

BI

Cours 3- OLAP

On-Line Analytical Processing

Sonia GUEHIS
Sonia.guehis@parisnanterre.fr

OLAP

- Définition (Caron, 1998) :
« Il s'agit d'une catégorie de logiciels axés sur l'exploration et l'analyse rapide des données selon une approche multidimensionnelle à plusieurs niveaux d'agrégation »
- Approche multidimensionnelle
 - Basée sur des thèmes d'analyse (dimensions)
 - Intuitive
- Plusieurs niveaux d'agrégation :
 - Différents niveaux de granularité

OLTP/OLAP

- OLTP : On-Line Transaction Processing :
 - Informatique opérationnelle
- OLAP : On-Line Analytical Processing :
 - Informatique décisionnelle et accès immédiat.

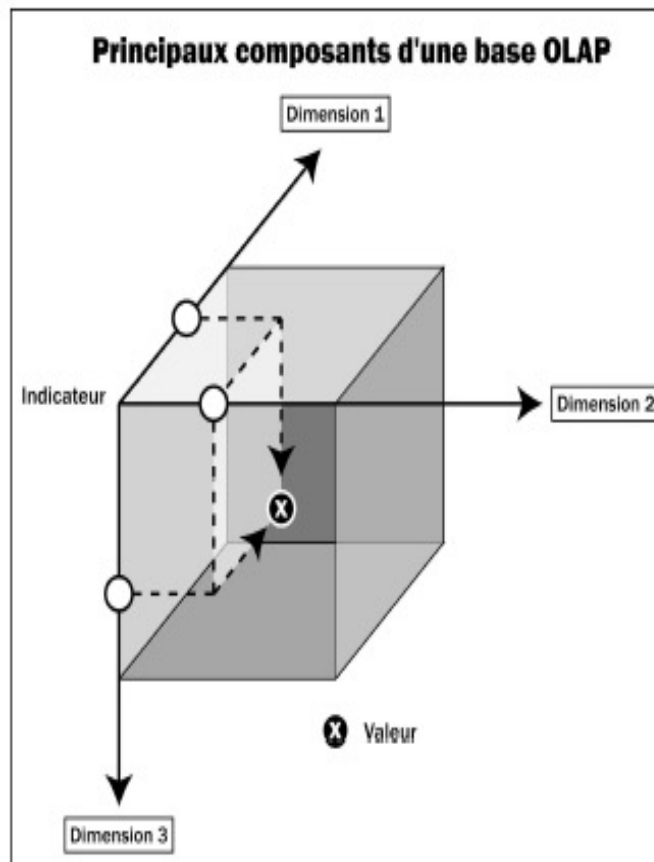
	OLTP	OLAP
Contenu	Données instantanées	Données historique
Utilisateurs	Production	Décision
Format	Accès rapide	Redondance, tables plates
Accès	Concurrentiel, RW	Lecture seule

