

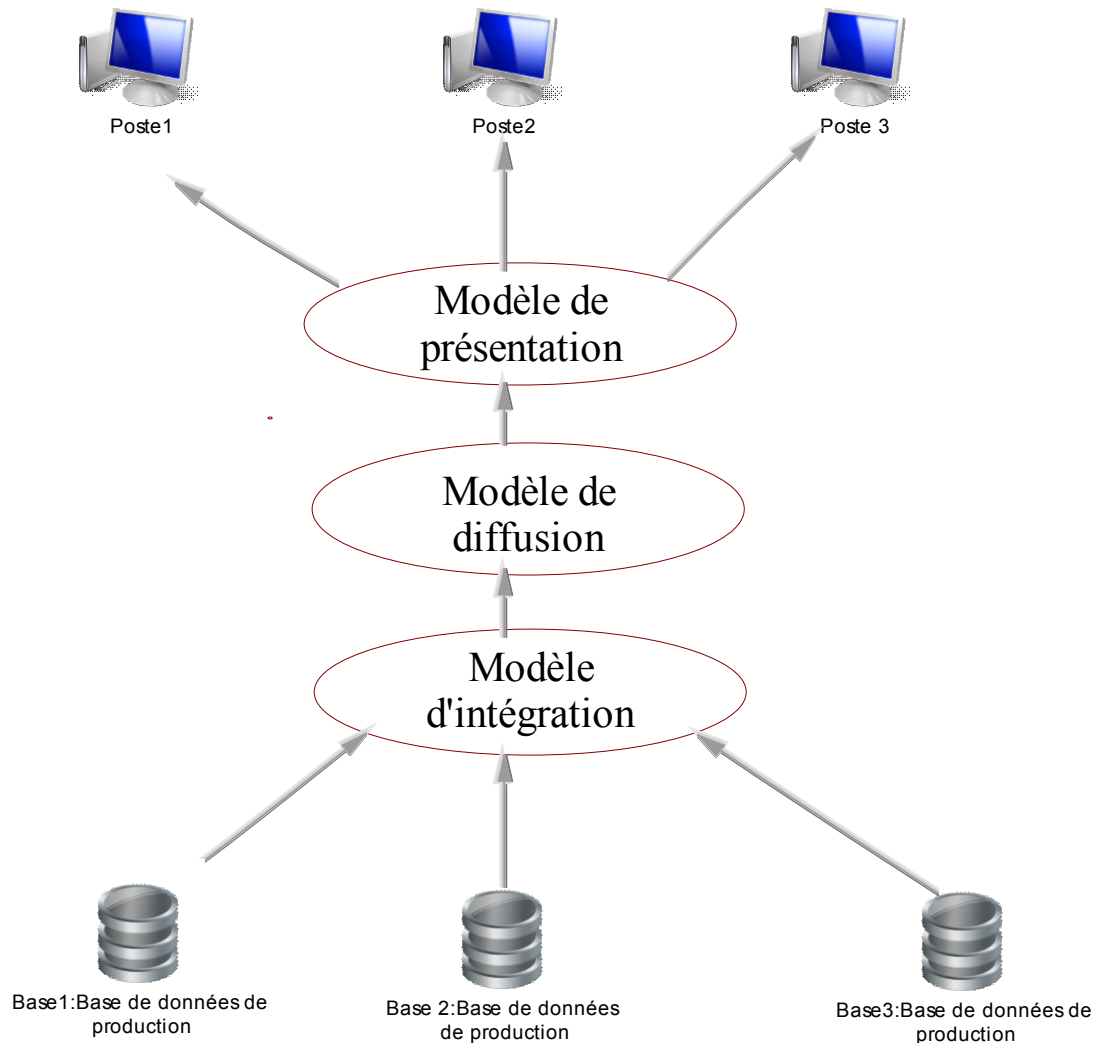
Bases de données
multidimensionnelles
OLAP

OnLine Analytical Processing

OLAP

- OLAP (On-Line Analytical Processing):
Ensemble des outils nécessaires pour la mise en place d'un Système d'information décisionnel (SID)

Modèles de données



- Le modèle d'intégration unifie les données
- Le modèle de diffusion modélise la structure de l'entrepôt de données (Serveur OLAP).
- Le modèle de présentation définit la manière dont les données seront présentées à l'utilisateur.

