


# Business Intelligence

Enseignante: Emna Ammar Elhadjamor


2019-2020

# Chapitre 1



## Généralités sur le BI

2



## Plan

- 1 Enjeux du décisionnel
- 2 Exemple
- 3 Définition
- 4 Informatique opérationnelle et décisionnelle
- 5 Les phases d'un projet décisionnel

3

## 1-Enjeux du décisionnel

- La prise de décisions stratégiques dans une organisation nécessite le recours et le croisement de multiples informations qui concernent tous les départements.
- Or ces données sont généralement :
  - ⚠ éparpillées au sein des départements et non connectées entre elles
  - ⚠ hétérogènes dans leurs formats techniques et leurs organisations structurelles
  - ⚠ volatiles, au sens où leur mise à jour peut conduire à oublier des informations obsolètes

4

## 1-Enjeux du décisionnel

Les domaines d'utilisation de la BI touchent la plupart des métiers de l'entreprise :

- Finance, avec les reportings financiers et budgétaires par exemple ;
- Vente et commercial, avec l'analyse des points de ventes, l'analyse de la rentabilité et de l'impact des promotions par exemple ;

5

## 1-Enjeux du décisionnel

Les domaines d'utilisation de la BI touchent la plupart des métiers de l'entreprise :

- Marketing, avec la segmentation clients, les analyses comportementales par exemple ;
- Logistique, avec l'optimisation de la gestion des stocks, le suivi des livraisons par exemple ;
- Ressources humaines, avec l'optimisation de l'allocation des ressources par exemple ;
- ...

6

## 2-Exemple

- Un catalogue de produits sera conçu pour permettre de trouver facilement un produit en fonction de caractéristiques précises, de faire des mises à jour rapides et fiables, de gérer des stocks...
- Mais un système décisionnel souhaitera :  
1-connaître l'organisation des produits selon certaines caractéristiques et regroupements qui ne sont pas forcément premiers dans la gestion quotidienne;  
2-croiser le catalogue avec les ventes...

7

## 2-Exemple

**Besoin:** nouveaux besoins pour l'aide à la décision

**Qui:** les décideurs

**Les moyens pour y parvenir sont :**

- une information riche, pertinente, détaillée, historisée, fiable;
- des outils d'analyse et de restitution puissants et adaptés;
- des indicateurs qui favorisent le pilotage et favorisent l'action.

**Pourquoi et comment le chiffre d'affaire a baissé?**

**Quels sont les résultats des ventes par gamme de produit et par région pour l'année dernière ?**



8

### 3-Définition

Le système d'information décisionnel est un ensemble de données organisées de façon spécifiques, facilement accessibles et appropriées à la prise de **décision**.

La finalité d'un système décisionnel est le **pilotage d'entreprise**.

Les systèmes décisionnels sont dédiés au **management de l'entreprise**.

9

### 3-Définition

La business intelligence est un système permettant aux dirigeants d'analyser et d'interpréter, à l'aide d'outils simples, les données complexes de l'entreprise et de son environnement économique.

10

### 4-Informatique opérationnelle et décisionnelle

- Traitement en ligne des données (**OLTP**): On line Transactional Processing
- Analyse en ligne de ces données (**OLAP**): On Line Analysis Processing

11

### 4-Informatique opérationnelle et décisionnelle

Systèmes d'information opérationnels :

- S'appuient sur des SGBD traditionnels pour gérer des BD « opérationnelles » ou de « production »
- Supportent en général une ou plusieurs grandes fonctions de l'entreprise (production,, ressources humaines, finance ...)
- Permettent des processus de traitement en ligne des données (Des données précises et à jour)

12

#### 4-Informatique opérationnelle et décisionnelle

Nouvelles attentes des SI :

- Considérer des quantités de données historisées de plus en plus importantes organisées selon différentes dimensions (temps, espace géographique, gammes de produit, ...)
- Passer du traitement en ligne des données (OLTP) à l'analyse en ligne de ces données (OLAP) selon différentes dimensions pour construire des indicateurs indispensables au pilotage de l'entreprise)

13

#### 4-Informatique opérationnelle et décisionnelle

Décisionnel	Opérationnel
Gros volumes de données à gérer. (Multidimensionnelle, Historisées, recalculées)	Petits volumes de données à gérer. (Relationnelle, Actualisées, mises à jour)
Nombre d'utilisateur restreint. (Analyste/décideur)	Utilisé par toute l'entreprise. (Agent opérationnel)
Processus ouverts pour permettre la génération de connaissance. (Requête complexe)	Processus fermés, transactionnels (Transaction simple)
Données en lecture seule.	Données en lecture – Écriture.
Rapidité moyenne comparée aux systèmes opérationnels.	Réponses très rapides.
Niveau de granularité très grand (Résumées, agrégées)	Niveau de granularité fin.

14

#### 5-Les phases d'un projet décisionnel

- La phase de collecte
- La phase d'intégration
- La phase de distribution
- La phase d'Exploitation

15

#### 5-Les phases d'un projet décisionnel

La phase de collecte : ETL (Extraction, Transformation, Loading)

- Un ETL extrait les données de sources hétérogènes issues des bases de production, les transforme (Rendre cohérentes les données des différentes sources) et les réinjecte dans une nouvelle base (le datawarehouse).
- C'est un système par lequel vont passer toutes les données des systèmes opérationnels avant d'arriver dans la forme souhaitée dans l'entrepôt (datawarehouse).

16

## 5-Les phases d'un projet décisionnel

- Les ETL sont basés sur un ensemble de connecteurs permettant la gestion de différentes sources de données
- Plusieurs logiciels sont disponibles sur le marché. Ils permettent d'effectuer de l'ETL sous forme de programmes "graphiques". Ils sont intégrés dans des suites de BI.

17

## 5-Les phases d'un projet décisionnel

Exemples ETL:

- Anatella2
- DataStudio (Data)
- Feature Manipulation Engine (FME)
- Hurence avec un ETL natif Hadoop
- IBM InfoSphere DataStage
- Informatica PowerCenter
- MapReport
- Microsoft SQL Server Integration Services (SSIS)
- Talend
- Etc.

18

## 5-Les phases d'un projet décisionnel

La phase de stockage consiste à :

- concentrer les données collectées dans un espace unifié : l'entrepôt de données. Il permet aux applications décisionnelles d'avoir une source d'information commune, homogène, normalisée et fiable.
- Centraliser les données structurées et traitées afin qu'elles soient disponibles pour un usage décisionnel.

19

## 5-Les phases d'un projet décisionnel

Les données, au préalable nettoyées et consolidées, seront stockées dans une base spécialisée : le data warehouse ou le datamart.

- Le data warehouse ou le datamart sont alimentés par l'outil d'ETL (Extract Transform load).
- Le data warehouse : récolte, stocke et gère efficacement des gros volumes de données pour la prise de décision.

