

### Informatique Décisionnelle (BI)

### ABDELTIF EL BYED

Professeur à l'université Hassan II – Faculté des sciences Ain Chok (FSAC)

<u>Aelbyed@gmail.com</u> / <u>Abdeltif.elbyed@univh2c.ac.ma</u>

Site: https://sites.google.com/site/aelbyed

Année: 2019-2020

# Plan général du Module

### Partie1: Informatique Décisionnelle

- 1. Introduction à l'Informatique Décisionnelle (BI)
- Entrepôt de données (DW)
- 3. OLAP: On-Line Analytical Processing
- Les système OLAP: ROLAP, MOLAP & HOLAP

### Partie 2: IBM Cognos Report Studio (En Anglais)

- 1. Introduction to the Reporting Application
- Create List Reports
- Focus Reports Using Filters
- Create Crosstab Reports
- Present Data Graphically
- Focus Reports Using Prompts

### Partie 3: Data Mining

- Objectifs de la fouille de données
- Les algorithemes de data mining
- 3. Le DW et le data mining

IA et l'apprentissage A.EL BYED: OLAP

# Plan général du Module

### Partie1: Informatique Décisionnelle

- 1. Introduction à l'Informatique Décisionnelle (BI)
- 2. Entrepôt de données (DW)
- 3. OLAP: On-Line Analytical Processing
- 4. Les système OLAP: ROLAP, MOLAP & HOLAP

### Partie 2: IBM Cognos Report Studio (En Anglais)

- Introduction to the Reporting Application
- 2. Create List Reports
- 3. Focus Reports Using Filters
- 4. Create Crosstab Reports
- 5. Present Data Graphically
- 6. Focus Reports Using Prompts

2

A.EL BYED: OLAP

# Plan général du Module

### Partie1: Informatique Décisionnelle

- 1. Informatique Décisionnelle Introduction
- 2. Entrepôt de données
- 3. OLAP: On-Line Analytical Processing
- 4. DW et les système OLAP

### Partie 2: IBM Cognos Report Studio (En Anglais)

- 1. Introduction to the Reporting Application
- 2. Create List Reports
- 3. Focus Reports Using Filters
- 4. Create Crosstab Reports
- 5. Present Data Graphically
- 6. Focus Reports Using Prompts

4



# CH3 OLAP: On-Line Analytical Processing

**Abdeltif EL BYED** 

Professeur à l'université Hassan II Abdeltif.elbyed@univh2c.ac.ma



### Introduction et problématique de l'OLAP

- 1. Entrepôt et OLAP
- 2. OLAP versus OLTP
- 3. Exemple d'analyses d'un entrepôt
- 4. Problématique de l'OLAP

### Plan

### 1. Introduction et problématique de l'OLAP

- 1. Entrepôt et OLAP
- 2. OLAP versus OLTP
- 3. Exemple d'analyses d'un entrepôt
- 4. Problématique de l'OLAP

### 2. Opérations élémentaires OLAP

- 1. Catégories d'opérations OLAP
- 2. Opérations de restructuration : rotate, switch, split, nest, push, pull
- 3. Opérations de granularité : roll-up, drill-down
- 4. Opérations ensemblistes : slide, dice, jointure(drill-across), data cube
- Modèles et langages pour l'OLAP
- 6. Les règles de Codd pour les produits OLAP
- 7. Problématique de la modélisation logique d'un ED



A.EL BYED: OLAP

### ED & 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

8

### **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 quasiinstantanée

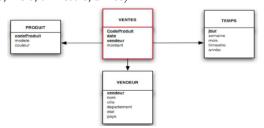
9

A.EL BYED: OLAP

# ED: Exemple

### Soit l'entrepôt en schéma étoile suivant :

- table faits
  - ventes(codeProduit, date, vendeur, montant)
- Tables dimension
  - produits(codeProduit, modèle, couleur)
  - vendeurs(nom, ville, département, état, pays)
  - temps(jour, semaine, mois, trimestre, année)



