

Le Langage MDX

Chapitre 4

Business Intelligence

Partie 1 : Présentation de MDX

Points traités :

Historique

Définitions

Définition

- Le MDX (**M**ulti-**D**imensional **eX**pressions) est un langage de requête pour les bases de données OLAP, analogue au langage SQL pour les bases de données relationnelles
 - Il est conçu pour rechercher et manipuler l'information dans les cubes OLAP
 - Il permet de naviguer dans les bases multidimensionnelles, et définir des requêtes sur tous leurs objets (dimensions, hiérarchies, niveaux, membres et cellules)
- C'est aussi un langage de calcul avec une syntaxe similaire à celle des tableurs
 - Une requête MDX retourne un rapport à plusieurs dimensions consistant en un ou plusieurs tableaux 2D

Définition

- Le langage MDX possède une syntaxe appropriée à l'interrogation et la manipulation des données multidimensionnelles dans un cube OLAP
 - Bien qu'il soit possible d'écrire certaines expressions dans le langage SQL traditionnel, cela nécessite une syntaxe SQL complexe
- MDX a été adopté par une large majorité de fournisseurs de la technologie OLAP et c'est devenu un standard
 - Microsoft SQL Server
 - SAP Netweaver BI
 - Hyperion Essbase (Oracle)
 - SAS OLAP Server

Historique

- Développé au sein de Microsoft
 - Présenté pour la première fois en 1997 comme une partie de la spécification OLE DB pour OLAP
- XML Analytics (XMLA) normalise MDX en tant que langage de requête
- Avec MS Analysis Services 2005, Microsoft a introduit des extensions au langage comme les sous-requêtes
 - Excel 2007 utilise déjà des extensions MDX

MDX vs SQL

- La syntaxe de MDX ressemble à celle de SQL par ses mots clé SELECT, FROM, WHERE, mais leurs sémantiques sont différentes :
 - SQL construit des vues relationnelles
 - MDX construits des vues multidimensionnelles des données
- Comparatif entre les termes multi-dimensionnels et relationnels :

Multidimensionnel (MDX)	Relationnel (SQL)
Cube	Table
Niveau (Level)	Colonne (chaîne de caractère ou valeur numérique)
Dimension	plusieurs colonnes liées ou une table de dimension
Mesure (Measure)	Colonne (discrète ou numérique)
Membre de dimension (Dimension member)	Valeur dans une colonne et une ligne particulière de la table

MDX vs SQL

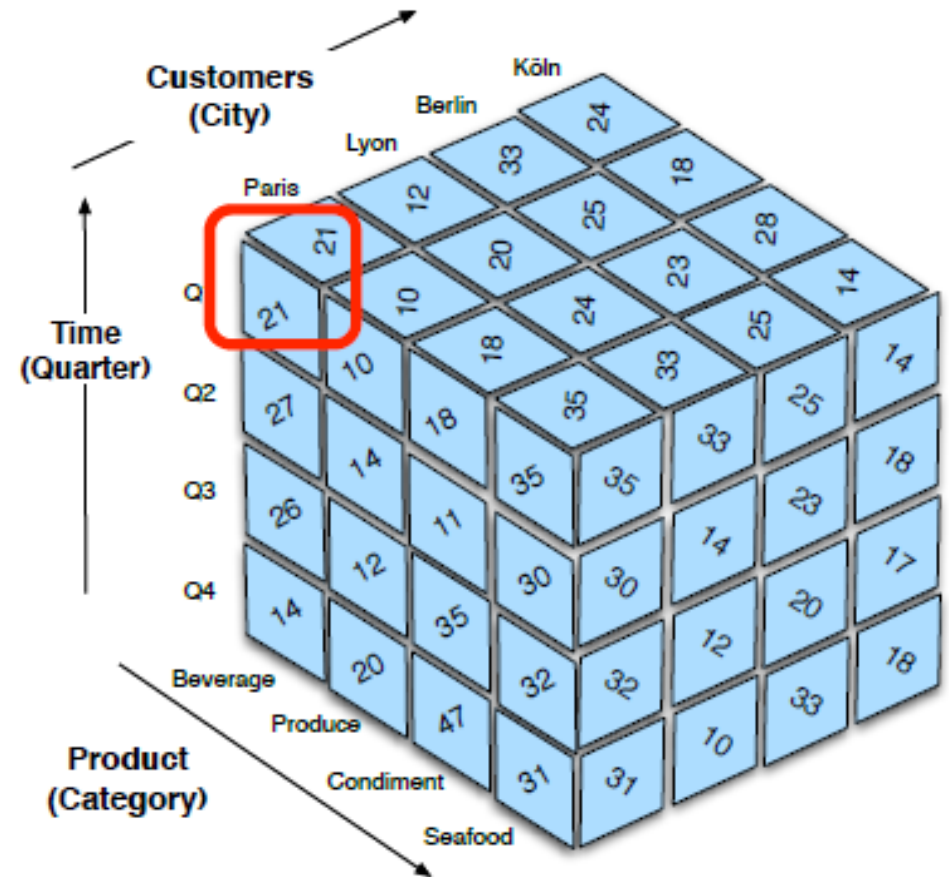
- Structure générale d'une requête :
 - SQL : `SELECT column1, column2, ..., columnN FROM table`
 - MDX : `SELECT axis1 ON COLUMNS, axis2 ON ROWS FROM cube`
- Clause `FROM` spécifie la source de données :
 - en SQL : une ou plusieurs tables
 - en MDX : un ou plusieurs cubes

