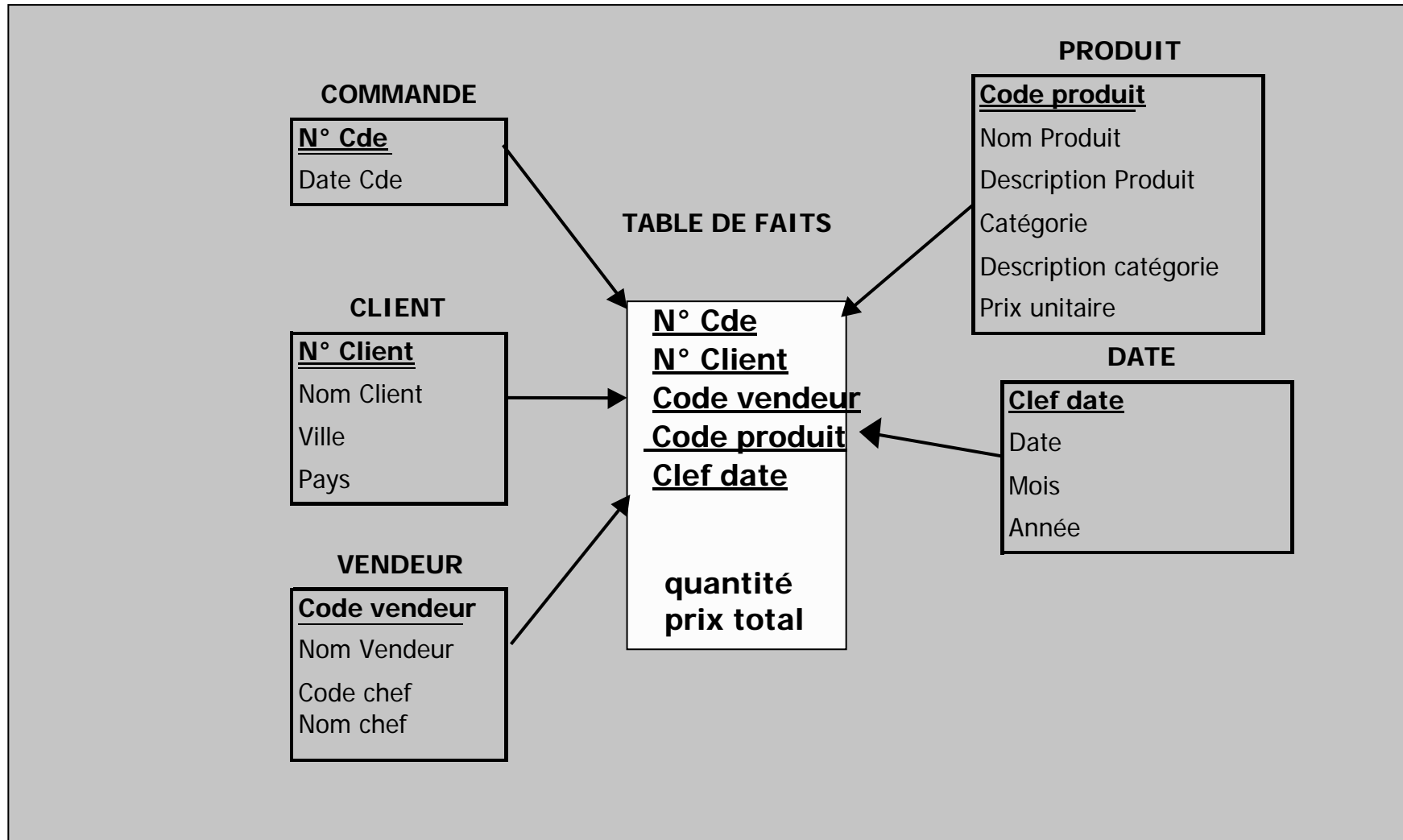
<u>Product _Code</u>	<u>Period _Code</u>	<u>Store _Code</u>	Units _Sold	Dollars _Sold	Dollars _Cost
110	002	S1	30	1500	1200
125	003	S2	50	1000	600
100	001	S1	40	1600	1000
110	002	S3	40	2000	1200
100	003	S2	30	1200	750
• • •					