### **OLTP versus OLAP**

	Caractéristiques	OLTP	OLAP
Conception	Orientation	Transaction	Analyse
	Conception	Entité-Relation	Etoile/flocon
Données	Granularité	Détail	Résumées, agrégées
	Nature	Relationnelle	Multidimentionnelle
	Actualisation	Actualisées, mises à jour	Historisées, recalculées
	Taille	100 Mo/Go	100 Go/To
Traitements	Unité de travail	Transaction simple	Requête complexe
	Accés	Lecture/écriture	Lecture
	Nb de tuples accédés	Dizaines	Millions
	Métrique	Débit de transactions	Temps de réponse
Utilisateurs	Utilisateur	Agent opérationnel	Analyste/décideur
	Nombre d'utilisateurs	Milliers	Centaines

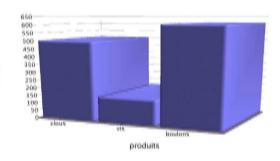
# Besoins d'analyse

- 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 (1)

• Analyse des ventes de divers produits :

SELECT modele, SUM(montant) FROM ventes, produits  $WHERE\ ventes.codeProduit = produits.codeProduit$ GROUP BY modele

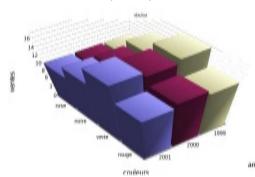


A.EL BYED: OLAP

# Exemple d'analyse (3)

• Les ventes de vis sont plus faibles que prévu... quelles années et quelles couleurs 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 (2)

• Les ventes de vis sont plus faibles que prévu... quelles couleurs sont responsables?

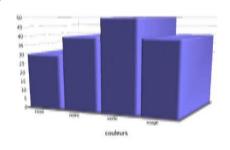
SELECT couleur, SUM(montant)

FROM ventes, produits

WHERE ventes.codeProduit = produits.codeProduit

AND modele = "vis"

GROUP BY couleur

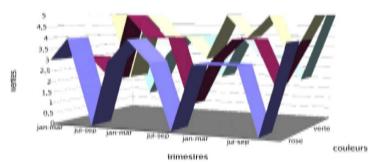


A.EL BYED: OLAP

# Exemple d'analyse (4)

• 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.jour AND modele = "vis" GROUP BY couleur, trimestre



# Exemple d'analyse (5)

 $\bullet \ \ Les \ ventes \ de \ vis \ sont \ plus \ faibles \ au \ 3 \ trimestre. \ Quels \ vendeurs \ sont \ responsables \ ?$ 

```
SELECT vendeur, somme
```

FROM(

SELECT trimestre, vendeur, SUM(montant) as somme FROM ventes, produits, temps, vendeur

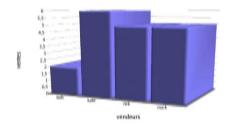
WHERE ventes.codeProduit = produits.codeProduit

AND ventes. date = temps. jour

AND ventes. vendeur = vendeurs. nom AND modele = "vis"

**GROUP BY** trimestre, vendeur)

**WHERE** trimestre = "jui-sep";



17

A.EL BYED: OLAP

# Problématiques d'OLAP

- Supporter des opérations "tableur" sur des BD de plusieurs Go (Chaudhuri et Dayal 97)
- Besoins spécifiques :
  - langages de manipulation
  - organisation des données
  - fonctions d'agrégation
  - ...
- Organisation des données proche des abstractions de l'analyste :
  - $\bullet\,$  selon plusieurs dimensions
  - selon différents niveaux de détail
  - en ensemble
  - donnée = point dans l'espace associé à des valeurs

# Exemple d'analyse (6)

• Quels sont les résultats cumulés des vendeurs par mois pour les vis rose ?

SELECT vendeur, mois, CSUM(resultat, vendeur, mois) as cumul

FROM (

SELECT vendeur, mois, Sum(montant) as resultat

FROM ventes, produits, temps, vendeur

WHERE ventes.codeProduit = produits.codeProduit AND ventes.vendeur = vendeurs.nom

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 pour l'année 2001?

SELECT date, montant, MAVG(montant, 2, date) as moy

FROM ventes, temps

WHERE ventes.date = temps.jour AND annee = 2001

ORDER BY date;

18

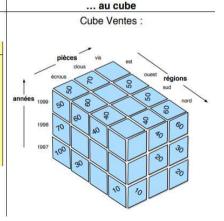
A.EL BYED: OLAP

### De la table au Cube

### Table Ventes: VENTES pièces Années 1999 50 1997 100 clous est vis 1998 50 ouest 220 écrous total écrous total

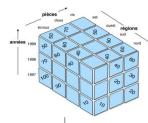
De la table ...