Analyse Multidimensionnelle

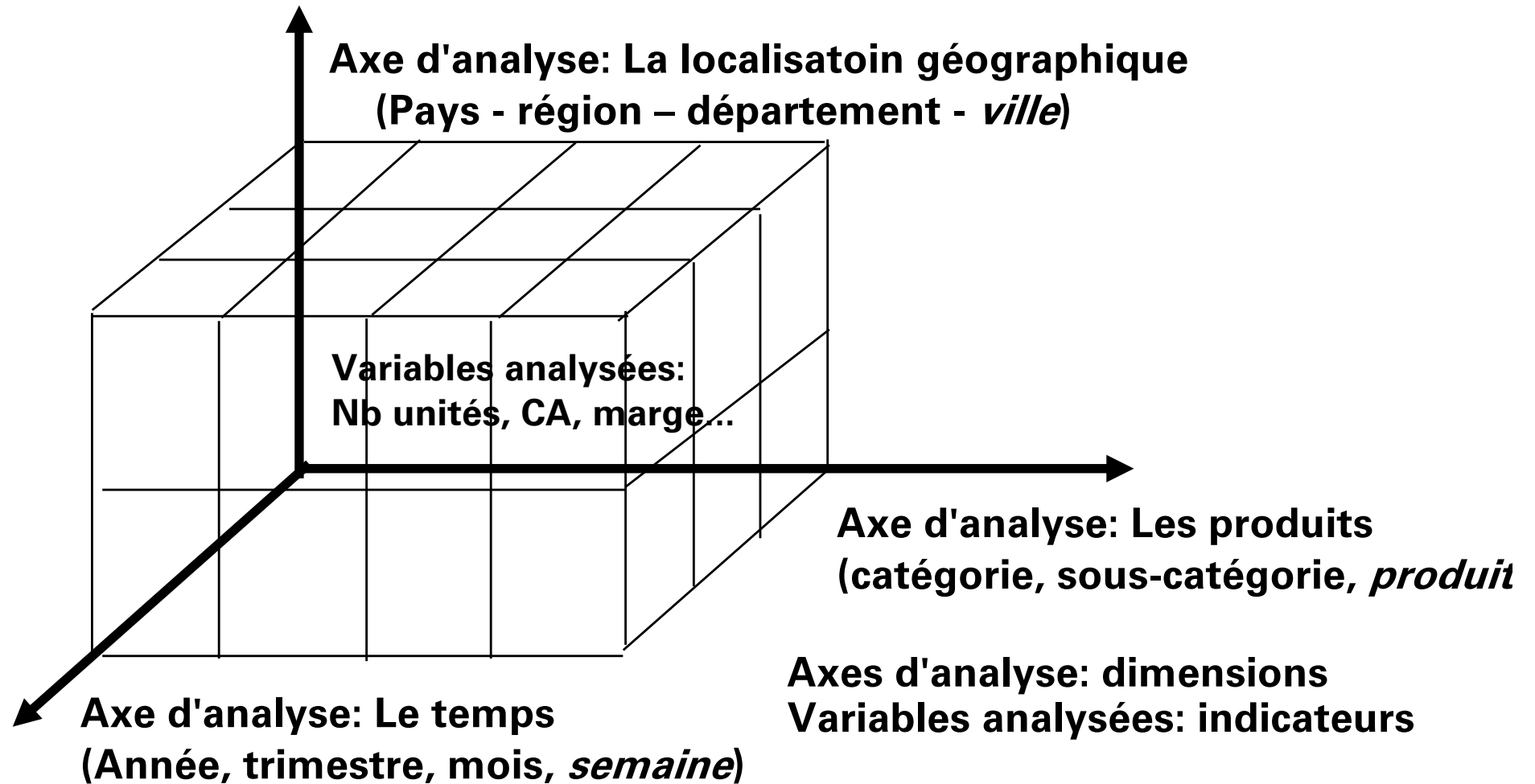
- Une collection de données orientées sujet, intégrées, non volatiles et historiées
- Organisées pour supporter un processus d'aide à la décision
- Comment : requêtes de type OLAP comportant
 - de nombreuses opérations de jointures et
 - d'agrégation sur des tables volumineuses
- Objectif : exploitation selon le besoin/simple/rapide d'accès