MDX vs SQL

- La clause **SELECT** indique les résultats (un cube) que l'on souhaite récupérer par la requête :
 - en SQL :
 - une vue des données en 2 dimensions : lignes (rows) et colonnes (columns)
 - les lignes ont la même structure définie par les colonnes
 - en MDX :
 - nombre quelconque de dimensions pour former les résultats de la requête
 - le terme « axe » pour éviter la confusion avec les dimensions du cube
 - pas de signification particulière pour les rows et les columns,
 - mais il faut définir chaque axe : axe1 définit l'axe horizontal et axe2 définit l'axe vertical

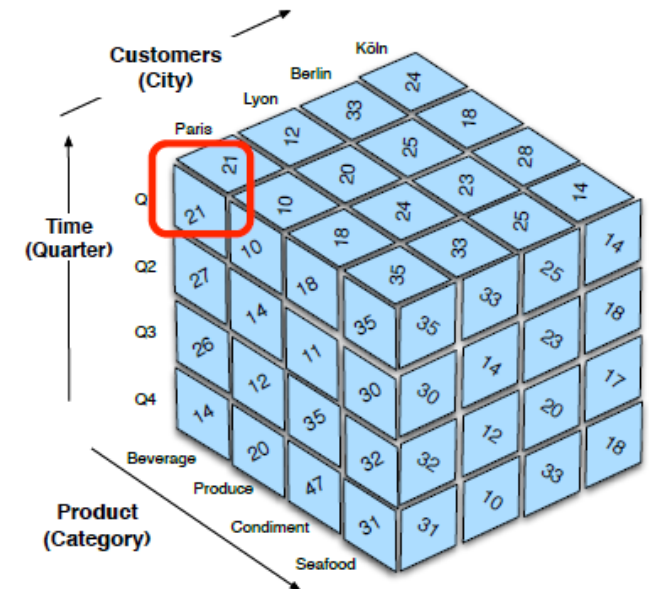
Exemple avec MDX

- Soit le cube « Sales » qui possède :
 - 3 dimensions : Time, Customer, Product
 - Plusieurs mesures : SalesAmount, ...
- La cellule sélectionnée correspond à :
 - La vente des produits de catégorie « Beverage » à Paris durant le premier trimestre (Q1)
- (Product.Category.Beverage ; Time.Quarter.Q1 ; Customer.City.Paris)



