

# OLAP: On Line Analytical Processing

Berghida Meryem, UMKB, 2019/2020

# Entrepôt et OLAP

✎ un entrepôt de données (ED) contient des données nombreuses, homogènes, exploitables, multidimensionnelles, consolidées

✎ comment exploiter ces données à des fins d'analyse ?

✎ traditionnellement : les requêtes OLTP sont exécutées sur les données sources

✎ l'ED est mis à jour chaque nuit

✎ les requêtes OLAP sont exécutées sur les données de l'ED

✎ analyser les données d'un ED c'est :

✎ résumer

✎ consolider

✎ observer

✎ appliquer des formules statistiques

✎ synthétiser des données selon plusieurs dimensions

# OLTP versus OLAP

- **OLTP (On Line Transaction Processing) :**
  - ✎ Les applications OLTP sont des applications opérationnelles (de production), constituées de traitements factuels concernant les produits, les ressources ou les clients de l'entreprise
  - ✎ Les requêtes OLTP sont exécutées sur les données sources
- **OLAP (On Line Analytical Processing) :**
  - ✎ Les applications OLAP sont des applications d'aide à la décision
  - ✎ Elles sont constituées de traitements ensemblistes réduisant une population à une valeur ou un comportement.
  - ✎ Les requêtes OLAP sont exécutées sur l'ED
- **Le terme OLAP désigne :**
  - ✎ L'ensemble des **moyens** et **techniques** à mettre en œuvre pour réaliser des systèmes d'aide à la décision efficaces
  - ✎ Des **traitements semi-automatiques** visant à **interroger**, **visualiser** et **synthétiser** les **données**, traitements **définis** et **mis en œuvre par les décideurs**
  - ✎ **On-Line** : signifie que le processus se fait **en ligne**, l'utilisateur doit avoir la **réponse** de façon **quasi-instantanée**

# Exemple ED

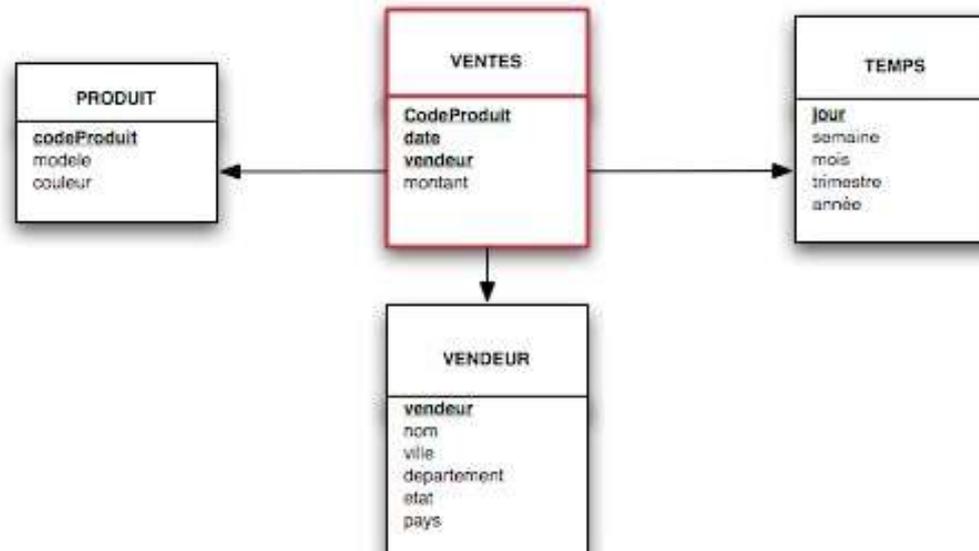
- **Soit l'entrepôt en schéma étoile suivant :**

✎ ventes(codeProduit, date, vendeur, montant) (*table faits*)

✎ produits(codeProduit, modèle, couleur) (*table dimension*)

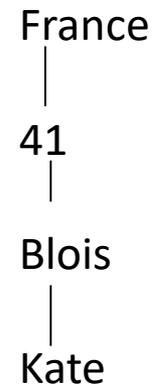
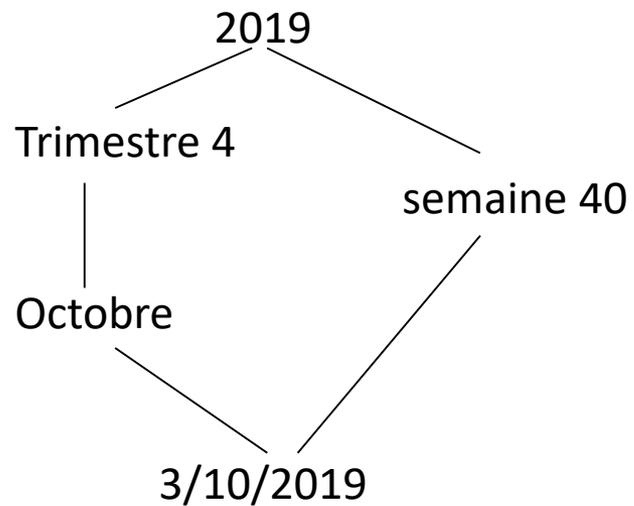
✎ vendeurs(nom, ville, département, état, pays) (*table dimension*)

✎ temps(jour, semaine, mois, trimestre, année) (*table dimension*)



# Exemple ED

- Hiérarchies:

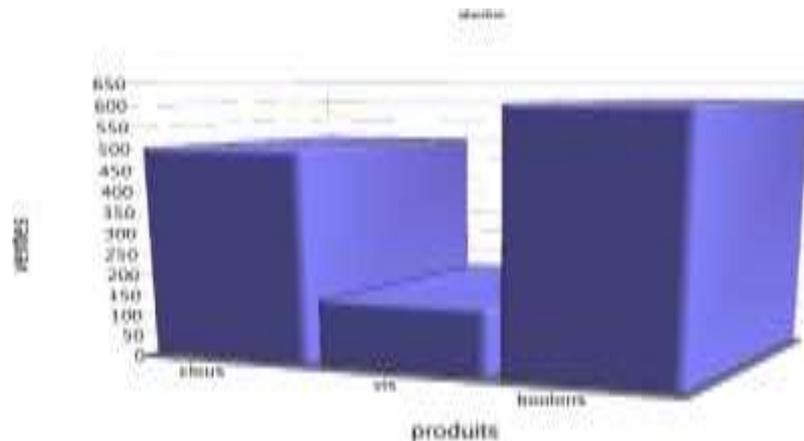


# Analyse des besoins

- **Analyse des ventes de divers produits**
- Exemple de questions associées :
  - ✎ Quels sont les produits dont les ventes ont chuté l'an dernier?
  - ✎ Quelles sont les quinze meilleures ventes par magasin et par semaine durant le premier trimestre de l'année 2001?
  - ✎ Quelle est la tendance des chiffres d'affaire (CA) par magasin depuis 3 ans?
  - ✎ Quelles prévisions peut-on faire sur les ventes d'une catégorie de produits dans les 6 mois à venir ?

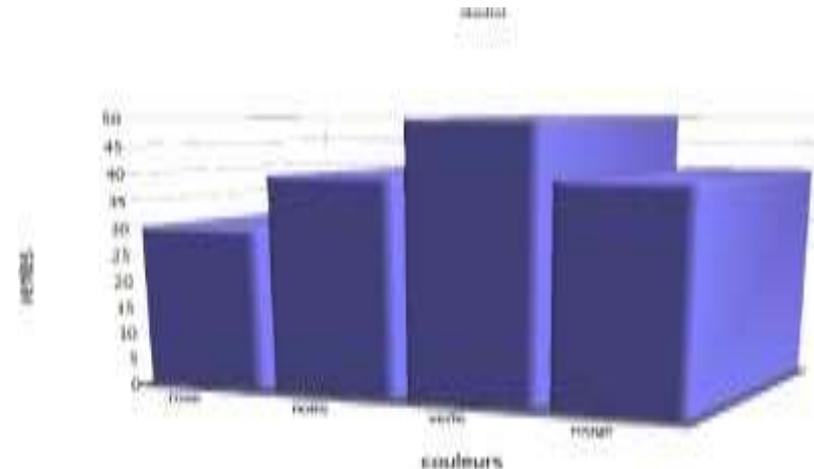
# Exemple d'analyse

- Analyse des ventes de divers produits :  
SELECT modele, SUM(montant)  
FROM ventes, produits  
WHERE ventes.codeProduit = produits.codeProduit  
GROUP BY modele ;



# Exemple d'analyse

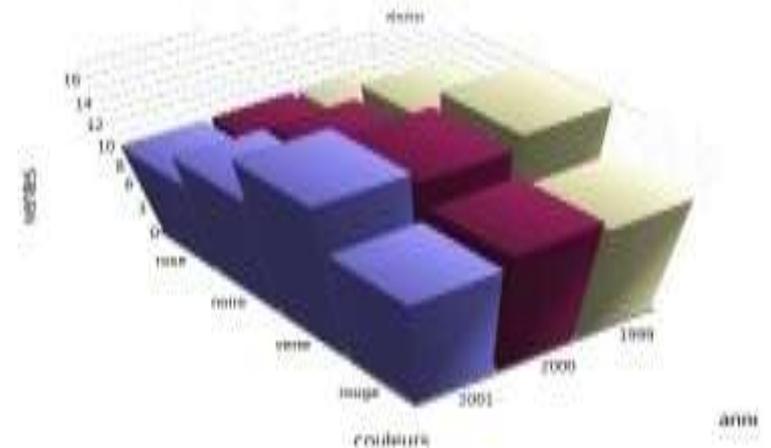
- Les ventes de vis sont plus faibles que prévu... quelles couleurs sont responsables ?  
SELECT couleur, SUM(montant)  
FROM ventes, produits  
WHERE ventes.codeProduit = produits.codeProduit  
AND modele = "vis"  
GROUP BY couleur ;



# Exemple d'analyse

- Les ventes de vis sont plus faibles que prévu...  
quelles années sont responsables ?

```
SELECT couleur, annees, SUM(montant)
FROM ventes, produits, temps
WHERE ventes.codeProduit = produits.codeProduit
AND ventes.date = temps.jour
AND modele = "vis"
GROUP BY couleur, annees
```



# Exemple d'analyse

- Les ventes de vis sont plus faibles que prévu...

Quels trimestres sont responsables ?

```
SELECT couleur, trimestre, SUM(montant)
```

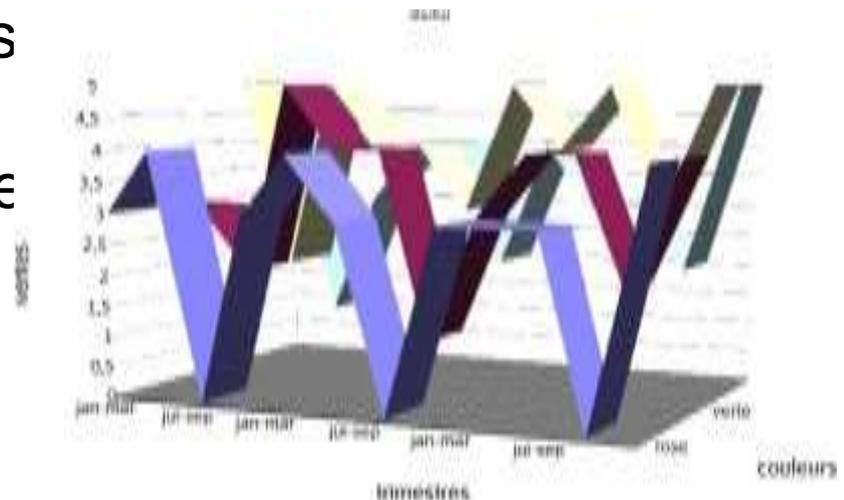
```
FROM ventes, produits, temps
```

```
WHERE ventes.codeProduit = produits.codeProduit
```

```
AND ventes.date = temps
```

```
AND modele = "vis"
```