(pièce, région, année) → quantité



# Hiérarchies de granularité régions villes Lyon Grenoble Dijon Nantes Bordeaux Poitiers Marseille Montpellier Paris Lille tout temps années 1997 1998 1999 pièces écrous vis clous

# Terminologies



Terme	Valeur
Cube	Ventes
Cellule	ecrous, est, 1997, 100
Référence	ecrous, est, 1997
mesure	100
Membre/paramètre	est
dimension	lieu
niveau	région
	3

22

A.EL BYED: OLAP

### Plan

- 1. Introduction et problématique de l'OLAP
  - 1. Entrepôt et OLAP
  - 2. OLAP versus OLTP
  - 3. Exemple d'analyses d'un entrepôt
  - 4. Problématique de l'OLAP

### 2. Opérations élémentaires OLAP

- 1. Catégories d'opérations OLAP
- 2. Opérations de restructuration :
- orotate, switch, split, nest, push, pull
- 3. Opérations de granularité :
  - roll-up, drill-down
- Opérations ensemblistes :
  - slide, dice, jointure(drill-across), data cube
- Modèles et langages pour l'OLAP
- Les règles de Codd pour les produits OLAP
- 7. Problématique de la modélisation logique d'un ED



# Opérations élémentaires OLAP

- 1. Catégories d'opérations OLAP
- Opérations de restructuration : rotate, switch, split, nest, push, pull
- 3. Opérations de granularité : roll-up, drill-down
- Opérations ensemblistes : slide, dice, jointure(drill-across), data cube
- 5. Modèles et langages pour l'OLAP
- 6. Les règles de Codd pour les produits OLAP
- 7. Problématique de la modélisation logique d'un ED

# Catégories d'opérations OLAP

### Il ya 3 catégories d'opérations élémentaires :

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



A.EL BYED: OLAP

# 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 combiner les membres d'une dimension aux mesures du cube,
  - i.e. de faire passer des membres comme contenu de cellules.

### Opérations de restructuration

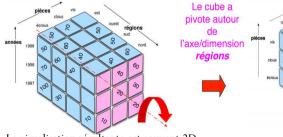
- Permettent un changement de points de vue, une réorientation selon différentes dimensions de la vue multidimensionnelle
- Opérations liées à: la structure, la manipulation et la visualisation du cube :
  - réorientation :
    - sélection graphique
    - flexibilité du schéma
    - membres complexes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



A.EL BYED: OLAP

# 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 :

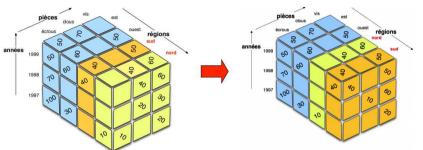
	nord	1999	1998	1997
	vis	60	30	20
	clous	40	20	
	écrous			10
A.EL BYED: OLA.				

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



La visualisation résultante est souvent 2D:

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

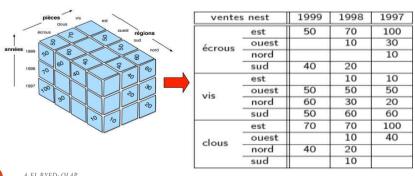
sud	1999	1998	1997
vis	50	60	60
clous		10	
écrous	40	20	

### A.EL BYED: OLAP

# Opérations de restructuration: Nest

Nest ou l'emboîtement: permet d'imbriquer des membres à partir du cube.