Exemple avec MDX

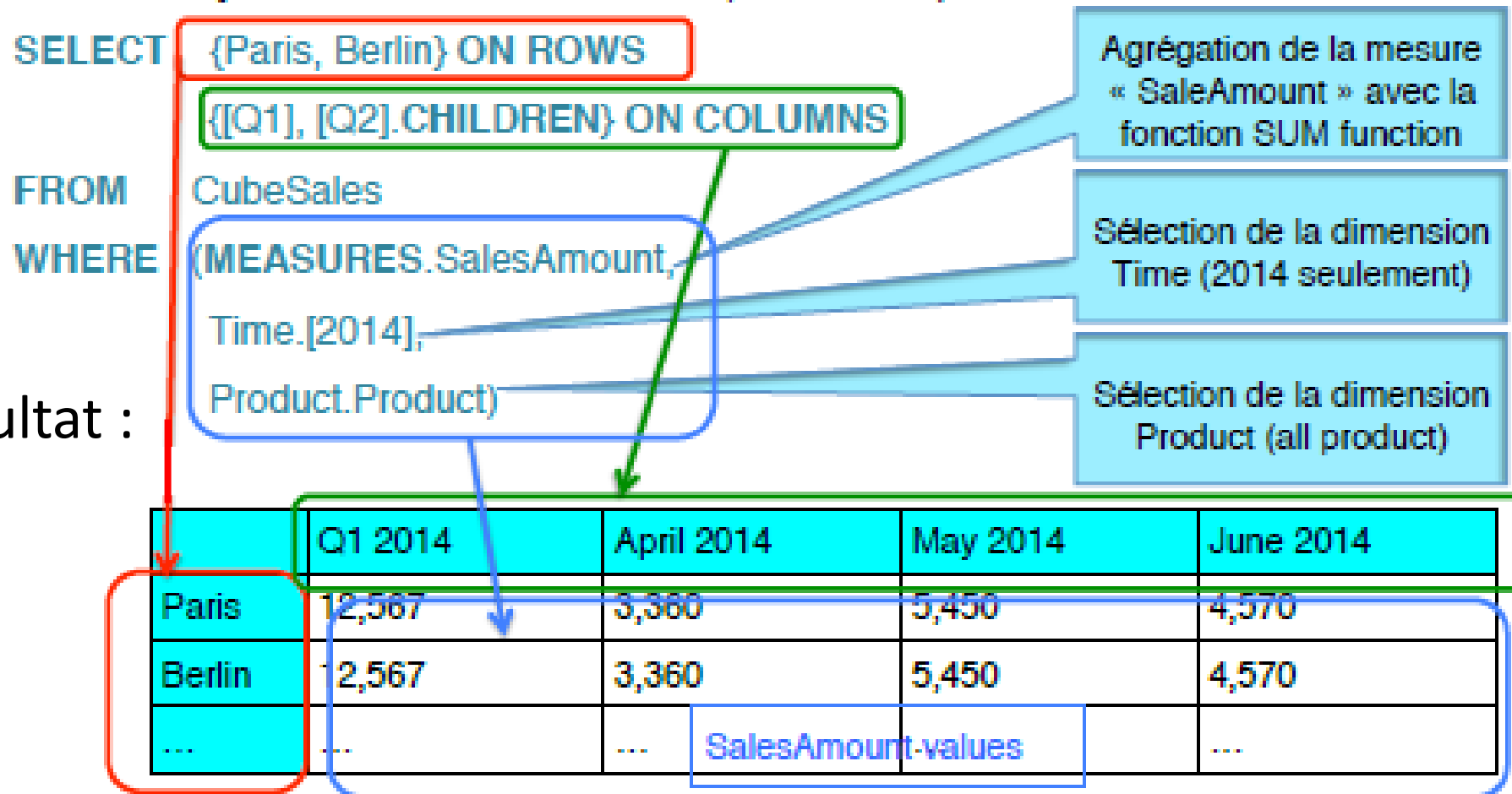
- Le cube « Sales » contient :
 - Measures : UnitPrice, Quantity, Discount, SalesAmount, Freight
 - Dimension : Time
 - hierarchy : Year > Quarter > Month > with members :
 - Year: 2010, 2011, 2012, 2013, 2014
 - Quarter: Q1, Q2, Q3, Q4
 - Month: January, February, March, ...
 - Dimension : Customer
 - hierarchy : Continent > Country > State > City with members :
 - City: Paris, Lyon, Berlin, Köln, Marseille, Nantes ...
 - State: Loire atlantique, Bouches du Rhône, Bas Rhin, Torino...
 - Country: Austria, Belgium, Danmark, France, ...
 - Continent : Europe, North America, Sud America, Asia
 - Dimension : Product
 - hierarchy : Category > Subcategory > product with members :
 - Category : Food, Drink ...
 - Subcategory : Baked_food
 - Product : ...,



Exemple avec MDX

```
SELECT {Paris, Berlin} ON ROWS,  
       {[Q1],[Q2].CHILDREN} ON COLUMNS  
FROM Sales  
WHERE (MEASURES.SalesAmount,  
       Time.[2014],  
       Product.Product)
```

- Soit la requête MDX suivante :



- Résultat :

Partie 2 : Composants d'une requête MDX

Points traités :

- Structure d'une requête MDX

- Membre

- Tuple

- Set

- Axe

Rappel sur les opérations OLAP

- Restructuration : permet un changement de points de vue selon différentes dimensions (opérations liées à la structure, manipulation et visualisation du cube) :
 - Rotate/pivot
 - Switch
 - Nest, Push, ...
- Granularité : concerne un changement de niveau de détail (opérations liées au niveau de granularité des données) :
 - Roll-up,
 - Drill-down
- Ensembliste : concerne l'extraction et l'OLTP classique :
 - Slice, Dice
 - Selection, Projection et Jointure