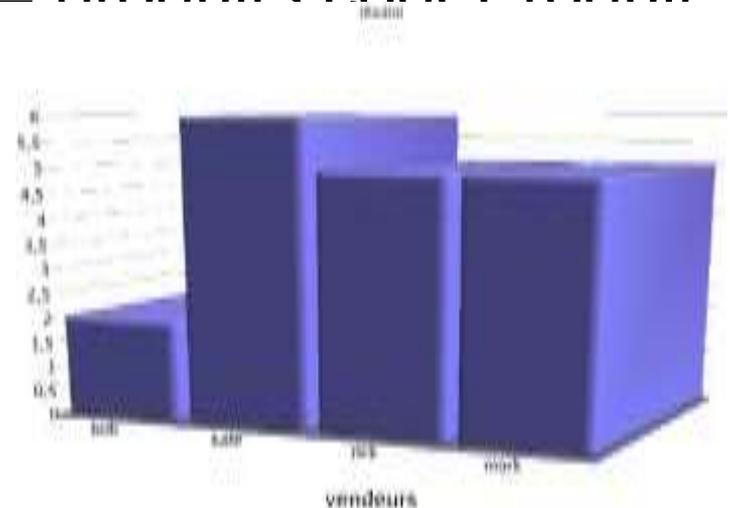
```
GROUP BY couleur, trime
```



# Exemple d'analyse

- Les ventes de vis sont plus faibles que prévu...  
Quels vendeurs sont responsables ?

```
SELECT vendeur, somme
FROM(
SELECT trimestre, vendeur, SUM(montant) as
somme
FROM ventes, produits, temps, vendeur
WHERE ventes.codeProduit = produits codeProduit
AND ventes.date = temps.j
AND ventes.vendeur = ven
AND modele = "vis"
GROUP BY trimestre, venci
WHERE trimestre = "jui-sep"
```



# Exemples de traitements

- **Quels sont les résultats cumulés des vendeurs par mois ?**  
SELECT vendeur, mois, CSUM(resultat,vendeur,mois) as cumul  
FROM (  
SELECT vendeur, mois, Sum(montant) as resultat  
FROM ventes, produits, temps  
WHERE ventes.codeProduit = produits.codeProduit  
AND ventes.date = temps.jour  
AND modele = "vis"  
AND couleur = "rose"  
GROUP BY mois, vendeurs)  
ORDER BY mois ;  
**Quelle est l'évolution de la moyenne des ventes pour une fenêtre de 2 jours ?**  
SELECT date, montant, MAVG(montant,2,date) as moy  
FROM ventes, temps  
WHERE ventes.date = temps.jour  
AND annee = 2001  
ORDER BY date ;

# Problématique de l'OLAP

- Supporter des opérations “tableur” sur des BD de plusieurs Go

## **Besoins spécifiques :**

- langages de manipulation
- organisation des données
- fonctions d'agrégation

## **Organisation des données proche des abstractions de l'analyste :**

- selon plusieurs dimensions
- selon différents niveaux de détail
- en ensemble
- donnée = point dans l'espace associé à des valeurs

# De la table au cube

De la table ...

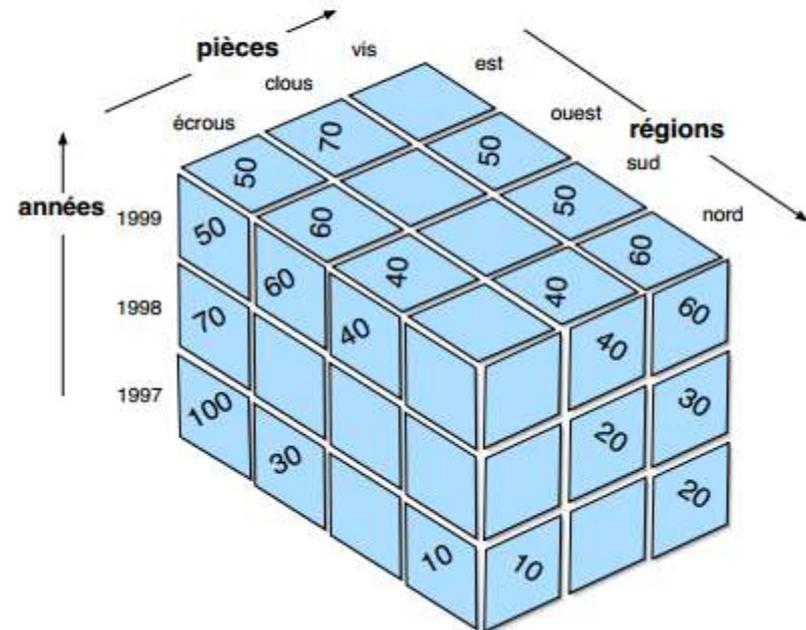
Table Ventes :