OLAP

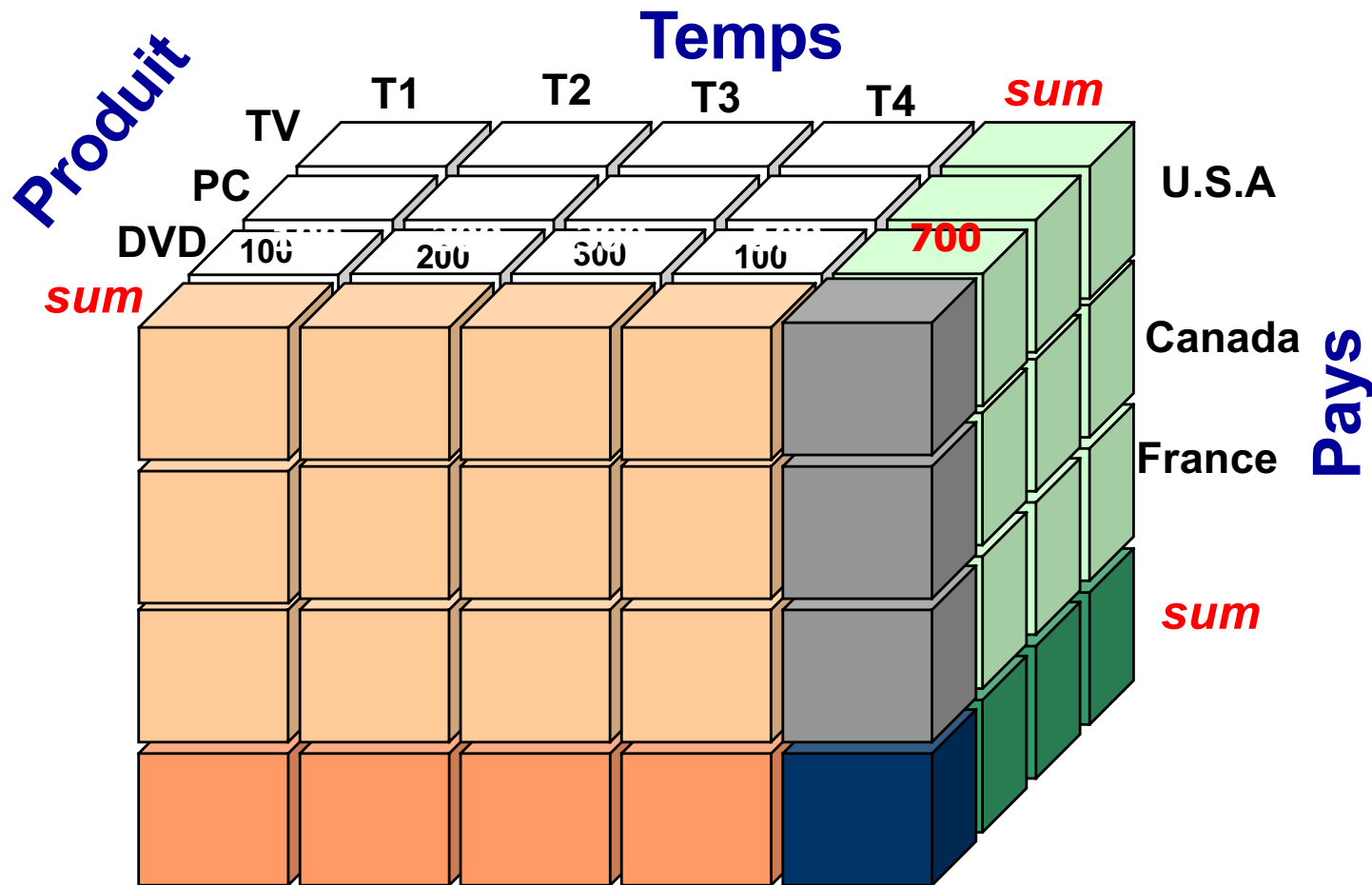
- Représentation de l'information dans un format multidimensionnel : cube à N dimensions (Hyper Cube)
- 4 composants de base :
 - L'indicateur : nombre d'unités vendues, CA, Coût, Marge...
 - Les dimensions : Temps, Géographie, Produits, Clients, Canaux de ventes....
 - Les hiérarchies
 - Les valeurs