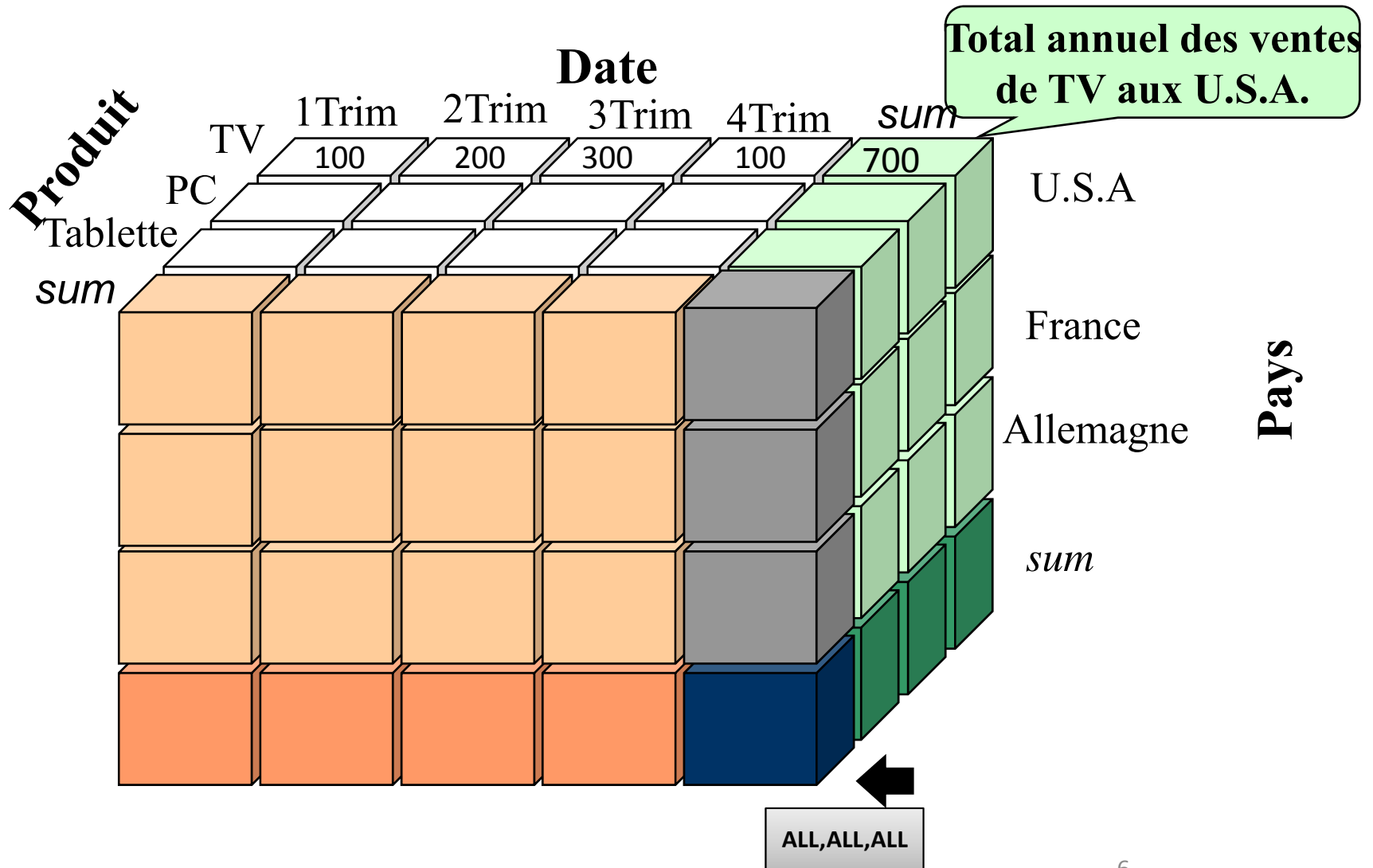
Les outils OLAP

- Les outils relationnels OLAP (ROLAP)
 - Les données sont stockées dans une base de données relationnelle, moteur OLAP permet de simuler le comportement d'un SGBD multidimensionnel
 - .
- Les outils MOLAP
 - Utilisent un système multidimensionnel « pur » qui gère les structures multidimensionnelles natives (les cubes)
 - Accès direct aux données dans le cube
- Les outils HOLAP (Hybrid OLAP)
 - tables de faits et tables de dimensions stockées dans SGBD relationnel (données de base)
 - données agrégées stockées dans des cubes