VENTES	pièces	Régions	Années	quantités
	écrous	est	1999	50
	clous	est	1997	100
	vis	ouest	1998	50
	...	...	...	...
	écrous	est	total	220
	...	...	...	...
	écrous	total	total	390
	...	...	...	...
	total	total	total	1200

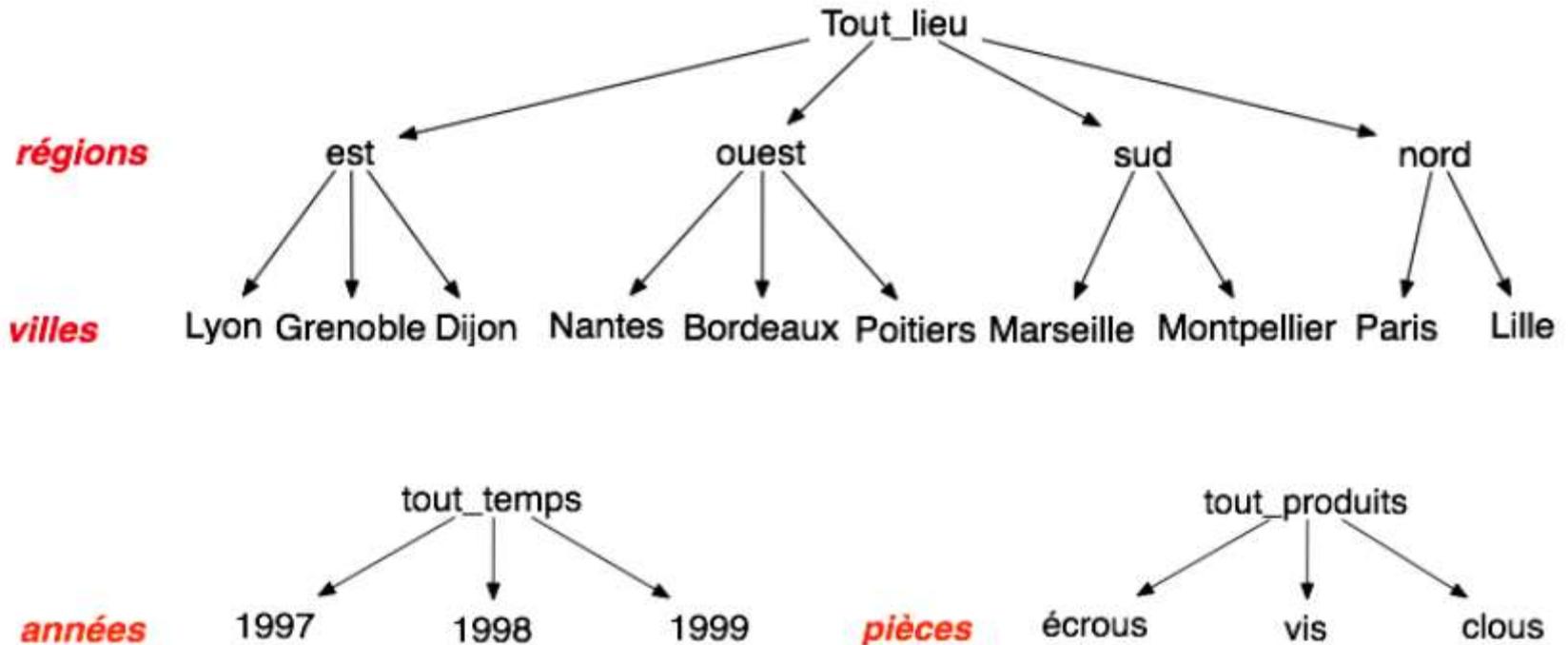
(pièce, région, année) → quantité

... au cube

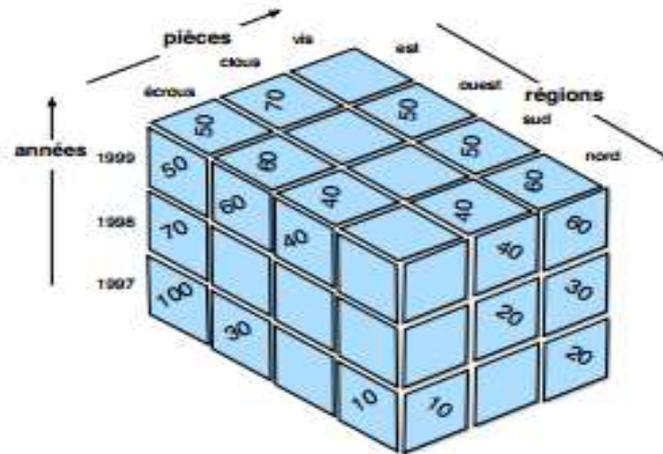
Cube Ventes :



# Hiérarchies de granularité



# Terminologie autour du cube



Terme	Valeur
Cube	Ventes
Cellule	écrous, est, 1997, 100
Référence	écrous, est, 1997
mesure	100
Membre/paramètre	est
dimension	lieu
niveau	région

# **OPÉRATIONS ÉLÉMENTAIRES**

## **OLAP**

# Catégories d'opérations OLAP

- **3 catégories d'opérations élémentaires :**
- **Restructuration** : concerne la représentation, permet un changement de points de vue selon différentes dimensions : opérations liées à la structure, manipulation et visualisation du cube :
  - ✎ **Rotate/pivot**
  - ✎ **Switch**
  - ✎ **Split, nest, push, pull**
- **Granularité** : concerne un changement de niveau de détail : opérations liées au niveau de granularité des données :
  - ✎ **roll-up,**
  - ✎ **drill-down**
- **Ensemble** : concerne l'extraction et l'OLTP classique :
  - ✎ **slice, dice**
  - ✎ **selection**
  - ✎ **projection**
  - ✎ **jointure (drill-across)**