Le data cube et les dimensions



Exemple de cube

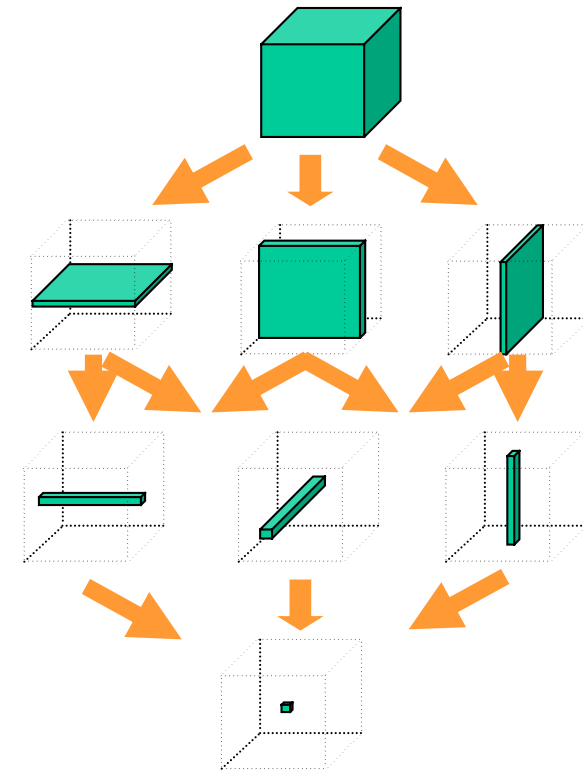
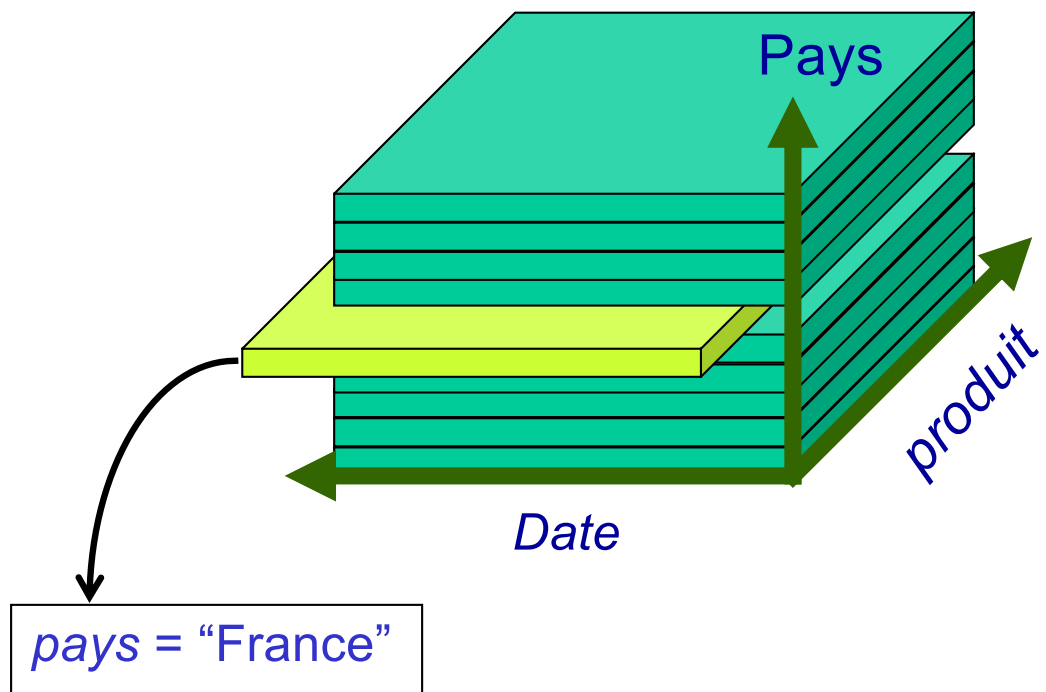


Opérations typiques de l'OLAP

« Navigation » dans les cubes multidimensionnels :

- ***Forer (drill-down, Roll-up) :***
 - descendre ou remonter dans la hiérarchie d'une dimension.
Ex. visualiser la vente par année, ensuite par mois et inversement.
- ***Pivoter (swap) :***
 - inter changer deux dimensions
- ***Projection et sélection (slice et dice)***
- ***Forer latéralement (drill-across) :***
 - Permet de passer d'un membre de dimension à un autre.
Ex. visualiser les données de Canada au lieu de celles de France