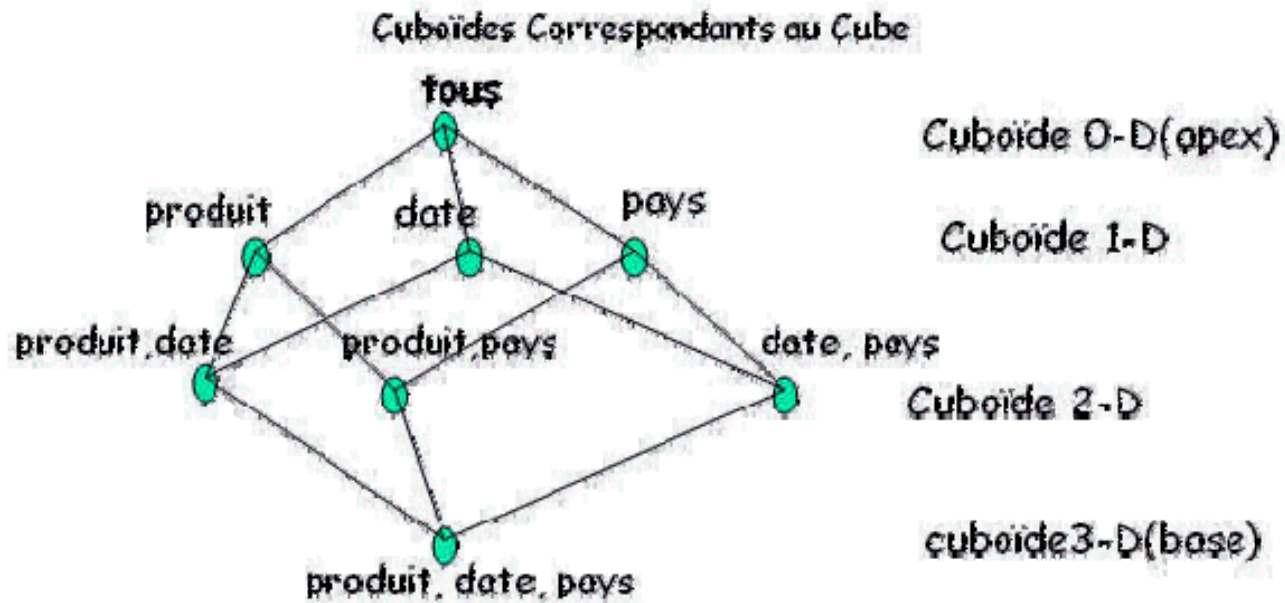
Le cube

- Modélisation multidimensionnelle des données facilitant l'analyse d'une quantité selon différentes dimensions:
 - Temps
 - Localisation géographique
 - Produits
 - Fournisseurs
 - Clients
 - ...
- Les calculs sont réalisés lors du chargement ou de la mise à jour du cube
- Un cube permet de visualiser les données selon plusieurs dimensions
- Un cuboïde est un cube de dimension n

Exemple de cube

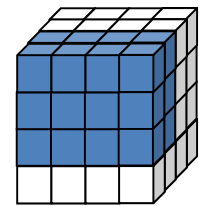
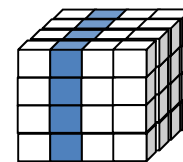


Treillis cuboïdes



Opérations typiques de l'OLAP

- Roll up : consolider (résumer) les données : Passer à un niveau supérieur dans la hiérarchie d'une dimension
- Drill down : l'inverse du Roll-up : descendre dans la hiérarchie d'une dimension
- Slice et Dice (tranche et extraction): Projection et sélection du modèle relationnel
- Pivot (rotate): Réoriente le cube pour visualisation



Modèles de stockage

- ROLAP: les données d'un groupe de mesures sont extraites dynamiquement à partir de la table des faits (le groupe de mesures est utilisé dans ce cas là uniquement pour définir les métadonnées).
- MOLAP: les données de la table des faits ainsi que les agrégats sont chargées dans une structure propre à SSAS.
- HOLAP: Base de données Multidimensionnelle hybride.