20

## 5-Les phases d'un projet décisionnel

La phase de distribution/diffusion:

- Le portail informationnel décisionnel de l'entreprise (EIP : Enterprise Information Portal) ouvre un accès sur l'entrepôt de données.
- Le portail permet de faciliter l'échange et le partage d'information entre les utilisateurs.
- La fonction de diffusion met les données à la disposition des utilisateurs.
- Les données doivent être distribuées et considérées comme un flux plutôt que comme de simples données.

21

## 5-Les phases d'un projet décisionnel

La phase Exploiter:

- L'analyse des données via l'interrogation de l'entrepôt de données grâce à des outils OLAP afin d'effectuer des analyses multidimensionnelles.
- Il est également possible dans cette phase d'exploitation, d'utiliser des méthodes de data mining afin de rechercher des corrélations ainsi que de produire des rapports d'analyse grâce à des outils de reporting.

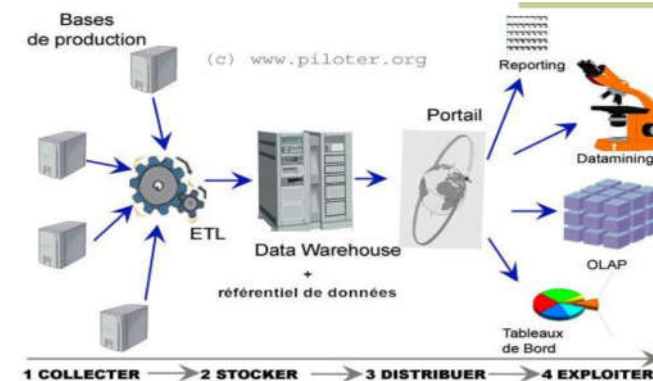
22

## 5-Les phases d'un projet décisionnel

- Analyser les données : la technologie la plus efficace pour effectuer des analyses à partir d'entrepôts de données est OLAP « OnLine Analytical Processing ».
- Rechercher des corrélations peu visibles avec le Data mining .
- Piloter la performance, aide à la décision des décideurs en situation avec les tableaux de bord présentant les indicateurs clés de l'activité.
- Produire des rapports d'analyse avec le Reporting

23

## 5-Les phases d'un projet décisionnel



24

# Chapitre 2



## Généralités sur les DW

25

## Plan

- 1 Introduction
- 2 Définitions
- 3 Exemple

26

### 1-Introduction

- Il est difficile d' avoir une vision globale homogène et cohérente des informations manipulées par l'ensemble des départements.
- L'information est inaccessible directement: il existe plusieurs sources utilisant des supports différents (papier, base de données, fichiers Excel).
- Les données de gestion peuvent avoir des significations différentes , exemples: la marge, le CA, le bénéfice, etc.

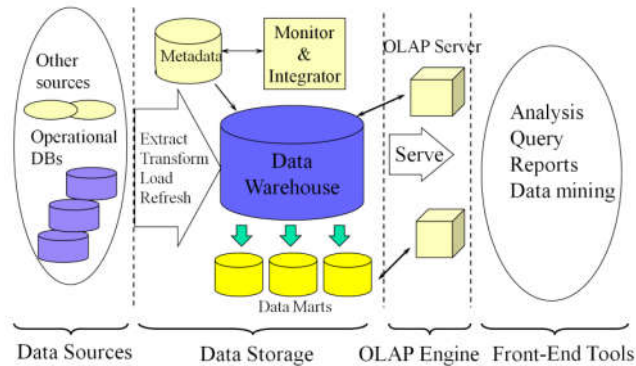
27

### 1-Introduction

- Toutes les données qu'elles proviennent des systèmes de production (ou systèmes transactionnels) de l'entreprise ou qu'elles soient achetées, vont devoir être organisées, coordonnées (ou regroupées), intégrées et stockées, pour donner à l'utilisateur une vue intégrée et orientée métier.

28

## 1-Introduction



29

## 2-Définition

Data Warehouse (DW) fournit un stockage permanent des données des sources hétérogènes pour une exploitation future (pour des fins d'analyse et de prise de décisions).

- il agrège de nombreuses données de l'entreprise (intégration)
- il mémorise les données dans le temps (historisation)
- il les organise pour faciliter les requêtes de prise de décision (optimisation).

**Synonymes :** entrepôt de données, base de données décisionnelle

30

## 2-Définition

W. H. Inmon (1996):

« Le data Warehouse est une collection de données **orientées sujet, intégrées, non volatiles et historisées**, organisées pour le support d'un processus d'aide à la décision »

31

## 2-Définition

### Orientation sujet :

- Le DW est organisé autour des sujets qui ont un intérêt majeur pour l'entreprise. Ainsi, les données sont structurées par thème, par opposition à celles des SI transactionnels, généralement organisés par processus fonctionnel.
- Le client, le produit sont des exemples de sujets d'intérêt.

32

## 2-Définition

### Données intégrées :

- L'intégration des données concerne celles qui sont internes à l'entreprise, mais aussi celles qui sont externes.
- Les données doivent être mises en forme et unifiées afin d'avoir un état cohérent pour l'ensemble du DW.

33

## 2-Définition

### Données historisées :

- Dans un DW, le système en préserve l'historique, c-à-d l'ensemble des valeurs que la donnée a pris au cours du temps.
- Une dimension temps doit donc être associée à la donnée afin de tracer son évolution.

34

## 2-Définition

### Données non volatiles :

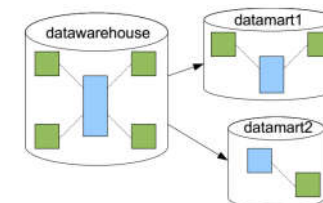
- La non volatilité est une conséquence de l'historisation.
- Une même requête effectuée à quelques mois d'intervalle, en précisant la date de prise de l'information recherchée donnera toujours le même résultat.

35

## 2-Définition

Magasin de données (datamart)

- Un sous-ensemble de l'ED contenant des informations se rapportant à un secteur d'activité (ou un métier) particulier de l'entreprise (finance, marketing, RH, etc.) .
- Un datamart est alimenté par un ED



36

## 2-Définition

OLAP

- Alimenté par un DW ou un DM
- Données organisées selon plusieurs axes d'analyse et selon différents niveaux de détail.
- Traitements permettant interactivement de changer de points de vue, de de niveau de détail, et effectuer les opérations OLTP classiques.
- Objectifs: génération des rapports (reporting) et analyse des données

37

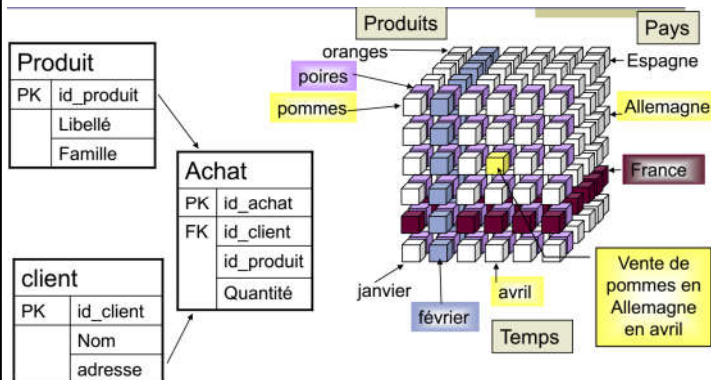
## 3-Exemple

Utilisation d'un entrepôt par une entreprise de distribution:

- Les données de vente sont enregistrées dans les différents magasins (OLTP).
- Chaque nuit, les données des différents magasins sont transférées dans un entrepôt de données au siège de la firme.
- Les données de l'entrepôt sont utilisée pour mettre au point des stratégies commerciales, des campagnes de promotion . . .