# Opérations de restructuration

- Permettent un **changement de points de vue**, une **réorientation selon différentes dimensions** de la vue multidimensionnelle
- Opérations liées à la **structure**, la **manipulation** et la **visualisation** du cube :
  - ✎ **réorientation** :
    - ✎ sélection graphique
    - ✎ flexibilité du schéma
    - ✎ membres complexes
    - ✎ symétrie membres/mesures
  - ✎ **manipulations** :
    - ✎ bijectives
    - ✎ relatives
    - ✎ à niveau d'information constant
- **Opérations de restructuration** :
  - ✎ **rotate/pivot**
  - ✎ **switch**
  - ✎ **split, nest, push, pull**

# Opérations de restructuration

- **Rotate ou Pivot :**

- ✎ effectuer à un cube une **rotation autour d'un de ses trois axes passant par le centre de 2 faces opposées**, de façon à présenter un ensemble de faces différent

- ✎ une sorte de **sélection de faces** et non des membres.

- **Switch ou permutation :**

- ✎ consiste à **inter-changer** la **position** des **membres** d'une **dimension**.

- **Split ou division :**

- ✎ consiste à **présenter chaque tranche du cube** et de passer d'une présentation tridimensionnelle d'un cube à sa présentation sous la forme d'un **ensemble de tables**

- ✎ sa généralisation permet de **découper un hypercube** de dimension 4 **en cubes**.

- **Nest ou l'emboîtement :**

- ✎ **imbrication des membres à partir du cube**.

- ✎ Permet de grouper sur une même représentation bi-dimensionnelle toutes les informations (mesures et membres) d'un cube quelque soit le nombre de ses

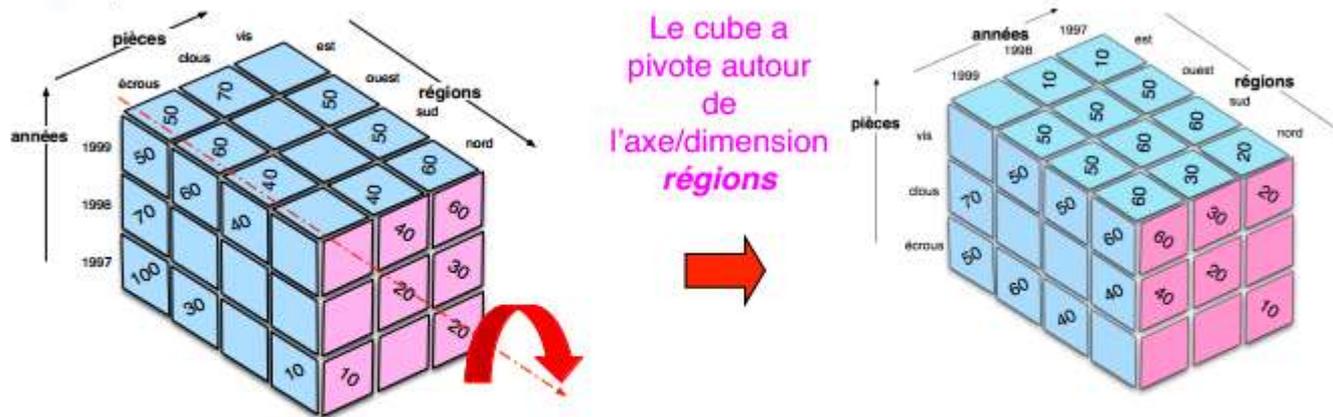
- dimensions.

- **Push ou l'enfoncement :**

- ✎ consiste à **combinaison des membres d'une dimension** aux mesures du cube, i.e. de faire passer des membres comme contenu de cellules.

# Opérations de restructuration: Rotate/Pivot

**Rotate/pivot** : effectue au cube une rotation autour d'un de ses 3 axes passant par le centre de 2 faces opposées, de façon à présenter un ensemble de faces différent (sélection de faces)



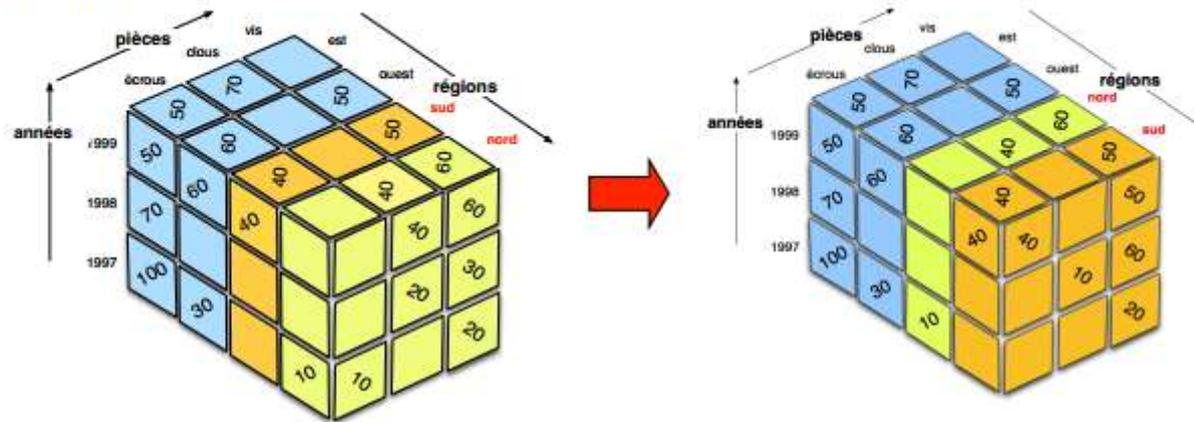
la visualisation résultante est souvent 2D :