Couche métadonnées (sémantique)

- Le cube peut contenir des méta informations pour:
 - Pour décrire les tables, colonnes et relations dans une base de données.
 - Définir des hiérarchies
 - Supporter la localisation
 - Mettre en forme les données

Dimensions

- Hiérarchies
 - Hiérarchie d'attributs: une hiérarchie d'attribut possède deux niveaux: la racine qui contient tous les membres (toutes les valeurs) et les feuilles de la hiérarchie sont composées par les membres.
 - Hiérarchie naturelle: une hiérarchie naturelle est le résultat d'une relation entre attributs. chaque attribut fils doit avoir un seul attribut parent (Exemple: Catégorie → Sous Catégorie → Produit)
 - Hiérarchie utilisateur:
 - Peut être basée sur une hiérarchie naturelle ou non.
 - Facilite le parcours d'une dimension.
 - N'affecte pas la structure de stockage du cube.
 - Exemple d'une hiérarchie utilisateur non basée sur une hiérarchie naturelle: Couleur d'un article → Taille de l'article

MDX

MultiDimensional Expressions

Présentation

- Langage de requêtes sur les bases de données OLAP, il comprend des instructions de manipulation de données et des instructions de définition de données
- Développé par Microsoft en 1997.
- Adopté par les plus importants éditeurs de solutions BI
 - SAP (Business Objects)
 - Oracle
 - IBM Cognos
 - Mondrian (Pentaho)
 - ...

Exemples

- Measures est considérée comme une dimension spéciale qui regroupe l'ensemble des mesures (faits) du cube.
- Il faut préciser les membres à afficher sur les colonnes(ON COLUMNS, ON Axis(0) ou ON 0, et les membres à afficher sur les lignes(ON ROWS, ON AXIS(1) ou ON 1).
- Exemple 1: afficher toutes les ventes:

```
SELECT Measures.[sales amount] ON COLUMNS,  
Client.[State Province Name] ON ROWS  
FROM [Adventure Works DW]
```

	Sales Amount
All	29 358 677

Exemples

- Ventes par pays:

```
SELECT [Measures].[sales amount] ON COLUMNS,  
[Client].[State Province Name].MEMBERS ON ROWS  
FROM [Adventure Works DW]
```

- Pour obtenir les valeurs non nulles uniquement il faut précéder [Client].[State Province Name] par **NON EMPTY**.
- Pour ne pas afficher le membre ALL, utiliser CHILDREN à la place de MEMBERS.

	Sales Amount
All	29 358 677
Abingdon	(NULL)
Albany	(NULL)
Alexandria	(NULL)
Alhambra	(NULL)
Alpine	(NULL)
Wood Dale	(NULL)
Woodburn	90 053
Woodbury	(NULL)
Woodinville	(NULL)
Woodland Hills	170 278
Woolston	66 758
Woonsocket	(NULL)
Wrentham	(NULL)
Yakima	94 460
York	226 396
Zealand	(NULL)

Exemples

- Si vous sélectionnez plusieurs colonnes alors ils doivent être séparées par des virgules et regroupés par des accolades {} (on crée alors un ensemble.)

```
SELECT { Measures.[sales amount] , [Measures].[Order Quantity]} ON  
COLUMNS,  
NON EMPTY Client.[State Province Name].CHILDREN ON ROWS  
FROM [Adventure Works DW]
```

- Total des ventes et quantité commandée par ville et par couleur de produit: les membres de lignes doivent être entre () (on crée alors un tuple).
- Le même résultat peut être obtenu en utilisant le produit des deux

```
SELECT { Measures.[sales amount] , [Measures].[Order Quantity]} ON  
COLUMNS,  
NON EMPTY (Client.[State Province Name].CHILDREN ,  
[Produit].[Color].CHILDREN )ON ROWS  
FROM [Adventure Works DW]
```