Structure générale d'une requête MDX

- Syntaxe générale d'une requête MDX :

```
SELECT [<specification d'un axe> [, <specification d'un axe>...]]  
FROM [<specification d'un cube>]  
[WHERE [<specification d'un filtre (slicer)>]]
```

- Parenthèses en MDX :

- { } : Ensemble des éléments servant à la création d'une dimension du résultat de la requête
- () : Sélection de tuples dans la clause WHERE
- [] : Représentation d'espaces, de caractères spéciaux et d'interprétation non numérique de chiffres
 - Optionnels, sauf pour un nom avec des espaces, avec des chiffres, ou des mots-clés

Structure générale d'une requête MDX

- SELECT : description des axes (dimensions) du cube résultat
- Chaque dimension du résultat :
 - est associée à un rôle correspondant à sa représentation dans le tableau retourné par la requête MDX :
 - Ex : `ON COLUMNS, ON ROWS, ON PAGES, ON SECTIONS, ON CHAPTERS`
 - sur un ou plusieurs niveaux de la hiérarchie :
 - Ex1 : `{Paris, Berlin}` de la dimension Customer, niveau Ville
 - Ex2 : `{[Q1], [Q2].CHILDREN}` de la dimension Time, niveaux trimestre et mois

Structure générale d'une requête MDX

- FROM : Spécification du (des) cube (s) de départ
 - Ensemble de cubes nécessaires à la création du cube résultat
 - Si plusieurs cubes sont nécessaires, cela implique une jointure multidimensionnelle : chaque paire de cubes doit alors posséder au moins une dimension concordante
- WHERE : Restriction sur les cubes
 - Restrictions sur les cubes de départ de la clause FROM
 - Spécification des restrictions par une liste de nœuds de la hiérarchie d'une dimension nommée la « slicer-dimension »
- En MDX les mesures sont des éléments d'une dimension spéciale nommée « Measures »
 - Ces mesures peuvent être utilisées aussi dans les clauses WHERE et FROM

Membres en MDX

- Un membre est une instance d'un niveau d'une dimension, généralement spécifié entre crochets [...]
 - Ex : `[Food]`, `[Drink]` = membres de la dimension "Products" de niveau 1
- Les membres sont des éléments accessibles dans les hiérarchies pouvant être référencés de différentes façons :
 - `[2012]`
 - `[Time].[2012]`
 - `[Product].[Food]`
 - `[Product].[Food].[Baked Goods]`
 - `[Product].[All Products].[Food].[Baked Goods]`

Membres en MDX

- Exemples d'utilisations des membres dans des requêtes simples :

```
SELECT [Time].[2012] ON COLUMNS FROM [Sales]
```

```
SELECT [Product].[Food] ON COLUMNS FROM [Sales]
```

```
SELECT [Product].[Food].Children ON COLUMNS FROM [Sales]
```

- Les enfants (CHILDREN) d'un membre sont les membres du niveau immédiatement en dessous de celui-ci dans l'hierarchie
 - Children est une fonction MDX

Tuples en MDX

- Un tuple est une suite de membres entre parenthèses séparés par une virgule :
 - Ex : ([Time].[2012] , [Product].[Food])
 - On peut omettre les parenthèses si on a un tuple avec un seul membre
- Un tuple permet d'identifier une ou plusieurs cellules dans un cube situées à l'intersection de ses membres :
 - `SELECT ([Time].[2012], [Product].[Food]) ON COLUMNS FROM [Sales]`
 - `SELECT ([Product].[All Products].[Food].[Baked Goods], [2012]) ON COLUMNS FROM [Sales]`
- Dans un tuple, les « mesures » sont traitées comme une dimension particulière, nommée [Measures] :
 - `SELECT ([Measures].[Unit Sales], [Product].[Food].[Baked Goods]) ON COLUMNS FROM [Sales]`

Tuples en MDX

- Tuples partiels :
 - Les valeurs de certains membres peuvent être manquantes, par exemple :
 - `([Date].[Année calendrier], [Produit].[Sous-Catégories])`
- Résolution de tuples partiels :
 - Pour chaque valeur manquante d'un membre, appliquer les règles suivantes :
 1. Si une valeur par défaut existe (définie lors de la conception du cube), utiliser cette valeur
 2. Sinon, si le membre (Tous) existe, utiliser cette valeur
 3. Sinon, si le membre (Tous) n'existe pas (ex : dimension Mesure), utiliser le premier membre