38

## 3-Exemple



39

# Chapitre 3



## Modélisation Multidimensionnelle

40

## Plan

- 1 Introduction
- 2 Le modèle en étoile
- 3 Le modèle en flocon
- 4 Le modèle en constellation

41

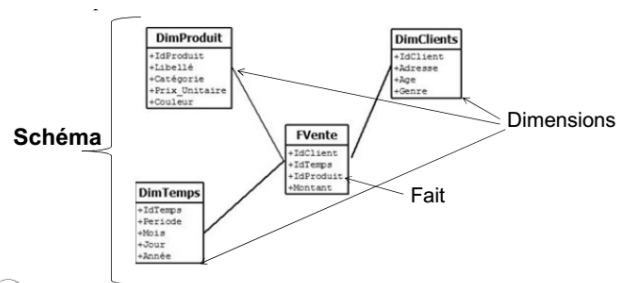
## 1-Introduction

- L'objectif principal d'un système décisionnel étant l'analyse de la performance, on mesure cette performance à travers des indicateurs qu'on a retenus.
- Les indicateurs vont donc être la base de la modélisation dimensionnelle et être regroupés dans une table dite **table de faits**.
- Ces faits (ou indicateurs ou métriques) peuvent alors être analysés suivant plusieurs dimensions.

42

## 1-Introduction

- Un schéma: Modélise des relations entre un (des) Fait(s) et des dimensions.



43

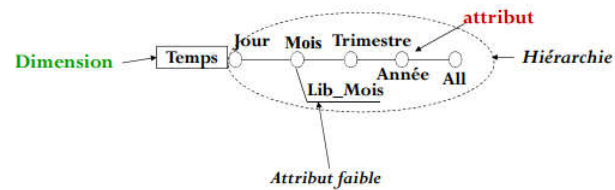
## 1-Introduction

- Les Dimensions:
  - Les axes de l'analyse selon lesquels sont visualisées les mesures d'activité d'un sujet d'analyse. Par exemple le temps (jour, mois, période, ...), la segmentation clientèle (tranche d'âge, marché, ...), etc.
  - Ils sont caractérisés par des attributs les plus souvent textuels.
  - Les attributs (ou paramètres) d'une dimension peuvent être organisés en hiérarchies, de la granularité la plus fine à la plus générale.

44

## 1-Introduction

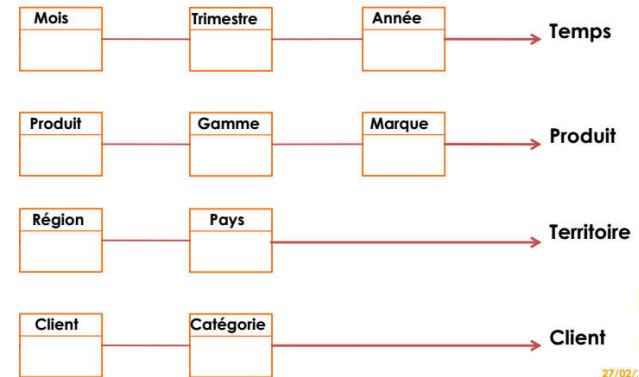
- Dimension : contient des membres organisés en hiérarchie
  - Chacun des membres appartient à un niveau hiérarchique particulier (ou niveau de granularité particulier)
  - Granularité d'une dimension : nombre de niveaux hiérarchiques



45

## 1-Introduction

Hiérarchies: Définition des niveaux de détail de l'analyse sur cette dimension



27/02/ 46

## 1-Introduction

Analyse multidimensionnelle: capacité à manipuler des données qui ont été agrégées selon différentes dimensions

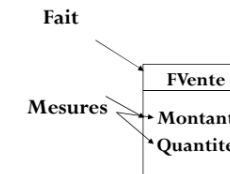
- ex. : analyse des ventes /catégorie de produit 1 dim.
  - + /année 2 dim.
  - + /département commercial 3 dim.
  - + / zone géographique 4 dim.



47

## 1-Introduction

- Un fait:
  - Sujet de l'analyse (sur quoi porte l'analyse)
  - Il comporte les mesures d'activité.
  - Une mesure est un indicateur d'analyse de type numérique et cumulable comme le prix, la quantités.



48

## 1-Introduction

- Un fait / mesure = critère d'évaluation du processus décisionnel. ex : chiffre d'affaires, quantité en stock
- Dimension = axe d'analyse associé à un indicateur.
- Un sujet d'intérêt. ex : temps, produit, localisation
- Hiérarchie = décomposition d'une dimension en une arborescence de niveaux. ex : temps décomposé en mois, trimestre, année, ...
- Attribut = caractéristique d'un niveau d'une hiérarchie. ex : prix d'un article
- Agrégat = résultat d'un indicateur par rapport à des niveaux. ex : chiffre d'affaires du mois.

49

## 1-Introduction

- Les schémas multidimensionnels:
  - o Le schéma en étoile (star schema)
  - o Le schéma en flocon (snowflake schema)
  - o Le schéma en constellation

50

## 2-Le modèle en étoile

- La table de faits est placée au centre du modèle, les tables de dimensions sont placées autour. Ce modèle présente visuellement une étoile.
- La table au centre de l'étoile est la table des faits. Les indicateurs sont groupés dans cette table, et ils partagent le même ensemble de dimensions.

51

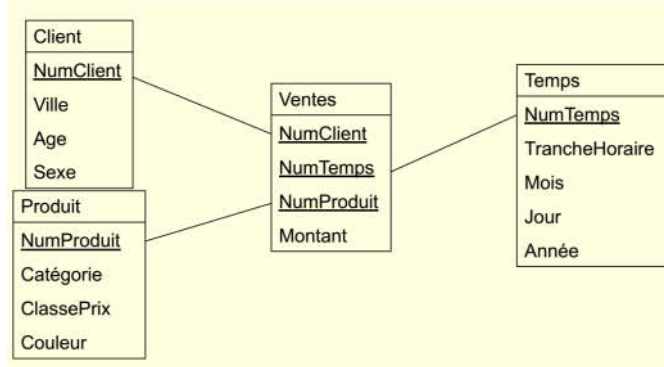
## 2-Le modèle en étoile

Exemple 1 :

- Un seul fait : le montant des ventes.
- Un enregistrement dans la table des faits «ventes » correspond à un total des ventes faites pour un client, dans une tranche horaire d'un jour précis, pour un produit choisi.