Tri

Pour ordonner les lignes par ordre croissant (décroissant) des ventes par pays, il faut utiliser la fonction order (ASC ou DESC)

```
SELECT Measures.[sales amount] ON COLUMNS,  
NON EMPTY order(Client.[State Province Name].CHILDREN , Measures.[sales  
amount]  
,ASC) ON ROWS  
FROM [Adventure Works DW]
```

Tri

Ventes par pays et par couleur de produit (tri ascendant)

```
SELECT [Measures].[Sales Amount] ON COLUMNS,  
NON EMPTY Order(  
[Client].[State Province Name].CHILDREN * [Produit].[Color].CHILDREN,  
[Measures].[Sales Amount],ASC)  
ON ROWS  
FROM [Adventure Works DW]
```

La fonction order est par défaut hiérarchique, il réorganise d'abord les lignes en fonction de leur position dans la hiérarchie, puis elle ordonne chaque niveau, pour ignorer la hiérarchie utilisez BASC, ou BDESC

		Sales Amount
Alabama	NA	37
Alberta	White	9
Alberta	Multi	68
Alberta	NA	379
Alberta	Blue	2 384
Alberta	Yellow	4 881
Alberta	Black	5 146
Alberta	Red	9 600
Arizona	NA	33

Afficher le total des ventes Internet par classe de produit:
(H=Haute, M=Moyenne, L=Basse)

```
SELECT Measures.[sales amount] ON COLUMNS,  
{Produit.class.H,Produit.class.L, Produit.class.M} ON ROWS  
FROM [Adventure Works DW]
```

Pour obtenir le même résultat, comme dans une feuille de calcul en utilisant la première et la dernière valeur séparées par « : »:

Produit.class.H:Produit.class.M

Filtre

- La clause prend en argument un tuple qui définit la condition du filtre

```
SELECT Measures.[sales amount] ON COLUMNS,  
{Produit.class.H,Produit.Class.L, Produit.class.M} ON ROWS  
FROM [Adventure Works DW]  
WHERE [Client].[French Country Region Name].FRANCE
```

```
SELECT { [Measures].[Sales Amount], [Measures].[Order Quantity]} ON  
COLUMNS,  
Filter([Client].[French Country Region Name].CHILDREN,[Measures].[Sales  
Amount]>5000000 AND [Measures].[Order Quantity]> 250)ON ROWS  
FROM [Adventure Works DW]
```

Exercices

Ecrire les requêtes MDX suivantes:

Ex1: Nombre de ventes par Gamme et par Pays de client.

Ex2: Nombre de ventes par Canal et par Pays de client

Ex3: Montant des ventes par Gamme et par Pays de Commercial

Ex4: Afficher le montant (Prix) et le nombre de ventes par mois (de l'année 2004) et par type de produit.

	Janvier	Janvier	Février	Février	Mars	Mars	Avril	Avril	Mai	Mai	Juin	Juin	Juillet
	Prix	Nb Ventes	Prix	Nb Ventes	Prix	Nb Ventes	Prix	Nb Ventes	Prix	Nb Ventes	Prix	Nb Ventes	Prix
Habillement	21879	6	116430	17	163770	23	151068	23	76158	12	88740	10	130530
Maison	29268	4	83199	15	186189	20	148788	18	89289	9	74424	6	145424
Sport	8064	2	41664	6	50880	9	61872	10	19608	5	5376	3	41376

Exercices

Ecrire les requêtes MDX suivantes:

Ex1: Nombre de ventes par Gamme et par Pays de client.

Ex2: Nombre de ventes par Canal et par Pays de client

Ex3: Montant des ventes par Gamme et par Pays de Commercial

Ex4: Afficher le montant (Prix) et le nombre de ventes par mois (de l'année 2004) et par type de produit.

	Janvier	Janvier	Février	Février	Mars	Mars	Avril	Avril	Mai	Mai	Juin	Juin	Juillet
	Prix	Nb Ventes	Prix	Nb Ventes	Prix	Nb Ventes	Prix	Nb Ventes	Prix	Nb Ventes	Prix	Nb Ventes	Prix
Habillement	21879	6	116430	17	163770	23	151068	23	76158	12	88740	10	130530
Maison	29268	4	83199	15	186189	20	148788	18	89289	9	74424	6	145424
Sport	8064	2	41664	6	50880	9	61872	10	19608	5	5376	3	41376