nord	1999	1998	1997
vis	60	30	20
clous	40	20	
écrous			10

vis	1999	1998	1997
est		10	10
ouest	50	50	50
sud	50	60	60
nord	60	30	20

# Opérations de restructuration: Switch

**Switch** ou **permutation** : consiste à interchanger la position des membres d'une dimension :



Ici sont interchangés les membres *nord* et *sud* de la dimension *régions*

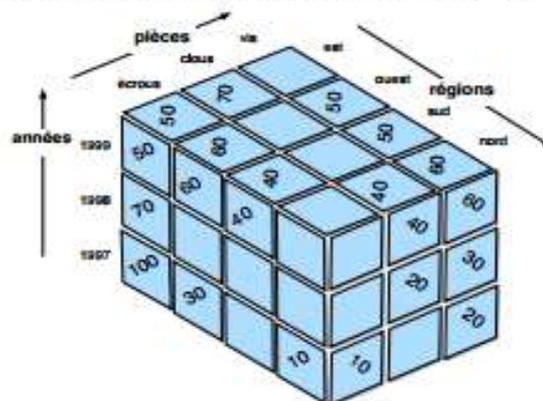
la visualisation résultante est souvent 2D :

nord	1999	1998	1997
vis	60	30	20
clous	40	20	
écrous			10

sud	1999	1998	1997
vis	50	60	60
clous		10	
écrous	40	20	

# Opérations de restructuration: Split

**Split ou division** : consiste à présenter chaque tranche du cube et de passer de sa présentation tridimensionnelle à sa présentation sous la forme d'un ensemble de tables.



ici un **split(region)** du cube Ventes conduit aux 4 tables suivantes :

ventes est	1999	1998	1997
écrous	50	70	100
vis		10	10
clous	70	70	100

ventes ouest	1999	1998	1997
écrous		10	30
vis	50	50	50
clous		10	40

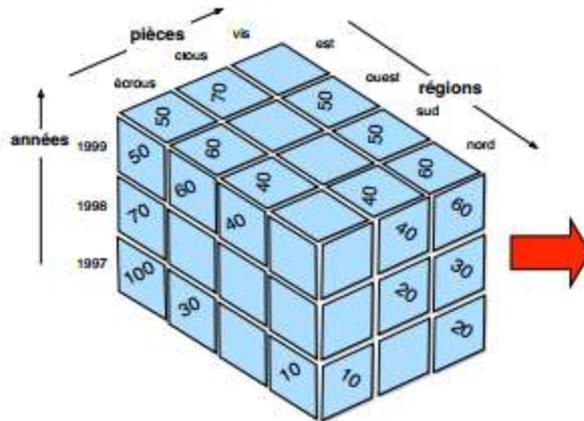
ventes sud	1999	1998	1997
écrous	40	20	
vis	50	60	60
clous		10	

ventes nord	1999	1998	1997
écrous			10
vis	60	30	20
clous	40	20	

# Opérations de restructuration: nest

**Nest ou l'emboîtement:** permet d'imbriquer des membres à partir du cube. L'intérêt de cette est qu'elle permet de grouper sur une même représentation bi-dimensionnelle toutes les informations (mesures et membres) d'un cube quelque soit le nombre de ses dimensions

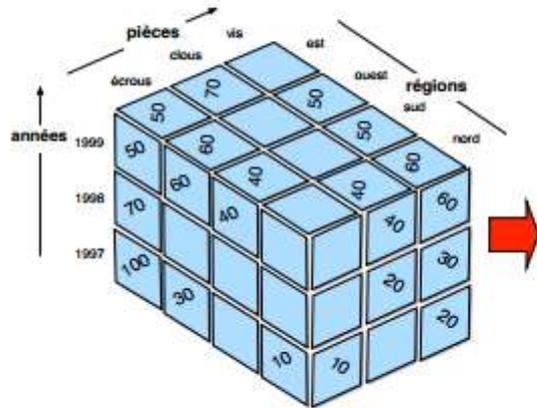
**nest(pièces, région) :**



ventes nest		1999	1998	1997
écrous	est	50	70	100
	ouest		10	30
	nord			10
	sud	40	20	
vis	est		10	10
	ouest	50	50	50
	nord	60	30	20
	sud	50	60	60
clous	est	70	70	100
	ouest		10	40
	nord	40	20	
	sud		10	

# Opérations de restructuration: push

**Push ou l'enfoncement:** consiste à combiner les membres d'une dimension aux mesures du cube, i.e. de faire passer des membres comme contenu de cellules.



**push(année) :**

ventes push	est	ouest	nord	sud
écrous	1999 50			1999 40
	1998 70	1998 10		1998 20
	1997 100	1997 30	1997 10	
vis		1999 50	1999 60	1999 50
	1998 10	1998 50	1998 30	1998 60
	1997 10	1997 50	1997 20	1997 60
clous	1999 70		1999 40	
	1998 70	1998 10	1998 20	1998 10
	1997 100	1997 40		

# Opérations de granularité

## Granularité :

- hiérarchisation de l'information en différents niveaux de détails appelés **niveaux de granularité**.
- un niveau est un ensemble nommé de **membres**
- le **niveau le plus bas est celui de l'entrepôt**

Des opérations **d'agrégation successives** sur ces données permettent **de nouveaux points de vue de moins en moins détaillés de l'information** et constituent autant de **niveaux supérieurs** :