Store

<u>Store _Code</u>	Store _Name	City	Telephone	Manager
S1	Jan's	San Antonio	683-192-1400	Burgess
S2	Bill's	Portland	943-681-2135	Thomas
S3	Ed's	Boulder	417-196-8037	Perry
• • •				



# Exemple de modèle en étoile



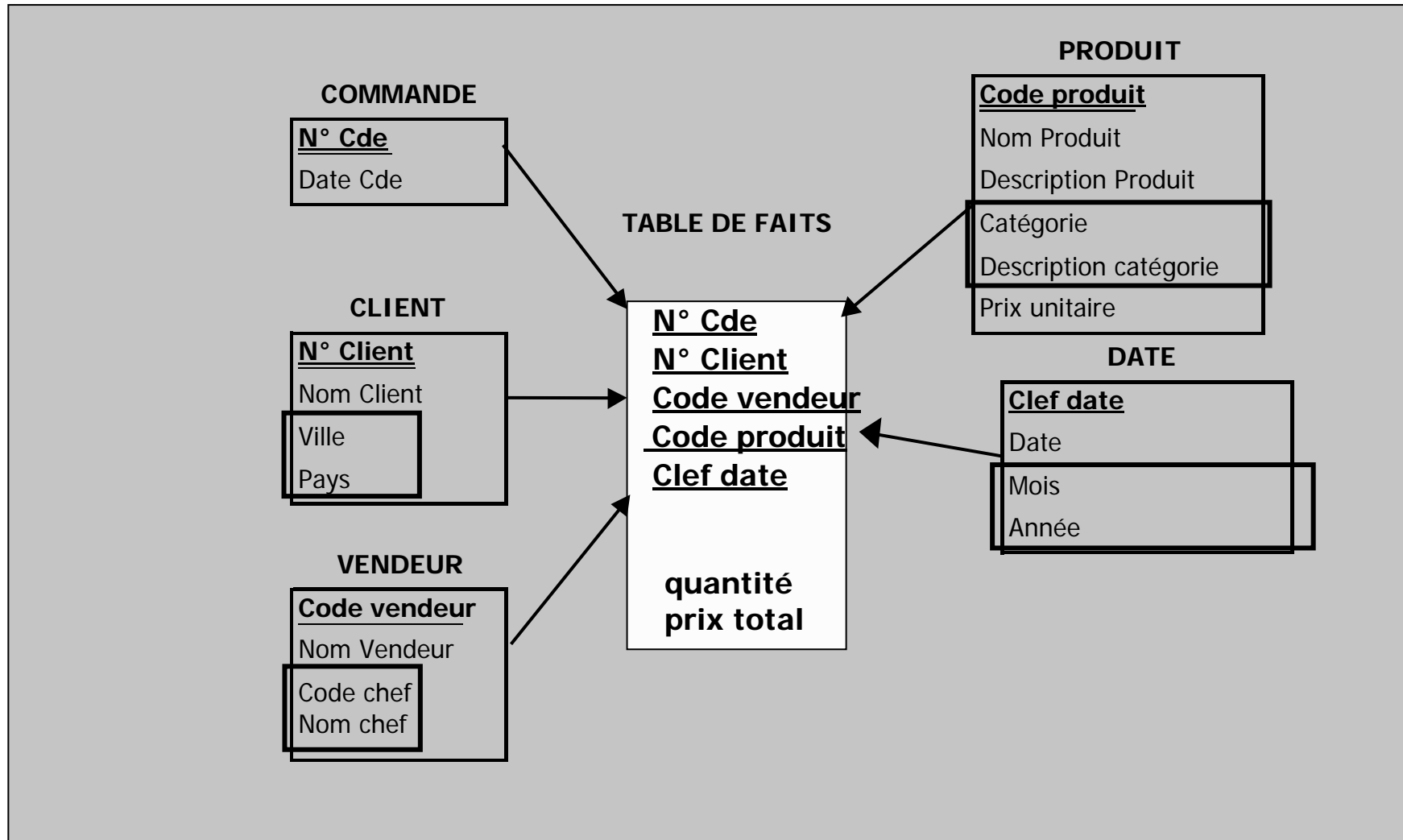
# Le modèle en flocons

---