Tuples en MDX

- Exemple de résolution de tuples partiels :
 - `[Date].[Année fiscale] ?`
→ `[Date].[Année fiscale].[Toutes périodes]`
 - `[Produit].[Catégorie] ?`
→ `[Produit].[Catégorie].[Tous les produits]`
 - `[Measures].[Mesure] ?`
→ `[Measures].[Mesure].[TotalAvantTaxes]`
- Remarque : On suppose ici qu'aucune valeur par défaut n'a été spécifiée

Ensembles (Sets) dans MDX

- Un « Set » est un ensemble ordonné de tuples défini sur les mêmes dimensions
 - Il commence par une accolade "{", dans laquelle sont énumérés les tuples séparés par des virgules, et se termine par une accolade appariée "}"

```
SELECT {  
    ([Measures].[Unit Sales], [Product].[All  
    Products].[Food].[Baked Goods]),  
    ([Measures].[Store Sales], [Product].[All  
    Products].[Food].[Baked Goods])  
} ON COLUMNS  
FROM [Sales]
```

- Cet ensemble contient :
 - 2 mesures différentes (Units sales et Store Sales) et
 - le même membre (Baked Goods) :

Ensembles (Sets) dans MDX

- Exemple d'un ensemble qui comporte 2 mesures et 2 membres différents de la même dimension sur 2 niveaux différents ([Food] et [Baked Goods]) :

```
SELECT
{ ([Measures].[Unit Sales], [Product].[Food]),
  ([Measures].[Store Sales], [Product].[Food].[Baked Goods]) }
ON COLUMNS
FROM [Sales]
```

Ensembles (Sets) dans MDX

- Exemple d'un ensemble qui a la même mesure et 2 membres contigus différents ([Food] et [Drink]) :

```
SELECT
  { ([Measures].[Unit Sales], [Product].[Food]),
    ([Measures].[Unit Sales], [Product].[Drink]) } ON COLUMNS
FROM [Sales]
```

- Exemple d'un ensemble qui ne contient qu'un seul membre ([2012]) :

```
SELECT
  { ([2012]) } ON COLUMNS          -- ou { [2012] } ON COLUMNS
FROM [Sales]
```


Axes en MDX

- Un axe est un Set suivi par le mot clé « ON », suivi d'un nom d'axe
 - Utilise un numéro s'il y a plus de 2 axes
- Ex: unités vendues "[Measures].[Unit Sales]" par an, en 2012 et 2013 pour les produits "Drink" et "Food" :

```
SELECT
{ ([Measures].[Unit Sales], [Product].[Food]),
  ([Measures].[Unit Sales], [Product].[Drink]) } ON COLUMNS,
{ ([Time].[2012]), ([Time].[2013]) } ON ROWS FROM [Sales]
```

- ou

```
SELECT
{ ([Measures].[Unit Sales], [Product].[Food]),
  ([Measures].[Unit Sales], [Product].[Drink]) } ON AXIS(0),
{ ([Time].[2012]), ([Time].[2013]) } ON AXIS(1) FROM [Sales]
```

Axes en MDX

- Une façon simple pour définir un axe est de présenter sur l'axe tous les membres d'une dimension :
 - `<dimension name>.MEMBERS`
- Si l'on veut voir apparaître tous les membres de la dimension à un certain niveau de cette dimension :
 - `<dimension name>.<level name>.MEMBERS`
- Exemple :

```
SELECT
Years.MEMBERS ON COLUMNS,
Regions.Continent.MEMBERS ON ROWS
FROM Sales
```

Axes en MDX

	2010	2011	2013	2014
<i>N. America</i>	120,000	200,000	400,000	600,000
<i>S. America</i>	-	10,000	30,000	70,000
<i>Europe</i>	55,000	95,000	160,000	310,000
<i>Asia</i>	30,000	80,000	220,000	200,000

- Soit la requête MDX suivante :

```
SELECT  
    Years.MEMBERS ON COLUMNS,  
    Regions.Continent.MEMBERS ON ROWS  
FROM Sales
```

- Son résultat est une table avec 2 axes :
 - l'axe horizontal est (i.e. COLUMNS) contient tous les membres de la dimension Years (2010, 2011, 2013, 2014)
 - l'axe vertical est (i.e. ROWS) contient tous les membres du niveau Continent de la dimension Regions (N & S America, Europe, Asia)
- Remarque :
 - Ici la mesure considérée (auxquelles correspondent les valeurs du résultat) n'est pas précisée, c'est la mesure par défaut (default Measure)