### ▪ navigation entre les niveaux :

- groupements
- agrégation

### ▪ manipulations :

- relatives
- nécessitant des informations non contenues dans le cube de départ

# Opérations de granularité

## Opérations de granularité :

- **roll-up,**
- **drill-down**

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**.
- une **fonction d'agrégation** (somme, moyenne, etc) en **paramètre** de l'**opération** indique **comment sont calculés les valeurs du niveau supérieur** à partir de celles du niveau inférieur

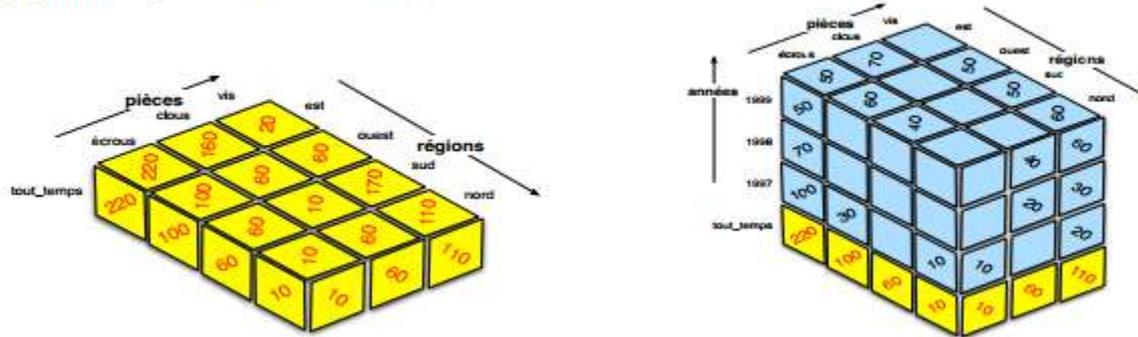
### **Drill-down** ou forage vers le bas :

- consiste à **représenter** les **données** du **cube** à un **niveau de granularité de niveau inférieur**, donc sous une forme **plus détaillée** (selon la hiérarchie définie de la dimension)



# Opérations de granularité: Roll-up

**roll-up(annee) : Ventes 97-99**



**roll-up(annees, pieces) : la visualisation est souvent 2D :**

nord	1999	1998	1997	tout_temps
vis	60	30	20	110
clous	40	20		60
écrous			10	10
tout_produit	100	50	30	180

**Remarque :** une fonction d'agrégation (somme, moyenne, ...) en paramètre de l'opération indique comment sont calculés les valeurs du niveau supérieur à partir de celles du niveau inférieur

# Opérations de granularité: Drill-down

**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.

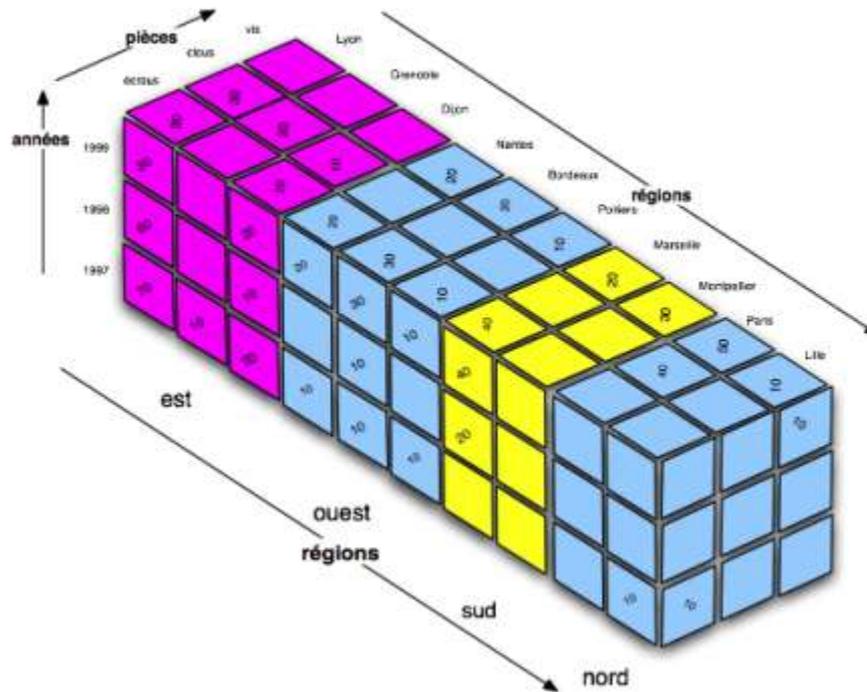
- opération **réciproque de roll-up**, drill-down permet d'**obtenir des détails sur la signification d'un résultat** en affinant une dimension ou en ajoutant une dimension
- opération **coûteuse** d'où son intégration dans le système

Exemple : un chiffre d'affaire suspect pour un produit donné :

- ajouter la dimension temps : envisager l'effet week-end
- ajouter la dimension magasin: envisager l'effet géographique

# Opérations de granularité: Drill-down

Drill-down du niveau des régions au niveau villes : Drill-down(regions) :



# Opérations ensemblistes

## Objet des opérations ensemblistes :

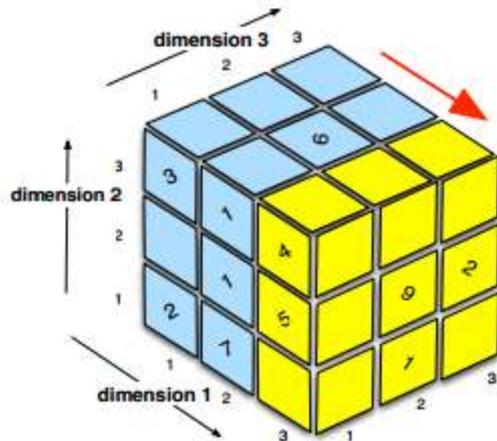
- concernent l'extraction
- manipulations classiques
- extension à plusieurs dimensions