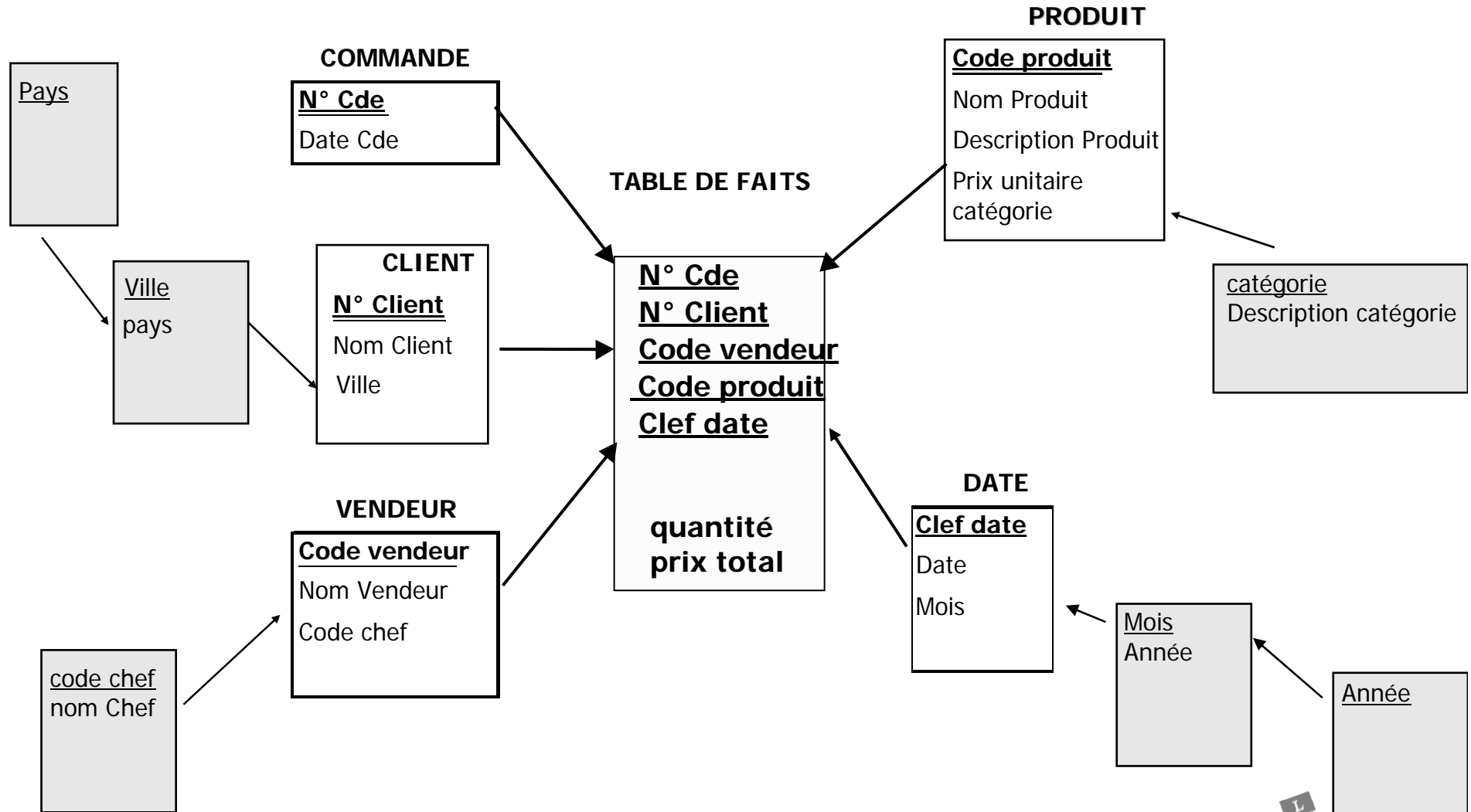
- Le schéma en flocon est dérivé du schéma en étoile où les tables de dimensions sont normalisées (la table des faits reste inchangée).
- Avec ce schéma, chacune des dimensions est décomposée selon sa (ou ses) hiérarchie(s).
- Exemple : Commune, Département, Région, Pays, Continent