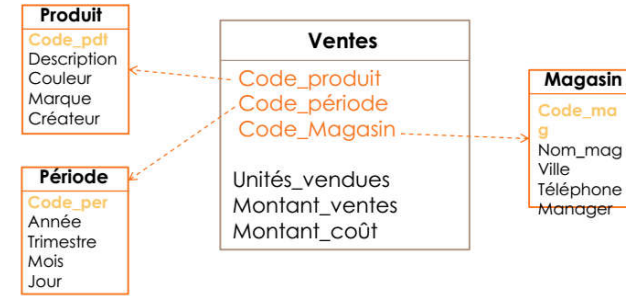
52

## 2-Le modèle en étoile



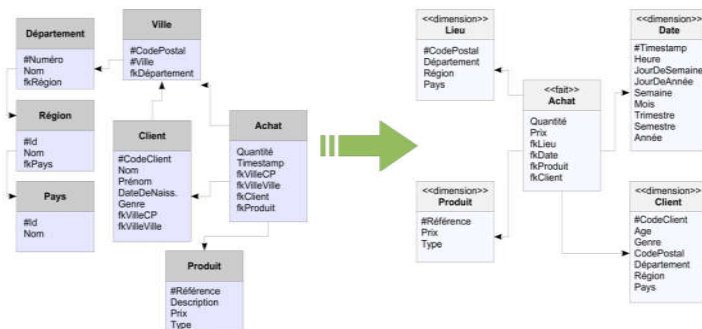
53

## 2-Le modèle en étoile



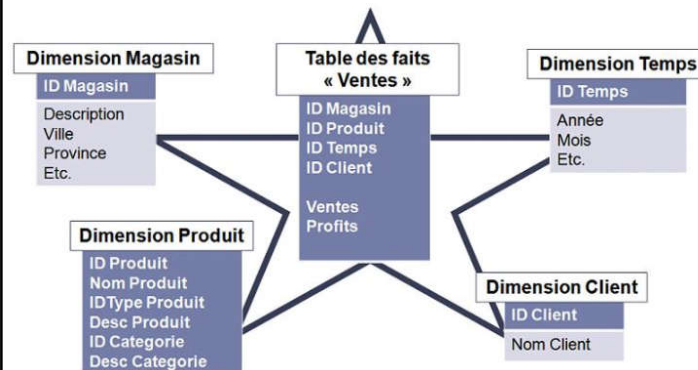
54

## 2-Le modèle en étoile



55

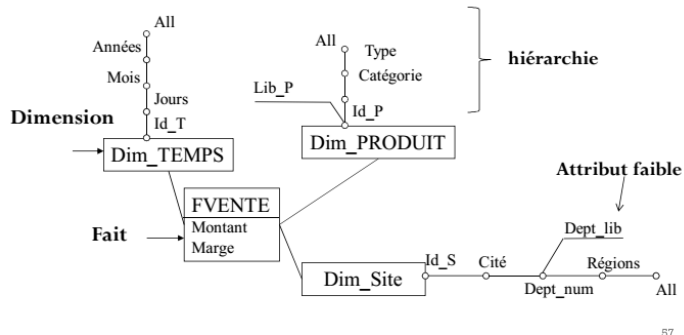
## 2-Le modèle en étoile



56

## 2-Le modèle en étoile

Exemple : « Modélisation des ventes en fonction du Temps, des produits et des sites de ventes »



57

## 2-Le modèle en étoile

### Exercice :

- Une entreprise commerciale de vente de matériels informatiques à plusieurs filiales internationales souhaite rassembler dans un Entrepôt de Données (ED) des informations sur les ventes du jour afin de dresser des tableaux de bord sur les chiffres d'affaires réalisés.

58

## 2-Le modèle en étoile

- L'entreprise dispose d'un système d'information complexe, constitué des éléments suivants:
- des applications et bases de données éparses et hétérogènes sur les produits qu'elle vend,
- des applications et bases de données variées sur les clients,
- des Bases de données sur les personnels de l'entreprise,
- des Bases de données sur les filiales de l'entreprise.

59

## 2-Le modèle en étoile

- Nous disposons des informations suivantes :
- Une vente correspond à un produit et un seul,
- Un produit est défini par: Refp, nom, description, type, famille.
- Une vente est effectuée par l'un des vendeurs du service de vente spécialisé dans le produit.
- Un client est défini par : RefC, nom, prénom, genre, adresse, catégorie.
- Un vendeur est défini par : RefV, nom, prénom, genre, adresse, filiale.

60

## 2-Le modèle en étoile

### Questions :

- Proposer des indicateurs d'analyse pour ce système de gestion
- Proposer des axes d'analyse pour ce système.
- Proposer un schéma en étoile dimensionnel de l'ED.

61

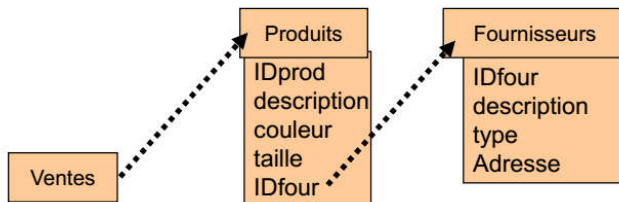
## 3-Le modèle en flocon

- Modèle en étoile + embranchement des dimensions.  
→ on éclate les tables de dimensions en sous-tables (selon la hiérarchie de cette dimension) selon le besoin.
- Avantages : permettre des analyses détaillées (drill down) sur la dimension hiérarchisée.
- Inconvénients : navigation difficile, nombreuses jointures.

62

## 3-Le modèle en flocon

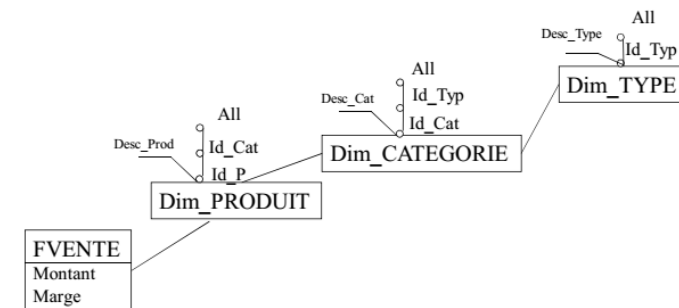
→ Raffinement du schéma étoile avec des tables normalisées par dimensions.