Axes en MDX

- Certaines dimensions ou niveaux ont plus de 1000 membres !
 - On peut souhaiter ne pas considérer tous les membres de la dimension ou du niveau,
 - Dans MDX on peut spécifier une liste de membres à considérer : { dim.member1, dim.member2, ... , dim.memberN }

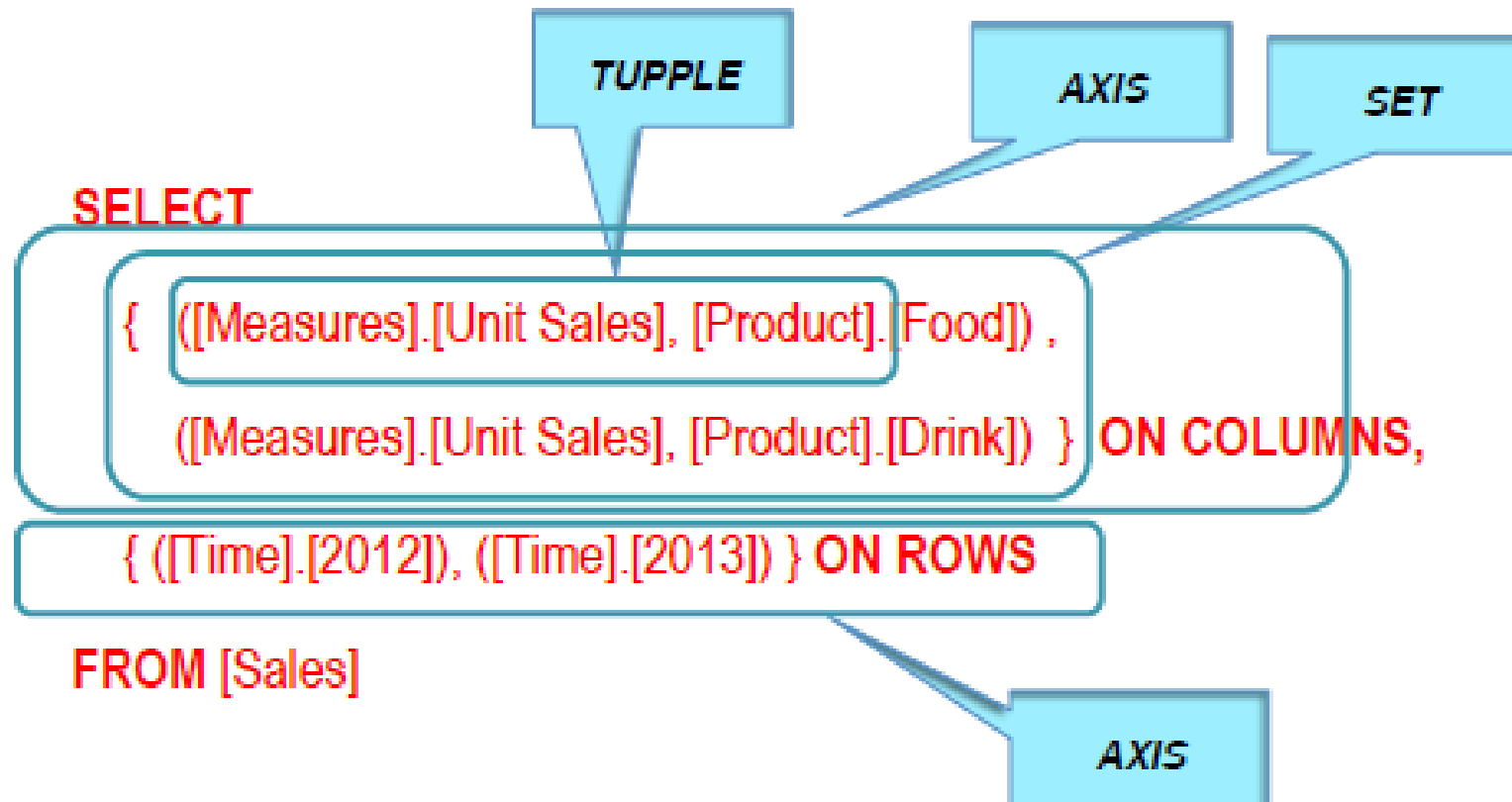
- Ex : On considère seulement les ventes sur les 02 années 2012 et 2013:

```
SELECT  
  { Years.[2012], Years.[2013] } ON COLUMNS,  
  Regions.Continent.MEMBERS ON ROWS  
FROM Sales
```