<b>Client</b>	<b>Continent</b>	<b>Pays</b>	<b>Region</b>	<b>Département</b>	<b>Commune</b>
Pepone	Europe	France	RhôneAlpes	Rhône	Lyon1
Testut	Europe	France	RhôneAlpes	Rhône	Lyon2
Soinin	Europe	France	RhôneAlpes	Rhône	Lyon3
Vepont	Europe	France	Ile de France	Paris	Paris1
Martin	Europe	France	Ile de France	Paris	Paris2
Elvert	Europe	France	Ile de France	Yvelines	Versailles

# Modèle en flocon



# Le modèle en flocons



# Modèle en flocons de neige

---

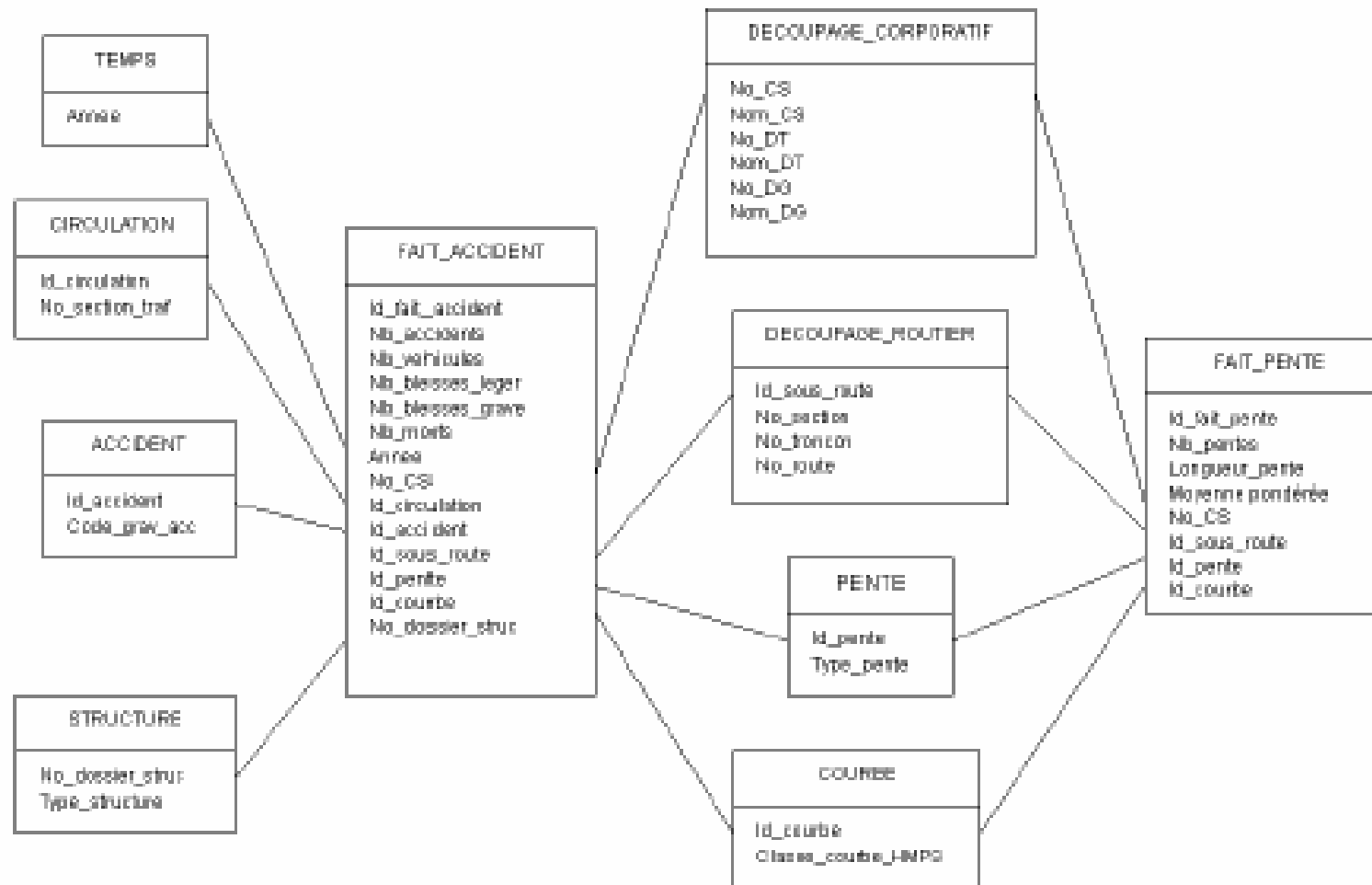
- Modèle en étoile + normalisation des dimensions
- Lorsque les tables sont trop volumineuses
  - ◆ Avantages :
    - réduction du volume,
  - ◆ Inconvénients :
    - navigation difficile,
    - nombreuses jointures.