63

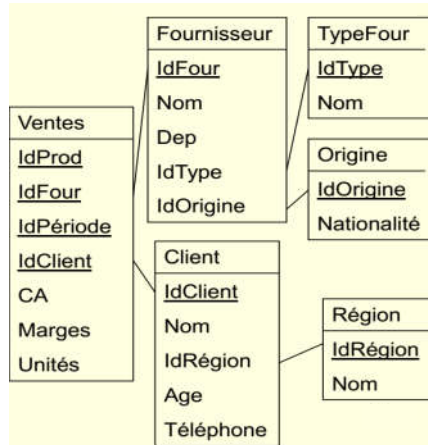
## 3-Le modèle en flocon

Exemple: Normalisation de la dimension Produits



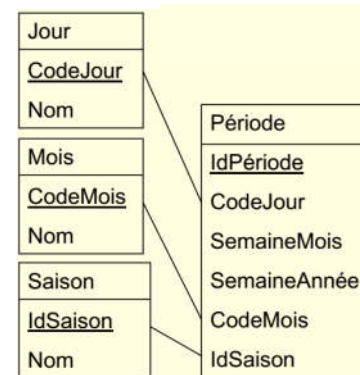
64

### 3-Le modèle en flocon



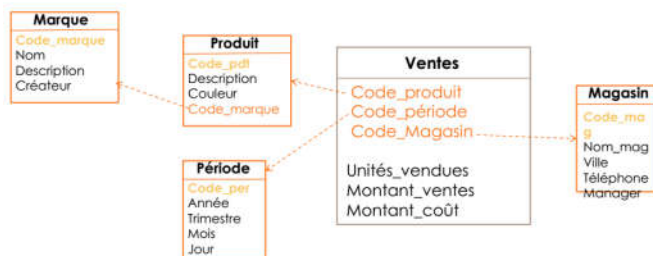
65

### 3-Le modèle en flocon



66

### 3-Le modèle en flocon



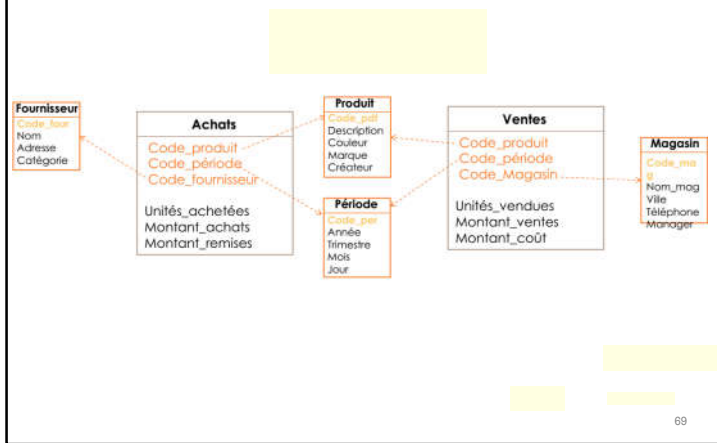
67

### 4-Le modèle en constellation

- Fusionner plusieurs modèles en étoile qui utilisent des dimensions communes
- Un modèle en constellation comprend donc :
  - o Plusieurs tables de faits
  - o Des tables de dimensions communes ou non à ces tables de faits.
  - o n sujet d'analyse (Faits)
  - o m axes d'analyse (Dimensions) pouvant être , partagés entre les différents faits

68

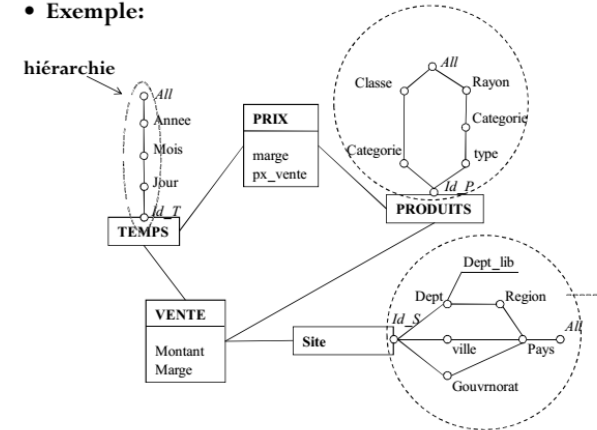
## 4-Le modèle en constellation



69

## 4-Le modèle en constellation

### Exemple:



70

# Chapitre 4



## Analyse Multidimensionnelle

71

## Plan

- 1 Introduction
- 2 Cube de donnée
- 3 L'Analyse OLAP

72

## 1-Introduction

- Le DW intègre des données à partir de sources multiples et hétérogènes afin de répondre aux requêtes du système d'aide à la décision.
- Ce type d'application est appelé On-Line Analytical Processing (OLAP)
- OLAP permet la transformation des données en informations stratégiques.

73

## 1-Introduction

- L'analyse multidimensionnelle est l'aptitude à analyser et exploiter des données qui ont été agrégées suivant plusieurs dimensions.
- Les données multidimensionnelles sont représentées par un cube (ou hypercube):
  - Les axes présentent les dimensions définies par l'utilisateur,
  - Les points dans l'espace (ou cellules) contiennent les mesures des faits calculées à partir des formules d'agrégation

74

## 2-Cube de donnée

- Un cube OLAP est une représentation abstraite d'informations multidimensionnelles exclusivement numérique utilisé par l'approche OLAP (acronyme de On-line Analytical Processing).
- Principe de base : ce sont les analyses des indicateurs qui intéressent l'utilisateur.

75

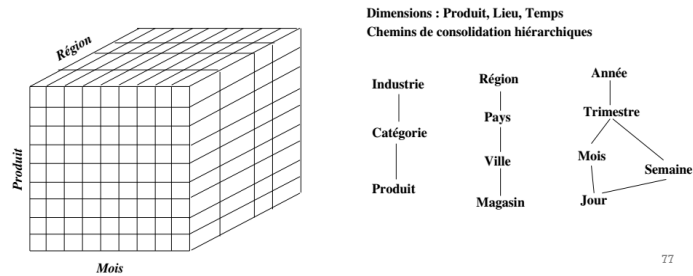
## 2-Cube de donnée

- Les cubes OLAP ont les caractéristiques suivantes :
  - 1- obtenir des informations déjà agrégées selon les besoins de l'utilisateur.
  - 2- simplicité et rapidité d'accès
  - 3- capacité à manipuler les données agrégées selon différentes dimensions
  - 4- un cube utilise les fonctions classiques d'agrégation : min, max, count, sum, avg, mais peut utiliser des fonctions d'agrégations spécifiques.