- Remarques :
 - Il est recommandé de choisir des noms de membres entre crochet [] sans espace, point, ...
 - Ici [2012] et [2013] sont des membres et pas des valeurs !!!
 - Là aussi, la mesure (Measure) n'est pas explicitement définie => mesure par défaut

Structure générale d'une requête MDX

- Exemple récapitulatif :



Exemples de requêtes

```
SELECT {[1997]} ON COLUMNS  
FROM [Sales]
```

ou :

```
SELECT {[Time].[1997]} ON COLUMNS  
FROM [Sales]
```

```
SELECT {([Time].[1997], [Product].[Food])}  
ON COLUMNS  
FROM [Sales]
```

Time
+ 1997
↓ 49 339

Table
Time
+ 1997
Product
+ Food
↓ 35 546

Exemples de requêtes

```
SELECT { ([Product].[All Products].[Food].[Baked Goods],  
          [Time].[1997]) }  
ON COLUMNS  
FROM [Sales]
```



A screenshot of a table viewer window titled 'Table - x'. It displays a single column of data. The column is labeled 'Product' and contains a single row with the value '+ Baked Goods'. Below this, the column is labeled 'Time' and contains a single row with the value '+ 1997'. At the bottom of the table, there is a green downward arrow followed by the number '1 491'.

Product
+ Baked Goods

Time
+ 1997

↓ 1 491



A screenshot of a table viewer window titled 'Table - x'. It displays a comparison of 'Unit Sales' and 'Store Sales' for 'Baked Goods' in 1997. The table has two columns: 'Unit Sales' and 'Store Sales'. Both columns have a 'Product' header and a data row for '+ Baked Goods'. The 'Unit Sales' data row shows a value of '1 491' and the 'Store Sales' data row shows a value of '3 096,87'. Both values are preceded by a green downward arrow.

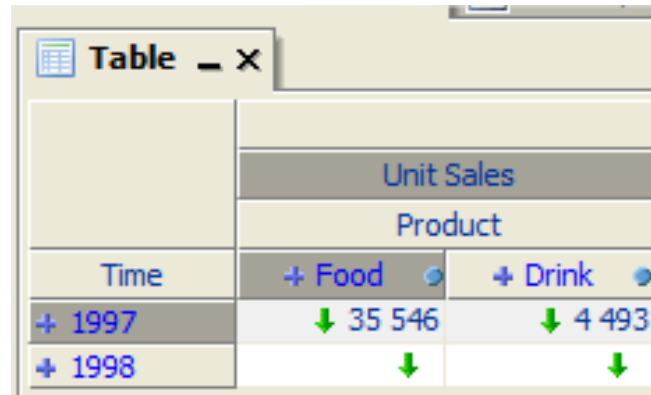
Unit Sales	Store Sales
Product	Product
+ Baked Goods	+ Baked Goods

↓ 1 491 ↓ 3 096,87

```
SELECT { ([Measures].[Unit Sales],  
          [Product].[All Products].[Food].[Baked Goods]),  
          ([Measures].[Store Sales],  
          [Product].[All Products].[Food].[Baked Goods]) }  
ON COLUMNS  
FROM [Sales]
```

Exemples de requêtes

```
SELECT { ([Measures].[Unit Sales], [Product].[Food]),  
        ([Measures].[Unit Sales], [Product].[Drink]) } ON COLUMNS ,  
        { ([Time].[1997]), ([Time].[1998]) } ON ROWS  
FROM [Sales]
```



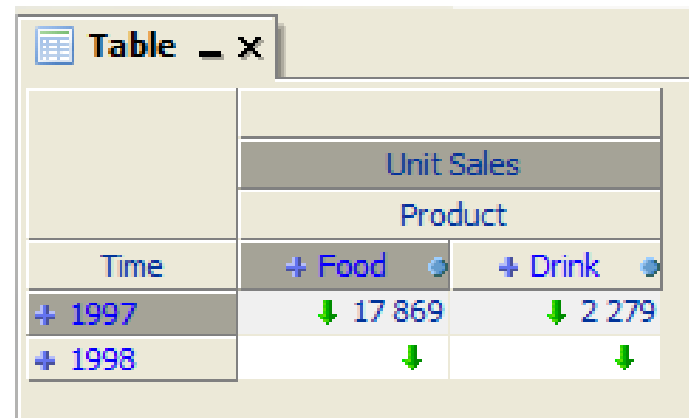
The screenshot shows a PivotTable window titled "Table". The PivotTable has a single data field "Unit Sales". The column labels are "Food" and "Drink", and the row labels are "1997" and "1998". The data values are 35,546 for Food in 1997 and 4,493 for Drink in 1997. The values for 1998 are represented by green downward arrows, indicating that the data is not displayed or is hidden.