# Le modèle en constellation

---

- La modélisation en constellation consiste à fusionner plusieurs modèles en étoile qui utilisent des dimensions communes.
- Un modèle en constellation comprend donc plusieurs tables de faits et des tables de dimensions communes ou non à ces tables de faits.

# Modèle en constellation





# Partie 5: Conception d'un DW

# Conception d'un data warehouse

---

- Différentes phases:
  - ◆ 1. Conception
    - 1. Définir la finalité du DW : Piloter quelle activité de l'entreprise ?
    - 2. Définition du modèle de données (modèle en étoile/flocon ou cubes)
  - ◆ 2. Acquisition des données
    - 1. Déterminer et recenser les données à entreposer: recherche des données dans les sources de l'entreprise
    - 2. Nettoyage des données
    - 2. Démarches d'alimentation
  - ◆ 3. Définir les aspects techniques de la réalisation
  - ◆ 4. Définir les modes de restitution, ...
  - ◆ 5. Stratégies d'administration, évolution, maintenance

# Acquisition des données

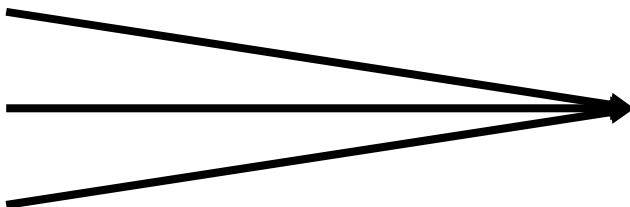
---

- Trois étapes:
  - ◆ Déterminer et recenser les données à entreposer:  
recherche des données dans les sources de données de l'entreprise
  - ◆ Nettoyage des données
    - conversions de données
    - filtrages
    - intégration
  - ◆ démarche d'alimentation
    - Incrémental ou total
    - Off-line ou. on-line
    - Fréquence de chargement: chaque nuit, 1/mois...
    - Détermination de la taille de l'historique (5 ans, 10 ans...)
    - Réalisée généralement par des outils dédiés (e.g. Powermart, Info Suite...)

# Exemple de nettoyage des données

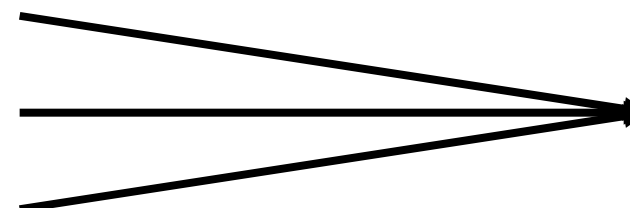
---

cm  
inches  
yards



cm

Jan.22,1999  
22/01/99  
01/22/99



Jan.22,1999

BD Facturation → Personne 1(Paul) → intégrrer (Paul)

BD Personnel → Personne 2(Paul) ↗

# Stockage

---

- Choix d'implémentation
  - ◆ MOLAP
  - ◆ ROLAP
  - ◆ HOLAP
- Implémentation du modèle en étoile et/ou des cubes et/ou des vues matérialisées
- Définition des index
- Stockage les données

# Restitution

---

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

# Différents types d'outils de restitution

---

- Différents types:
  - ◆ Requêteurs et Outils d'analyse
  - ◆ Outils de data mining