76

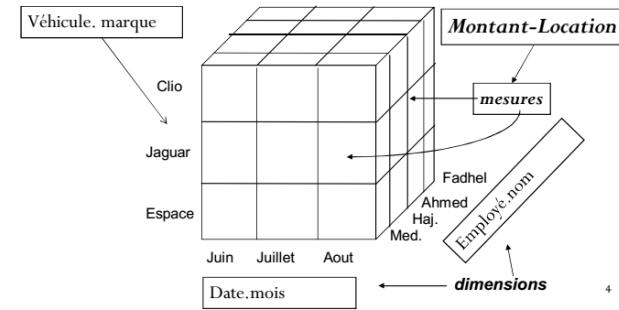
## 2-Cube de donnée

- Modèle multidimensionnel : les données sont vues comme des data cubes
- Exemple: Montant des ventes comme une fonction des paramètres produits, mois, région



## 2-Cube de donnée

- Une vue des données multidimensionnelle et logique à l'utilisateur.

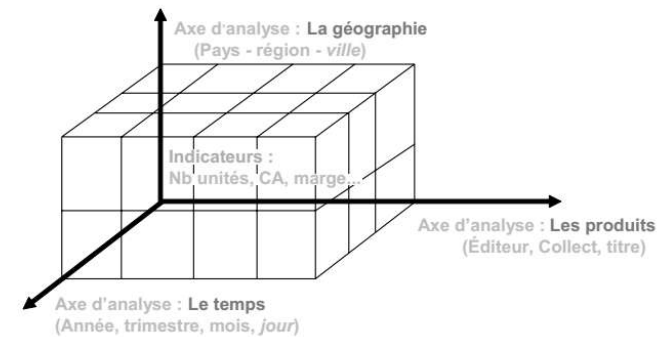


## 2-Cube de donnée

- Sujet analysé : un point dans un espace à plusieurs dimensions.
- Organisation des données pour mettre en évidence le sujet analysé et les différentes perspectives de l'analyse.
- data cube (par exemple, les ventes) : vision des données sur plusieurs dimensions.

79

## 2-Cube de donnée



## 2-Cube de donnée

- Exemple Cube : représentation des données dans le cas de deux dimensions

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

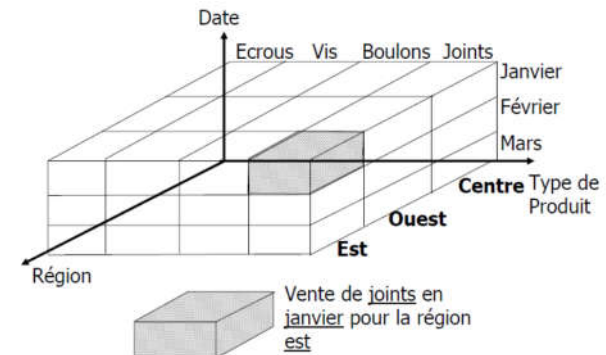


	Est	Ouest	Centre
écrous	50	60	110
vis	70	80	90
boulons	120	10	20
joints	50	40	70

81

## 2-Cube de donnée

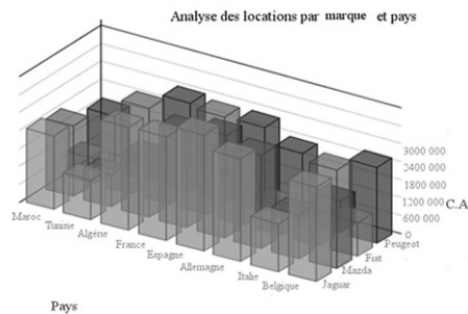
- Exemple Cube : représentation des données sous forme multidimensionnelle



82

## 3-L'Analyse OLAP

- Méthode d'analyse et d'exploitation des données (reflète l'activité d'une entreprise) fondée sur des critères combinés des dimensions.



83

## 3-L'Analyse OLAP

L'outil OLAP propose des méthodes de navigation dans les données:

- Drill-up/down
- Rotate
- Slicing
- Scoping
- Etc.

84

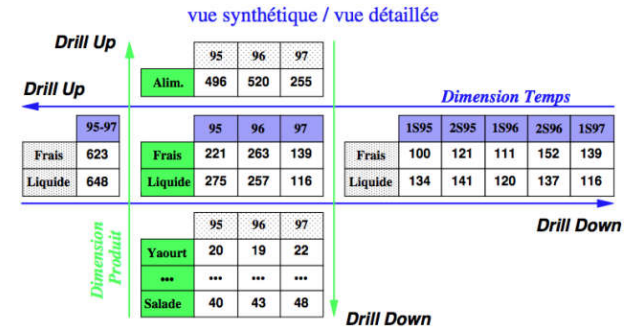
### 3-L'Analyse OLAP

- Drill Down: Forage vers le Bas: Ajout d'un (des) paramètre(s) d'une (des) dimension(s) afin de visualiser des données ayant un niveau de détail plus fin. c'est l'équivalent d'un zoom, opération inverse du drill-up.
- Drill Up (ou Roll Up): Forage vers le haut: visualiser les données ayant un niveau de détail moins fin. synthèse des informations en fonction d'une dimension (exemple de drill-up sur l'axe temps : passer de la présentation de l'information jour par jour sur une année, à une valeur synthétique pour l'année)

85

### 3-L'Analyse OLAP

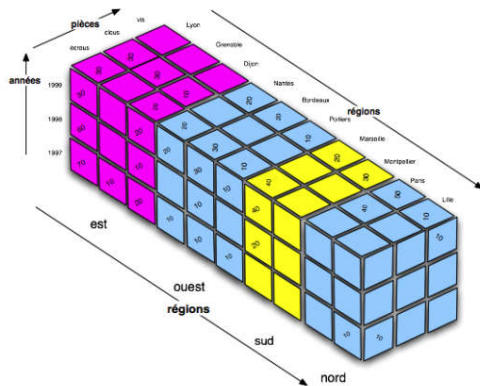
- Drill-up/down (vue synthétique / vue détaillée)



86

### 3-L'Analyse OLAP

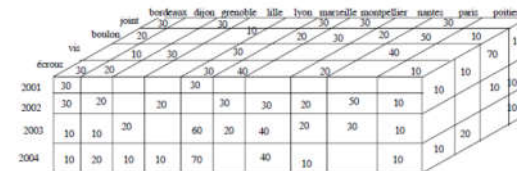
Drill-down du niveau des régions au niveau villes : Drill-down(regions) :



87

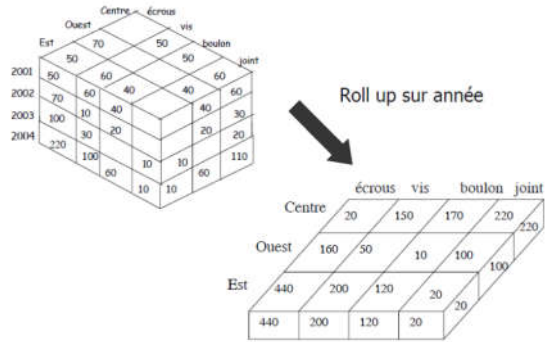
### 3-L'Analyse OLAP

Drill-down ~ opération inverse de Roll-up  
Drill-down du niveau des régions au niveau villes