- 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.
- Exemple: nest(pièces, région):



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

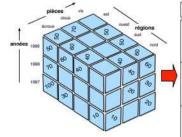
ventes ouest	1999	1998	1997
écrous		10	30
vis	50	50	50
clous		10	40

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

# 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.
- Exemple: push(année):



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

# Catégories d'opérations OLAP

### Il ya 3 catégories d'opérations élémentaires :

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



A.EL BYED: OLAP

# 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.
  - sous une forme plus détaillée (selon la hiérarchie définie de la dimension)

# Opérations de granularité

### Granularité:

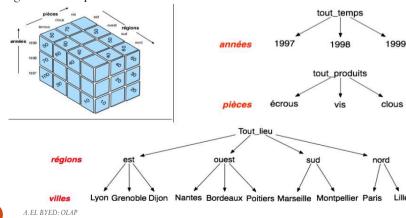
- hiérarchisation de l'information en différents niveaux de détails appelés niveaux de granularité.
- un niveau est un ensemble nommé de membres
- le niveau le plus bas est celui de l'entrepôt
- Des opérations d'agrégation successives sur ces données permettent de nouveaux points de vue de moins en moins 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



A.EL BYED: OLAP

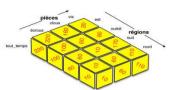
# Opérations de granularité : Roll-up

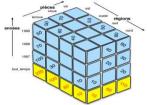
• *Roll-up* : consiste à représenter les données du cube à un niveau de granularité supérieur conformément à la hiérarchie définie sur la dimension.



# Roll-up (suite)

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



A.EL BYED: OLAP

# 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 : Cube

*L'opération CUBE* (représentation cubique généralisée du roll-up)

- Consiste à calculer tous les agrégats suivant tous les niveaux de toutes les dimensions :
- L'union de plusieurs group-by donne naissance à un cube :

Select ALL, ALL, ALL, Sum(quantité) From VENTES

Select pièces, ALL, ALL, Sum(quantité) From VENTES Group-By pièces;

Select pièces, années, ALL, Sum(quantité) FromVENTES Group-By pièces, années; IINION

Select pièces, années, régions, Sum(quantité) From VENTES Group-By pièces, années, régions;

 L'opérateur cube est une généralisation N-dimensionnelle de fonctions d'agrégations simples. C'est un opérateur relationnel:

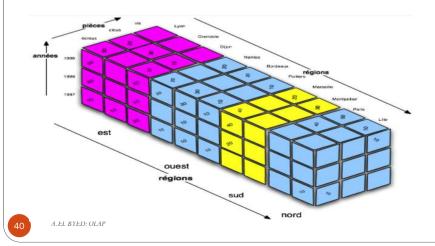
Select pièces, années, régions, Sum(quantité) From VENTES Group-By CUBE pièces, années, régions;

38

A.EL BYED: OLAP

# Opérations de granularité: Drill-down

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



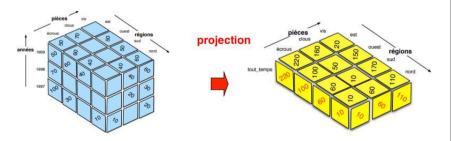
# Opérations ensemblistes

- Les opérations ensemblistes concernent :
  - L'extraction
  - Manipulations classiques
  - L'extension à plusieurs dimensions
- Opérations OLAP ensemblistes :
  - slice et dice (sélection et projection)
  - drill-across (jointure)

41 A.EL BYED: OLAP

# Opérations ensemblistes: Slice & Dice Slice: correspond à une projection selon une dimension du cube: Dice: correspond à une sélection du cube: dimension 2 dimension 2 dimension 1 dimension 1 A.EL BYED: OLAP

# Opérations ensemblistes: Slice



### Π piece, region:

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

Opérations ensemblistes: Dice

Selection 1

vente ≥ 50

Sélection 2

(regions = nord ou regions = sud) et (pieces = clous ou pieces = ecrous) et (annees = 1998 ou années = 1999)

A.EL BYED: OLAP

# Opérations ensemblistes: drill-across

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



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



A.EL BYED: OLAP

# Problème des opérations binaires

prix	97-99	р	rix	97-99
écrous	1	ь	oulons	8.0
clous	0.7	fo	orets	1.1
vis	0.8	Vi	is	0.7

Quelle mesure pour les vis ?

### **Exemple de traitements**

- Quels sont les 10 produits les plus performants ?
- Calculer la moyenne glissante des ventes par région et par pièces, pour une fenêtre de 2 années
- Calculer les prévisions de ventes pour les années 2000 `a 2002 avec comme hypothèse un accroissement annuels des ventes de 10%