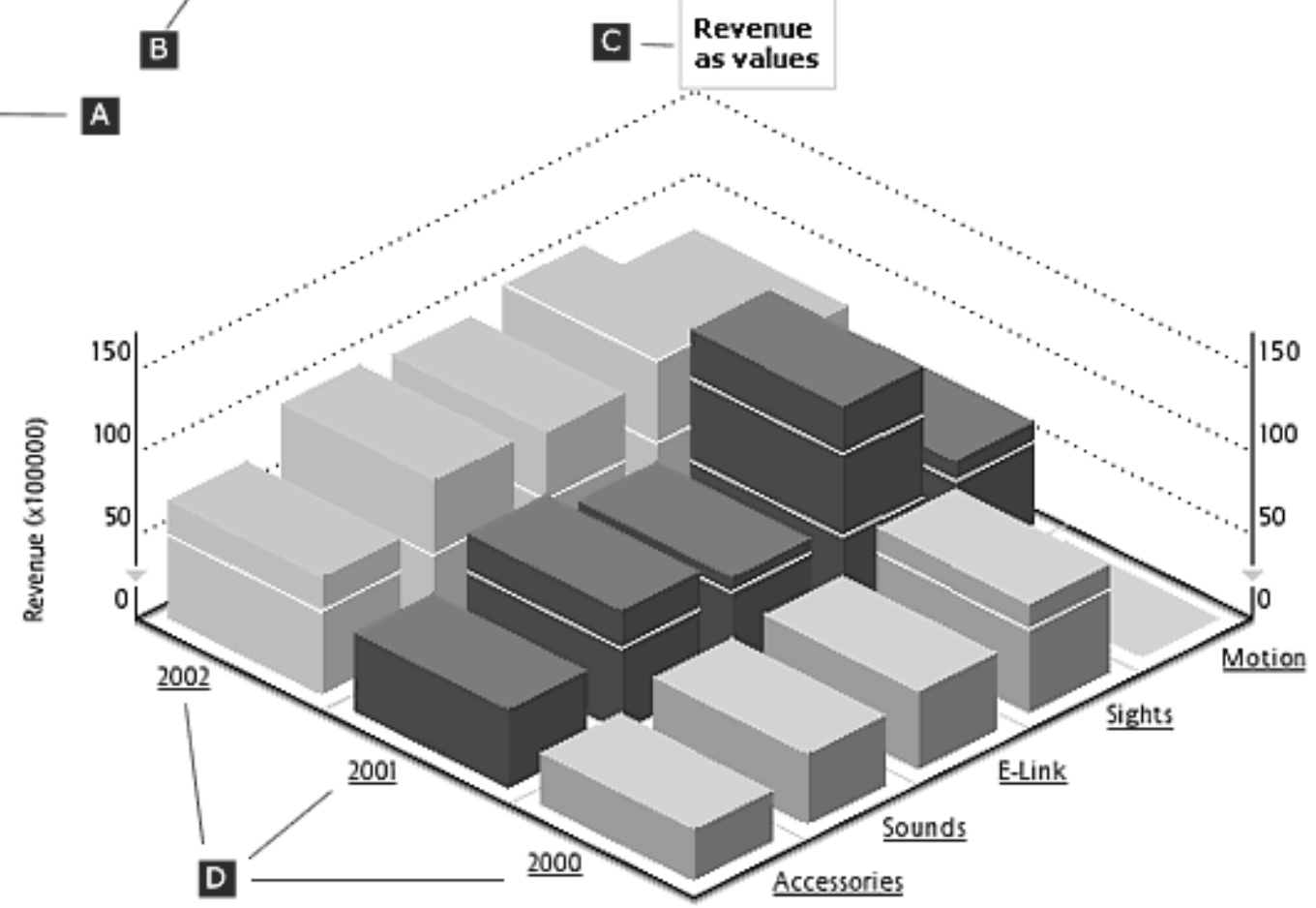
- [-] Sales
  - [-] Dates
  - [-] Products
  - [-] Locations
  - [-] Branch
  - [-] Order Method
  - [-] Margin Range
  - [-] LoyaltyType
  - [-] MEASURES

Navigation controls: [Back] [Forward] [Refresh] [Home] [Print] [Help] [Close]

Filter tabs: Dates | Products | Locations | Branch | Order Method | Margin Range

Dates

- 2000
- 2001
- 2002



Bottom toolbar with various icons for navigation and actions.