88

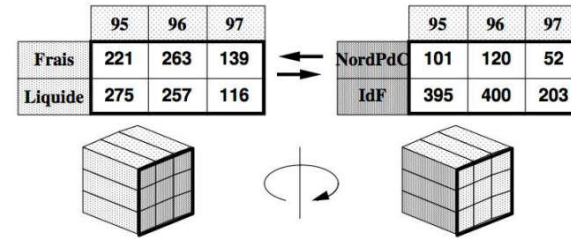
### 3-L'Analyse OLAP



89

### 3-L'Analyse OLAP

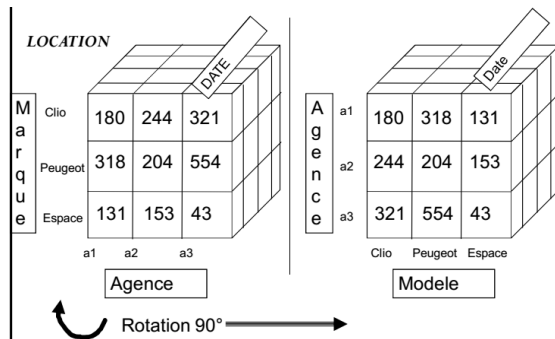
- Rotate



90

### 3-L'Analyse OLAP

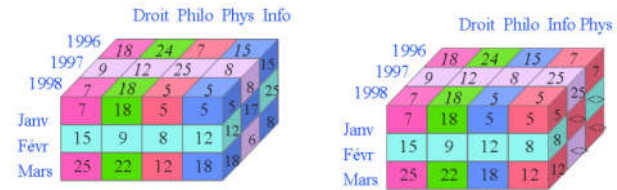
- Rotate: sélection du couple de dimensions à cibler



91

### 3-L'Analyse OLAP

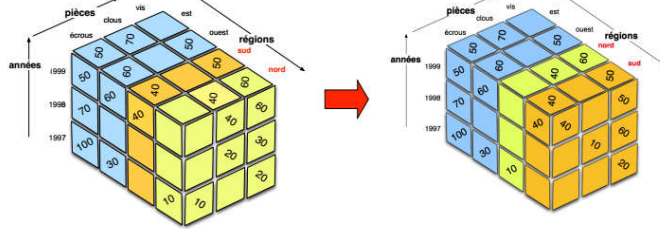
- Switch (Permutation): consiste à interchanger la position des membres d'une dimension.



92

### 3-L'Analyse OLAP

- Switch (Permutation)



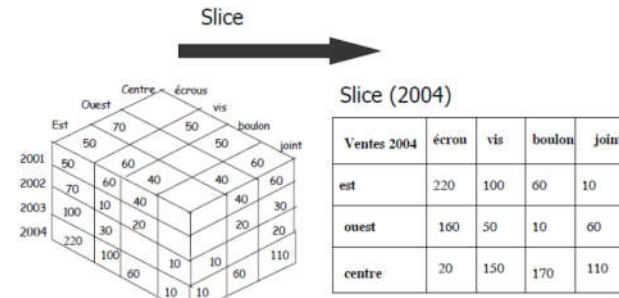
Ici sont interchangés les membres nord et sud de la dimension régions

nord	1999	1998	1997	sud	1999	1998	1997
vis	60	30	20	vis	50	60	60
clous	40	20		clous		10	
écrous			10	écrous	40	20	

93

### 3-L'Analyse OLAP

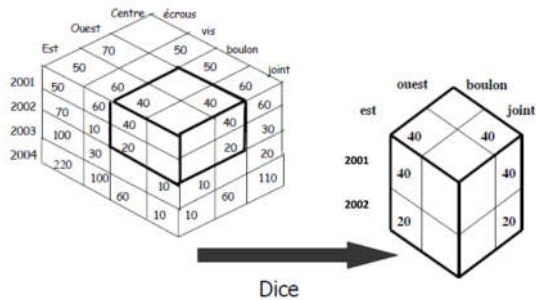
- Slicing (projection) : sélection de tranches du cube pour des valeurs prises par une dimension → exp: permet de retenir la partie du cube qui correspond à cette date.



94

### 3-L'Analyse OLAP

- Dicing : extraction d'un sous-cube

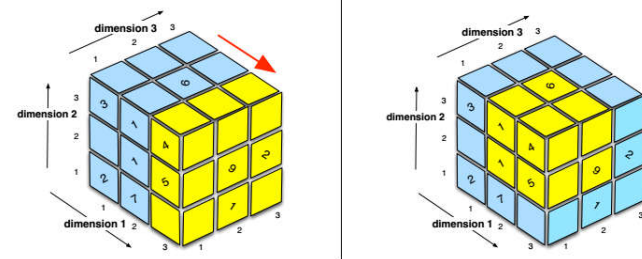


Dice

95

### 3-L'Analyse OLAP

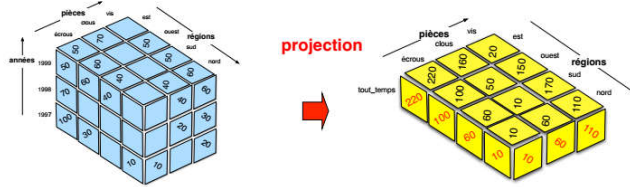
- Slicing/Dicing



96

### 3-L'Analyse OLAP

- Slicing



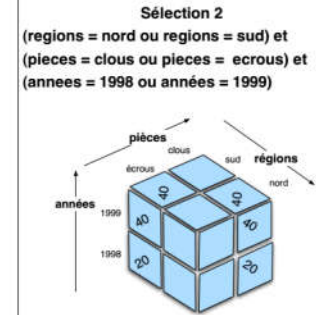
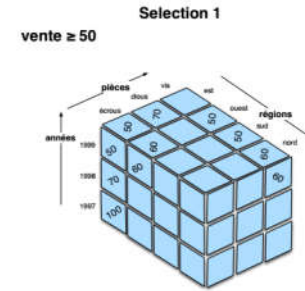
$\Pi$  pièce, région :

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

97

### 3-L'Analyse OLAP

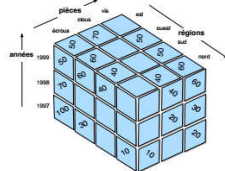
- Dicing



98

### 3-L'Analyse OLAP

- 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
vis	50	50	50
clous			10

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	

99

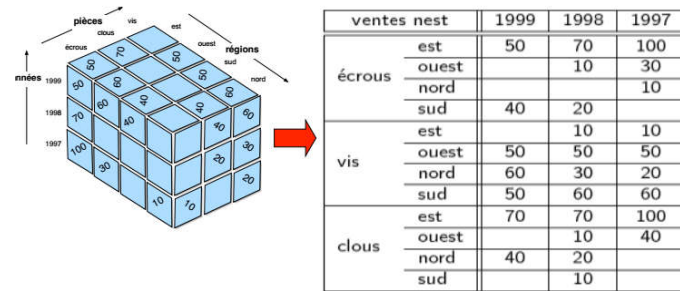
### 3-L'Analyse OLAP

- 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

100

### 3-L'Analyse OLAP

Nest (pièces, région) :