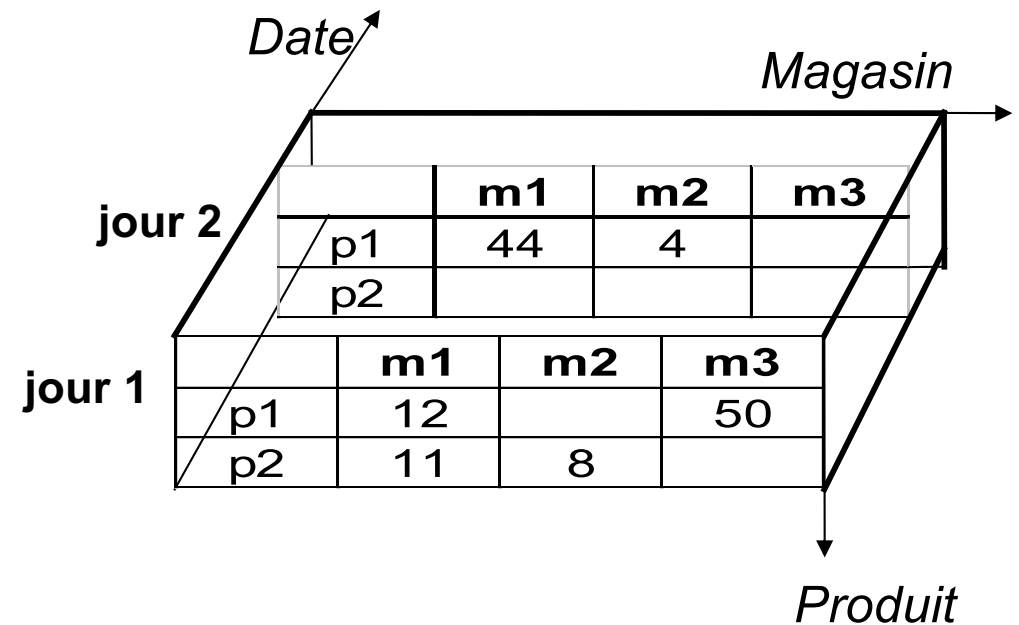
Opérations typiques de l'OLAP



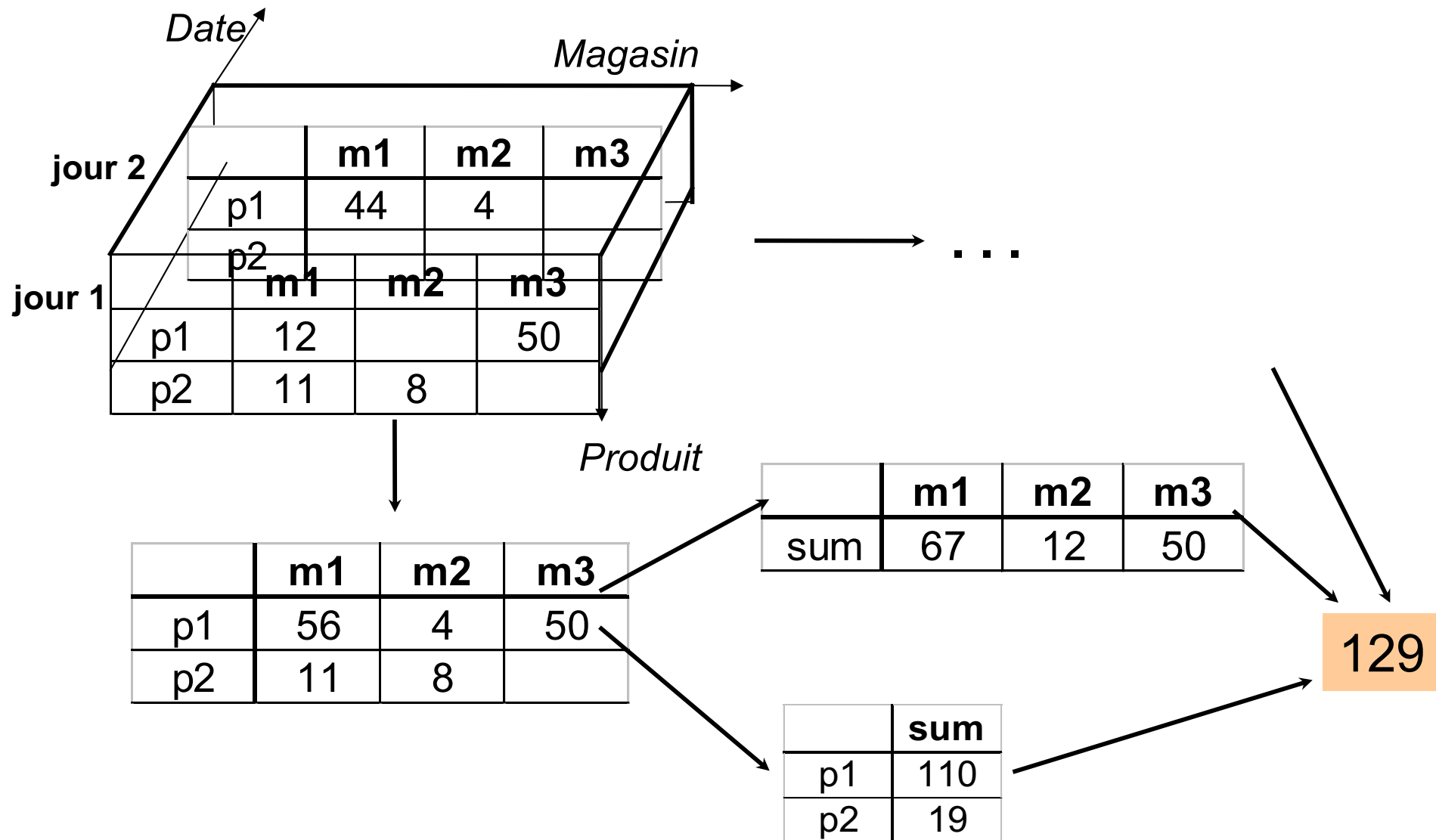
Exemple

Représentation multidimensionnelle: 3-D Data cube

Vente	idprod	idmag	date	CA
	p1	m1	1	12
	p2	m1	1	11
	p1	m3	1	50
	p2	m2	1	8
	p1	m1	2	44
	p1	m2	2	4



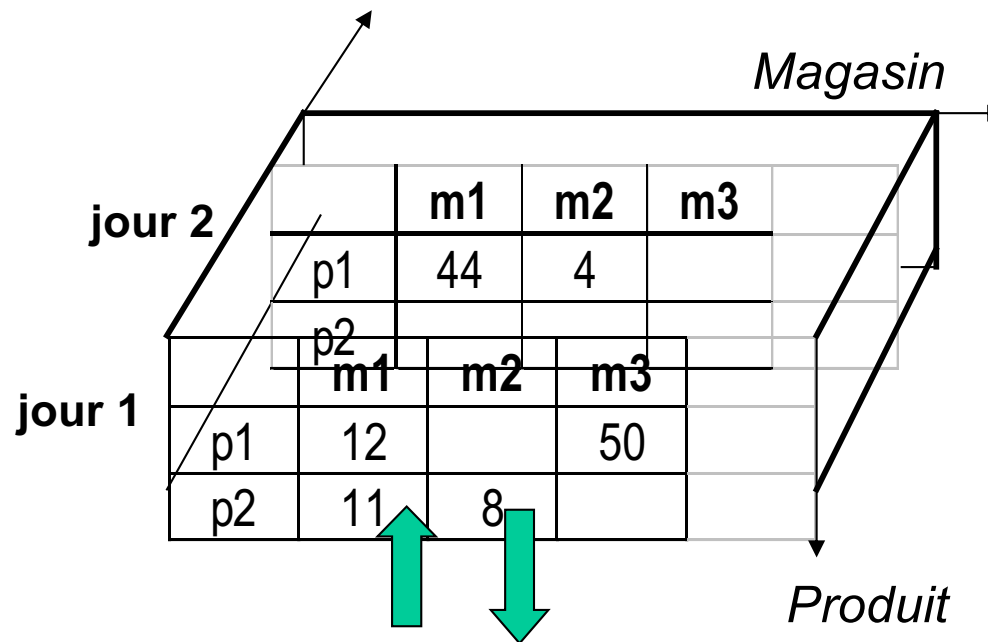
Exemple



Drill-down/ Roll-up

- **Roll-up: forage vers le haut**
 - Représente les données à un niveau de granularité supérieur selon la hiérarchie de la dimension désirée
 - Agréger selon une dimension:
 - Semaine -> mois
- **Drill-down: forage vers le bas**
 - L'inverse de l'opération Roll-up
 - Représente les données à un niveau de granularité inférieur
 - Détailler selon une dimension:
 - Mois -> semaine

Drill-down/ Roll-up



	region A	region B
p1	56	54
p2	11	8

	France		
p1	110		
p2	19		

magasin
|
region
|
pays

(Magasin m1 dans Region A;
Magasins m2, m3 dans Region B)

Opérations de Sélection /Projection

- **Slice**

- Sélection
- Tranche de cube obtenue par prédicats selon une dimension
- Ex: Année= 2020

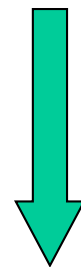
- **Dice**

- Projection selon un axe
- Sorte de cumuls de sélections
- Résultat: un sous-cube
- Ex: Projeter (Clients, Région)

Slicing

A 3D cube diagram representing a data cube. The vertical axis is labeled *Date*, the horizontal axis is labeled *Magasin*, and the depth axis is labeled *Produit*. The cube is divided into two slices along the *Date* axis: *jour 1* and *jour 2*. Each slice is a 2x4 grid of cells. The columns are labeled *m1*, *m2*, and *m3* (with an empty header cell). The rows are labeled *p1* and *p2*. The data values are as follows:

		<i>m1</i>	<i>m2</i>	<i>m3</i>
<i>jour 2</i>	<i>p1</i>	44	4	
	<i>p2</i>			
<i>jour 1</i>	<i>p1</i>	12		50
	<i>p2</i>	11	8	



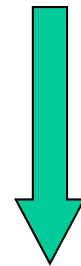
Date = jour 1

	m1	m2	m3
p1	12		50
p2	11	8	

Dicing

Diagram illustrating a 3D data cube structure for dicing. The axes are labeled: *Date* (vertical), *Magasin* (horizontal), and *Produit* (depth). The cube is divided into cells representing data points.

Date	Produit	m1	m2	m3
jour 2	p1	44	4	
	p2			
jour 1	p1	12		50
	p2	11	8	



Date = jour 1 ou jour 2,
magasin= m1

	Jour 1	Jour 2
P1	12	44
P2	11	

Matrice dense / creuse

- Matrice dense :
 - Toutes les cellules ont une valeur
- Matrice creuse :
 - Certaines valeurs absentes
 - Certaines dimensions n'ont pas de valeur pour chaque occurrence
 - Dimensions comportant des valeurs exclusives
 - Analyse des ventes par clients et par sexe

Avantages d'OLAP

- Facilité

- Ne nécessite pas de maîtriser les langages d'interrogation et des interfaces complexes
- Interrogation directement des données, en interagissant avec celles-ci

- Rapidité

- Exploite une dénormalisation maximale des données, sous la forme d'une pré-agrégation stockée
- L'utilisateur devient opérationnel très vite

→ L'utilisateur se concentre sur son analyse et non sur le processus (les moyens utilisés pour l'analyse)

Implémentation

- **Multidimensional OLAP (MOLAP)**
 - Basé sur un stockage par matrice en mémoire
 - Indexation rapide de données calculées
 - Opérations : manipulation des vecteurs
 - Accéder aux données via une API spécifique
- **Relationnel OLAP (ROLAP)**
 - Utilise un SGBD relationnel pour stocker les données ainsi qu'un middleware pour implémenter les opérations spécifiques de l'OLAP
 - Schéma en étoile ou flocon
 - Accéder aux données via SQL
 - Indexe bitmap, index de jointure
- **Hybride OLAP (HOLAP)**
 - Combinaison de ROLAP et MOLAP

Avantages / Inconvénients

MOLAP

– Avantages

- Toutes les agrégations sont calculées d'avance
- Rapidité d'exécution
- Accès direct aux données dans le cube

– Inconvénients

- Plus complexe à mettre en place
- Temps de chargement
- Volume important
- Analyses non-prevue

Avantages / Inconvénients

- ROLAP
 - Avantages
 - Définir des données multidimensionnelles avec un modèle simple
 - Permettre au DW d'évoluer avec relativement peu de maintenance
 - Peu coûteux
 - Stockage de gros volumes
 - Inconvénient
 - Temps de calcul
 - Moins performant dans les étapes de calcul

Quelques outils OLAP

- Oracle
 - OLAP API = Datacube
 - Express = Analyse
 - Report = Reporting
- Cognos
 - Impromptu = Reporting
 - Powerplay = Datacube
 - Query = Requêtage
- Business Object
 - BusinessQuery = Requêtage
 - BusinessObject = Requêtage + Analyse + Reporting
 - WebIntelligence = Datacube
- Hyperion
 - ESS Base = Base MOLAP
 - ESS Analysis = Analyse + Datacube

12 règles d'OLAP (Codd, 1993)

- **Multidimensionnalité** : Vue multidimensionnelle
- **Transparence** : L'emplacement physique du serveur OLAP est transparent
- **Accessibilité** : accès à toutes les données nécessaires à l'analyse
- **Stabilité** : Performance du système de Reporting est restée stable indépendamment du nombre de dimensions
- **Architecture** client/serveur
- **Dimensionnement** : générique
- **Gestion complète** : Gestion dynamique des matrices creuses
- **Support multi-utilisateurs**
- **Inter Dimensions** : Calculs à travers les dimensions
- **Intuitif** : Manipulation intuitive des données
- **Flexibilité** : Souplesse et facilité de constitution de rapports
- **Analyse sans limites** : Nombre illimité de niveaux d'agrégation et de dimensions