- [-] Sales
  - [+] Dates
  - [+] Products
  - [+] Locations
  - [+] Branch
  - [+] Order Method
    - [-] Direct Mail
    - [-] E-mail
    - [-] Live Semina
    - [-] Special
    - [-] Telesales
    - [-] Tradeshow
    - [-] Web Banne
  - [-] Margin Range
    - [-] Over 65 Pe
    - [-] 30 to 65 Pe
    - [-] Under 30 P
  - [+] LoyaltyType
  - [-] MEASURES
    - [-] Revenue
    - [-] Margin
    - [-] Quantity
    - [-] MarketingC
    - [-] Discount

Navigation: [Back] [Forward] [Refresh] [Home] [Print] [Help] [Close]

Filters: Dates | Products | Locations | Branch | Order Method

Revenue as values		Locations			Dates
<b>Accessories</b>	<b>Headphones</b>	Central Europe	Austria		
	<b>Microphones</b>	Americas	England		
	<b>Mobile Accessories</b>	Asia Pacific	France	Paris	
	<b>Remotes</b>	Northern Europe	Germany		
	<b>Cables</b>				
	<b>Accessories</b>	\$395,715.00	\$652,484.00	\$899,079.00	\$1,947,278.00
	<b>Remotes</b>	\$454,340.00	\$816,455.00	\$1,074,748.00	\$2,345,543.00
	<b>Cables</b>	\$320,270.00	\$575,028.00	\$766,284.00	\$1,661,582.00
	<b>Accessories</b>	<b>\$3,087,727.00</b>	<b>\$4,771,606.00</b>	<b>\$7,126,467.00</b>	<b>\$14,985,800.00</b>
<b>Sounds</b>	<b>Receivers</b>	\$1,365,228.00	\$2,018,599.00	\$2,579,830.00	\$5,963,657.00
	<b>CD Players</b>	\$731,145.00	\$1,209,855.00	\$1,199,367.00	\$3,140,367.00
	<b>Speakers</b>	\$886,344.00	\$1,537,376.00	\$2,114,998.00	\$4,538,718.00
	<b>Bookshelf Stereos</b>	\$583,281.00	\$1,145,299.00	\$1,671,423.00	\$3,400,003.00
	<b>Cassette Recorders</b>	\$752,449.00	\$1,438,654.00	\$2,025,994.00	\$4,217,097.00
	<b>Sounds</b>	<b>\$4,318,447.00</b>	<b>\$7,349,783.00</b>	<b>\$9,591,612.00</b>	<b>\$21,259,842.00</b>
<b>E-Link</b>	<b>GPS</b>	\$2,289,142.00	\$2,947,832.00	\$4,563,875.00	\$9,800,849.00
	<b>MP3 Players</b>	\$1,492,083.00	\$1,924,116.00	\$3,091,206.00	\$6,507,405.00
	<b>PDA's</b>	\$930,028.00	\$1,082,420.00	\$1,444,762.00	\$3,457,210.00
	<b>E-Link</b>	<b>\$4,711,253.00</b>	<b>\$5,954,368.00</b>	<b>\$9,099,843.00</b>	<b>\$19,765,464.00</b>