101

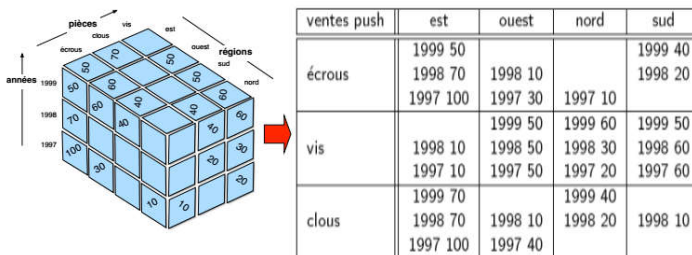
### 3-L'Analyse OLAP

- 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

102

### 3-L'Analyse OLAP

- puch(année) :



103

### 3-L'Analyse OLAP

- drill-across (jointure)

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

104

### 3-L'Analyse OLAP

Opérations OLAP ensemblistes :

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

Opérations de granularité :

- roll-up
- drill-down

Operations de restructuration

- Rotate ou Pivot
- Switch ou permutation
- Split ou division
- Nest ou l'emboîtement
- Push ou l'enfoncement

105

# Chapitre 5



## Restitution des Informations

106

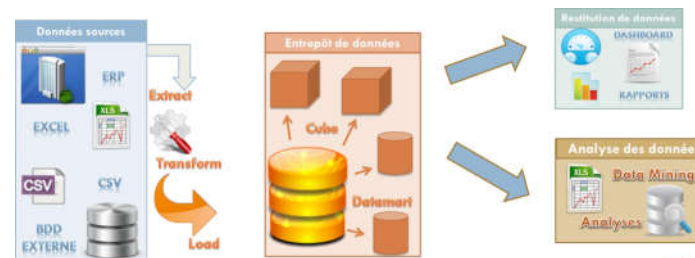


## Plan

- 1 Introduction
- 2 Exemple
- 3 Fouille de données
- 4 Arbres de décision
- 5 Règles d'association

107

### 1-Introduction



108

## 1-Introduction

- Les données restituées peuvent former des tableaux de bord, des rapports, des courbes d'évaluation via des outils d'analyse décisionnelle.
- Présente les informations (à valeur ajoutée) d'une manière lisible et conviviale dans le cadre de l'aide à la décision



109

## 2-Les applications de Reporting

Les applications de « Reporting » permettant de:

- sélectionner des données par période, production, secteur de clientèle, etc.,
- trier, regrouper ou répartir ces données selon des critères de choix,
- réaliser des calculs (totaux, moyennes, sommes, pourcentages, écarts, comparatif, ...),
- présenter les résultats de manière synthétique ou détaillée, généralement sous forme de graphiques.

110

## 2-Les applications de Reporting

Exemple:

- Pentaho Report Designer (PRD)
- JasperServer, JasperReport, iReport
- Business Intelligence And ReportingTools (BIRT)
- Business Intelligence (Oracle)
- Etc.

111

## 3-Fouille de données

Ensemble de techniques d'exploration de données permettant d'extraire d'une base de données des connaissances sous la forme de modèles de description afin de :

- décrire le comportement actuel des données et/ou
- prédire le comportement futur des données

112

### 3-Fouille de données

Deux types de techniques de Data Mining:

- les techniques descriptives: Elles s'intéressent à mettre en évidence des informations présentes mais cachées par le volume important de données.
- les techniques Prédictives : Elles visent à évoquer de nouvelles informations à partir des informations présentes.

113

### 3-Fouille de données

Exemple de méthodes descriptives (non supervisées):

- Classification automatiques ou le **clustering**: regroupe des objets en groupes, classes, familles, segments ou clusters afin que 2 objets d'un même groupe se ressemblent le plus possible
- Recherche **d'associations**: cette technique consiste à rechercher des règles d'associations entre les variables.

114

### 3-Fouille de données

Exemple de méthodes prédictives (supervisées):

- **Régression** logistique: produire un modèle permettant de prédire avec le plus de précision possible les valeurs prises par une variable partir d'une série de variables explicatives continues.
- **Arbres de décision**: méthode récursive (basée sur diviser pour mieux régner) pour créer des sous-groupes à partir des groupes de base (les algorithmes les plus utilisés: le CART et le C5).

115

### 4-Arbres de décision

- Objectif général :
  - A partir d'un ensemble de valeurs d'attributs (variables prédictives ou variable endogènes) il s'agit de prédire la valeur d'un autre attribut (variable cible ou variable exogène)
- une des méthodes supervisée (apprentissage) les plus connues de classification et de prédiction
- un arbre est équivalent à un ensemble de règles de décision

116

## 4-Arbres de décision

- un arbre est composé :
  - de noeuds = classes d'individus de plus en plus fine depuis la racine
  - d'arcs = prédicats de partitionnement de la classe source
- Un noeud qui n'a pas de parents est appelé « noeud racine » ou « racine ».
- Un noeud qui n'a pas de noeuds enfants est appelé « noeud feuille » ou « feuille ».
- arbres binaires : 2 branches partent de chaque noeuds
- arbres n-aire : n branches partent de chaque noeud
- algorithmes d'apprentissage d'arbre : ID3, CART, C4.5

117

## 5-Règles d'association

- **Qu'est ce qu'un motif fréquent ?**
  - Un motif (ensemble d'items, séquences, arbres, ...) qui interviennent fréquemment ensemble dans une base de données
- **Analyse des associations : Corrélacion ou analyse de causalité**
  - Panier de la ménagère, conception de catalogue, etc.
  - Quels produits sont souvent achetés ensemble ?
- **Règles de la forme R : X → Y (A%, B%)**
  - **ANTECEDENT** → **CONSEQUENT** [Support, Confiance](support et confiance sont des mesures d'intérêt définies par l'utilisateur)

118

## 5-Règles d'association

### ANALYSE DES TICKETS

espr

{ "crème" } → { "pain" }



ID	PRODUITS
1	pain, crème, eau
2	crème
3	pain, crème, vin
4	eau
5	crème, eau

Support = Prob. (crème et pain) :

$$\text{Sup} = \frac{\text{nom}(\text{tran. contenant crème et pain})}{\text{nom\_total}(\text{tran.})} = \frac{2}{5} = 0,4$$

Confiance = Prob(crème et pain / crème) :

$$\text{Conf} = \frac{\text{nom}(\text{tran. contenant crème et pain})}{\text{nom}(\text{tran. contenant crème})} = \frac{2}{4} = 0,5 = \frac{\text{sup}(\text{crème et pain})}{\text{sup}(\text{crème})}$$

119