## Opérations OLAP ensemblistes :

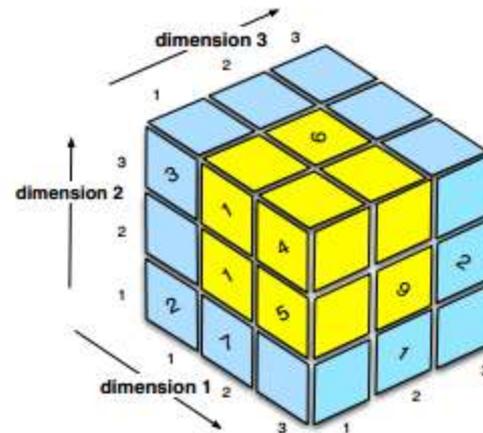
- **slice et dice (sélection et projection)**
- **drill-across (jointure)**

# Opérations ensembliste: Slide et Dice

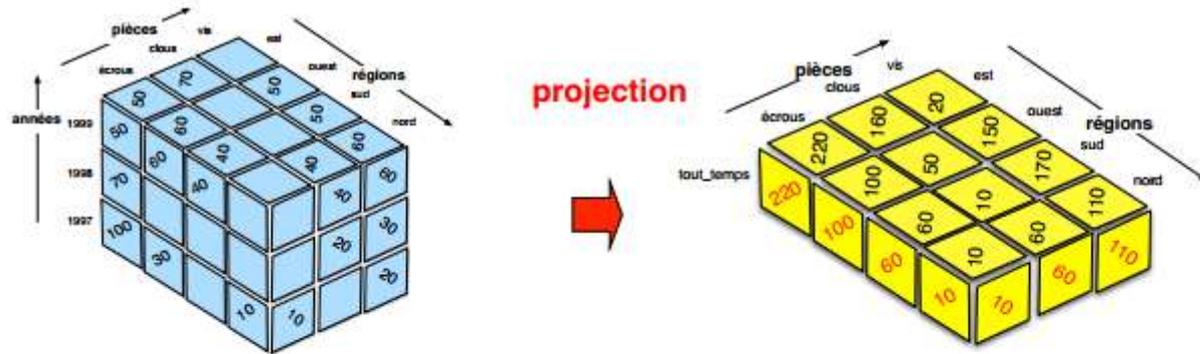
**slide** : correspond à une **projection** selon une dimension du cube :



**dice** : correspond à une **sélection** du cube :



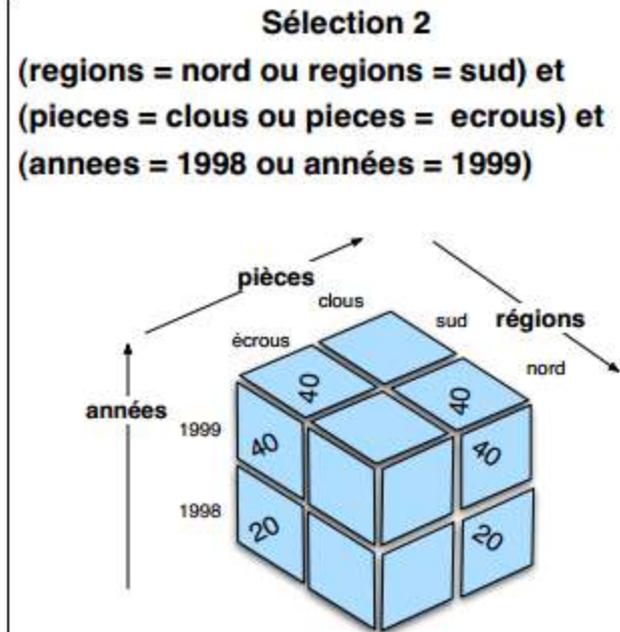
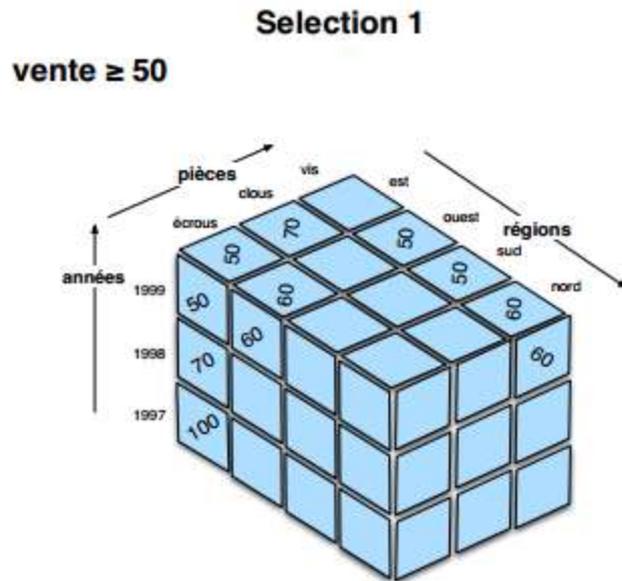
# Opérations ensembliste: projection (slide)



$\Pi$  pièce, région :

ventes 97-99	est	ouest	sud	nord
écrous	220	100	60	10
clous	160	50	10	60
vis	20	150	170	110

# Opérations ensembliste: sélection (dice)



# Opérations ensemblistes: jointures (drill-across)

ventes 97-99 ⋈

prix	97-99
écrous	1
clous	0.7
vis	0.8

=

ventes 97-99	est	ouest	sud	nord
écrous	220 1	100 1	60 1	10 1
clous	160 0.7	50 0.7	10 0.7	60 0.7
vis	20 0.8	150 0.8	170 0.8	110 0.8