	Food	Drink
1997	35 546	4 493
1998	↓	↓

Exemples de requêtes

- Utilisation d'un prédicat de slice

```
SELECT { ([Measures].[Unit Sales], [Product].[Food]),  
        ([Measures].[Unit Sales], [Product].[Drink]) } ON COLUMNS,  
       { ([Time].[1997]), ([Time].[1998]) } ON ROWS  
FROM [Sales]  
WHERE { ([Gender].[M]) }
```



The screenshot shows a PivotTable titled 'Table - X'. The table has a hierarchical structure with 'Time' as the row label and 'Product' as the column label. The 'Unit Sales' measure is displayed in the center. The data is filtered by 'Gender' to 'M'. The table shows data for the years 1997 and 1998, with Food and Drink products. The values are 17,869 for Food in 1997 and 2,279 for Drink in 1997. The values for 1998 are not displayed, only green arrows indicating the data is present.

Unit Sales		
Product		
Time	+ Food	+ Drink
+ 1997	↓ 17 869	↓ 2 279
+ 1998	↓	↓

Partie 3 : Filtres & opérateurs

Points traités :

- Filtres (slicer)

- Commentaires en MDX

Filtres (Slicers) en MDX

- Supposons qu'on ne s'intéresse pas aux ventes de tous les produits, mais uniquement aux ventes d'ordinateurs. On définit alors le nouvel axe :

```
{ Products.[Product Group].Computers }
```

- On pourrait ajouter cet axe à notre requête, mais elle contiendrait alors 3 axes, et tous les outils OLAP ne pourraient pas la visualiser.
- Solution : utiliser une opération de Slice (filtre)

Filtres (Slicers) en MDX

- Dans MDX l'opération de Slice est traitée par une clause WHERE :

```
SELECT
    { Years.[2012], Years.[2013] } ON COLUMNS,
    Regions.Continent.MEMBERS ON ROWS
FROM Sales
WHERE
    ( Products.[Product Group].[Computers] )
```

- La clause WHERE a la syntaxe suivante:

```
WHERE (member-of-dim1, member-of-dim2, ..., member-of-dimn)
```

Filtres (Slicers) en MDX

- Dans un Slice en MDX on peut avoir plusieurs membres, mais ils doivent appartenir à des dimensions différentes (pas de Slice sur 2 produits différents, par ex. computer et printers)
- Exemple : Slice sur un produit (computer) et un client particulier (Mobilis) :
 - Ventes de Computers à Mobilis pour les années 2012 et 2013 pour tous les continents :

```
SELECT
{ Years.[2012], Years.[2013] } ON COLUMNS,
  Regions.Continent.MEMBERS ON ROWS
FROM Sales
WHERE
( Products.[Product Group].[Computers],
  Customers.[Mobilis] )
```

Filtres (Slicers) en MDX

- On peut souhaiter voir plutôt que les ventes (la mesure par défaut) de computer, le nombre d'unité expédiées
 - On doit juste ajouter la mesure « Units » dans le Slice
- Exemple : Nombre d'unités de Computers expédiées à Mobilis pour les années 2012 et 2013 pour tous les continents :

```
SELECT
    { Years.[2012], Years.[2013] } ON COLUMNS,
    Regions.Continent.MEMBERS ON ROWS
FROM Sales
WHERE
    ( Products.[Product Group].[Computers],
      Customers.[AT&T],
      Measures.[Units] )
```

Filtres (Slicers) en MDX

- Résumé :
 - Le WHERE permet de sélectionner une tranche du cube
 - On spécifie la tranche en utilisant des membres qui ne font pas partie des dimensions utilisées au niveau des axes ON ROWS et ON COLUMNS
 - Il n'est pas autorisé de préciser une tranche avec plus d'un membre d'une même dimension

```
SELECT { ([Measures].[Unit Sales]) } ON COLUMNS,  
        { ([Time].[Year].Members) } ON ROWS  
FROM SALES  
WHERE   ([Store].[USA].[WA])
```

Commentaires en MDX

- Les commandes MDX peuvent être commentées de trois façons différentes :
 - `//` Commentaire en fin de ligne
 - `--` Commentaire en fin de ligne
 - `/*` Commentaire sur plusieurs lignes
`*/`
- Les commentaires peuvent être imbriqués :
 - `/*` Commentaire sur
 - plusieurs lignes `/*` Commentaire imbriqué `*/`
 - `*/`