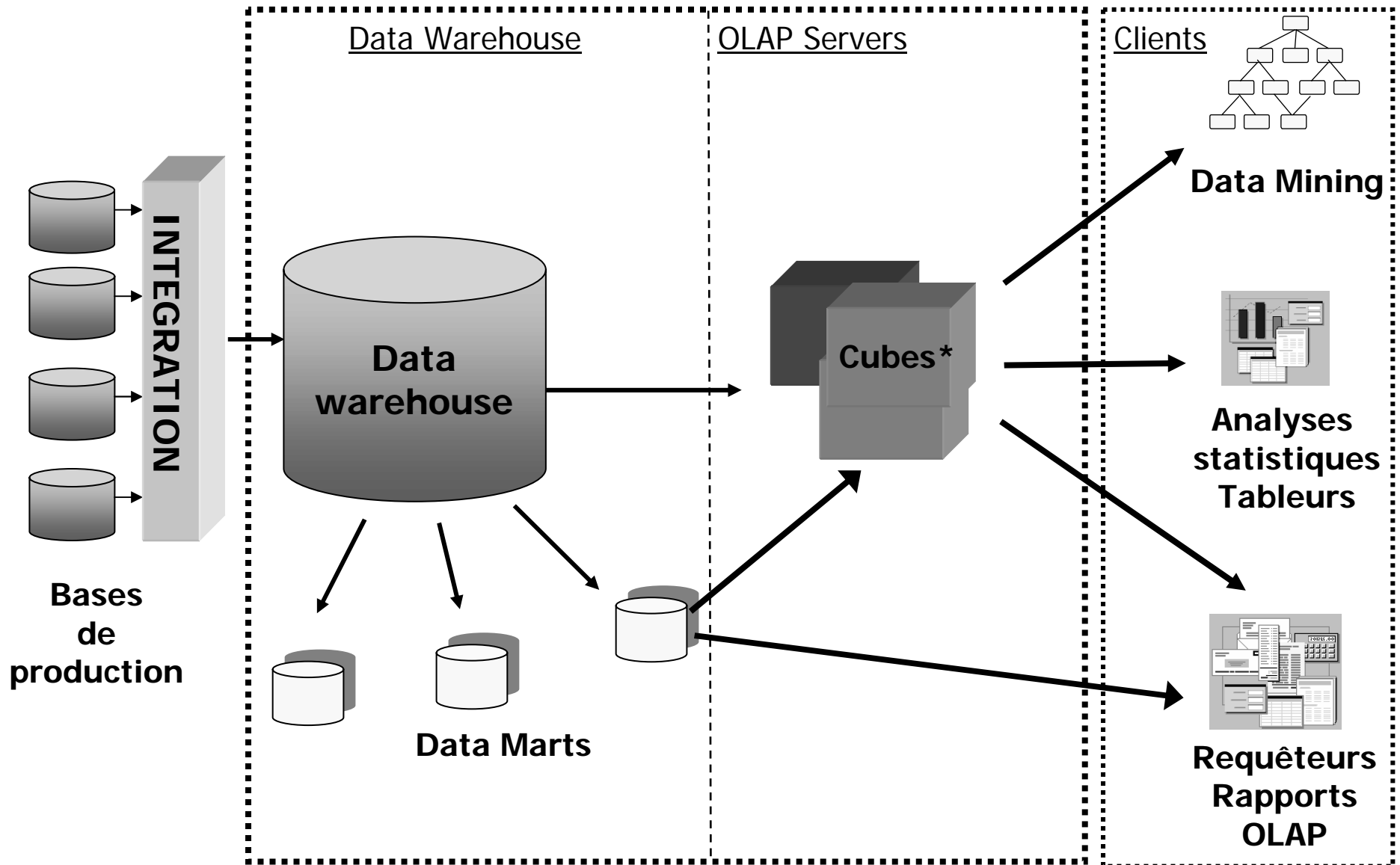
Bottom Panel: [Grid] [Table] [Chart] [Print] [Zoom] [Refresh] [Help] [Close]

# 2 formes de data warehouse

---

- Datawarehouse
  - ◆ Contient les données concernant l'ensemble des composantes de l'organisation
  - ◆ Plusieurs BD opérationnelles et sources extérieures
  
- Data mart
  - ◆ Un sous-ensemble de l'entrepôt global concernant un groupe spécifique d'utilisateurs
    - Ex : datamart concernant les ressources humaines
  - ◆ Peut être dépendant ou non de l'entrepôt de données

# Architecture



\*MOLAP, ROLAP, HOLAP

# Datawarehouse - outils

---

- OLAP:

- ◆ Monde BD : Oracle (Express server), Informix, IBM DB2 (Arbor software)
- ◆ Monde décisionnel: Essbase, Business-object, Powerplay, MMDB (SAS institute )

- Restitution:

- ◆ Impromptu (rapports),
- ◆ Dbminer (data mining), ...

# Références

---

- Transparents:
  - ◆ Sophie Montiès: "Data Warehousing"
  - ◆ M-A Aufaure: "Les entrepôts de données"
  - ◆ Enrico Franconi: "Data Warehouse Models and OLAP Operations"
  - ◆ Hector Garcia-Molina : "Data Warehousing and OLAP"
- Livre ou chapitre de livres:
  - ◆ J-F. Goglin: "La construction du data warehouse", édition Hermès
  - ◆ G. Gardarin: "Internet/intranet et bases de données", chapitre 4