



UNIVERSITAT DE BARCELONA



## Tema 2: Bases de dades Multidimensionals

### Bases de dades avançades curs 17/18

Enric Biosca Trias [ebiosca@maia.ub.es](mailto:ebiosca@maia.ub.es)

Dept. Matemàtica Aplicada i Anàlisi.

**Universitat de Barcelona**

**OLAP** es online analytical processing. Es tracta d'una forma d'emmagatzemar la informació en una base de dades que permeti realitzar de forma més efectiva les consultes. És una definició abreujada, és clar, la realitat és més complexa.

**MOLAP**, Multidimensional OLAP. Tant les dades font com les dades agregades o precalculades resideixen en el mateix format multidimensional. Optimitza les queries, però requereix més espai de disc i diferent programari. El primer punt aquesta deixant de ser un problema: l'espai de disc cada vegada és més barat.

**ROLAP**, Relational OLAP. Tant les dades precalculades i agregades com les dades font resideixen a la mateixa base de dades relacional. Si el DataWarehouse és molt gran o es necessita rapidesa per part dels usuaris pot ser un problema.

**HOLAP**, Hybrid OLAP: És una combinació dels dos anteriors. Les dades agregades i precalculades s'emmagatzemen en estructures multidimensionals i els de menor nivell de detall en el relacional. Requereix una bona feina d'anàlisi per identificar cada tipus de dada.

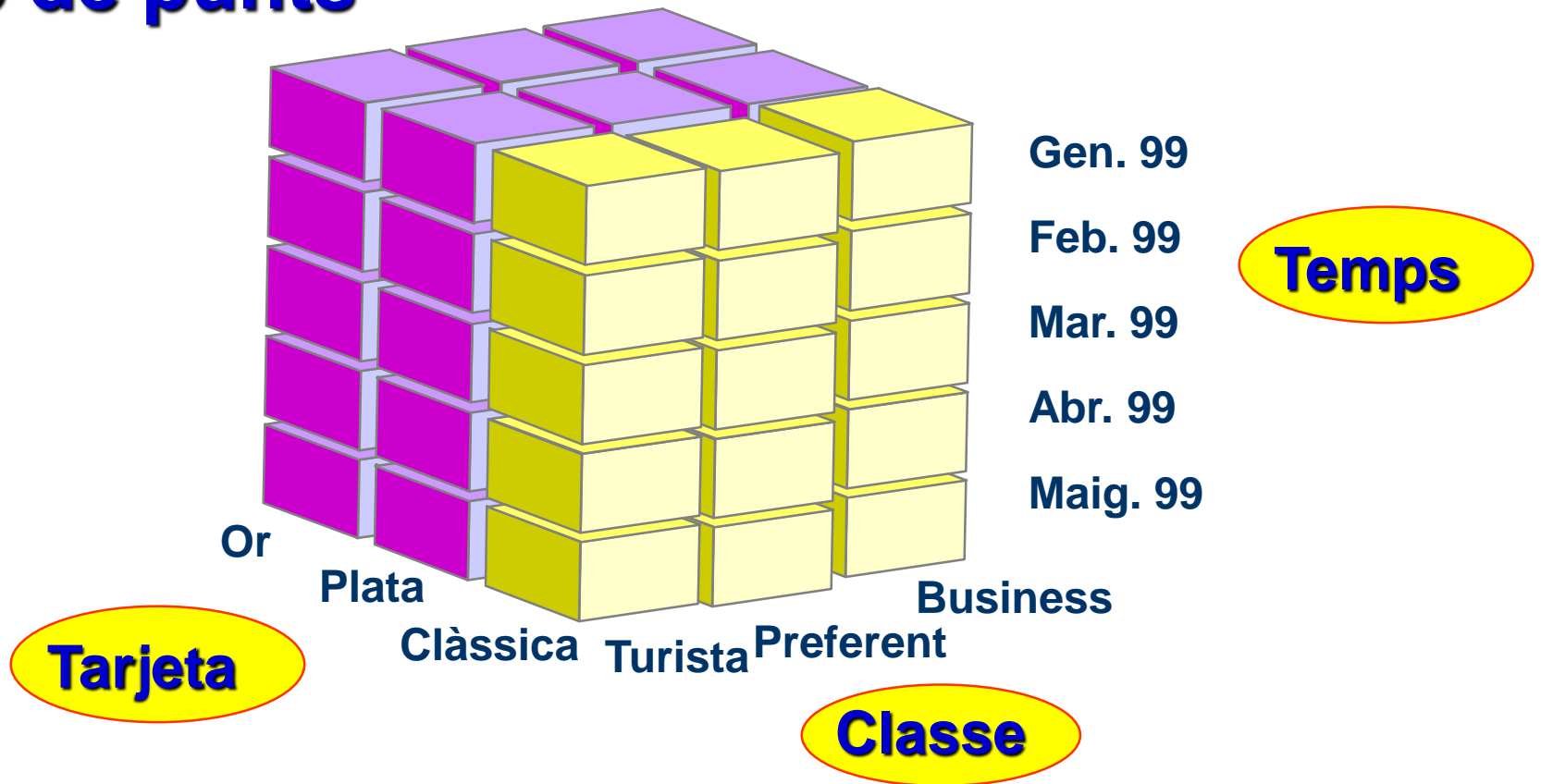


**Els 6 elements bàsics OLAP:**

- **Dimensions**
- **Métriques**
- **Jerarquies**
- **Nivells**
- **Atributs**
- **Indicadors**



## Cub de punts





UNIVERSITAT DE BARCELONA

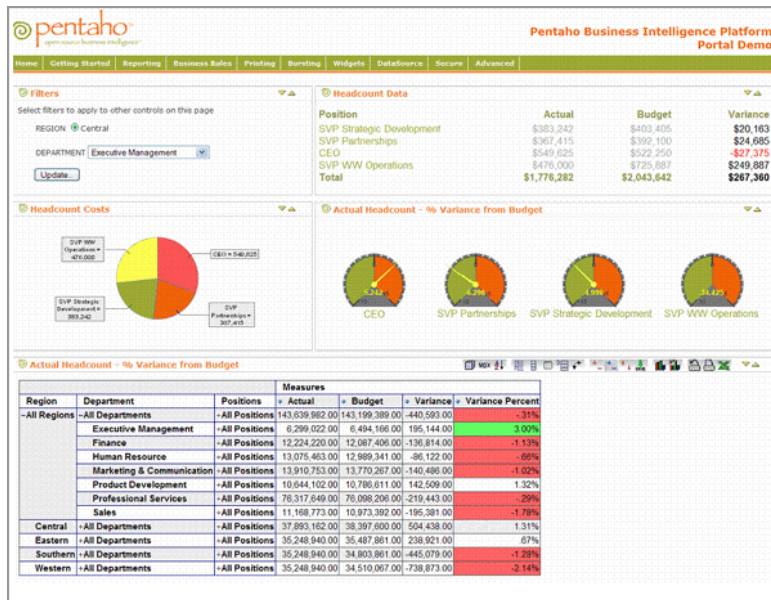


# Anàlisi OLAP

## Cub de punts



Per obtenir la funcionalitat de processament (OLAP) s'utilitzen dues aplicacions: el servidor OLAP Mondrian, que combinat amb jpivot, permeten realitzar querys a Datamarts, que els resultats siguin presentats mitjançant un browser i que l'usuari pugui realitzar drill down i altres navegacions típiques.

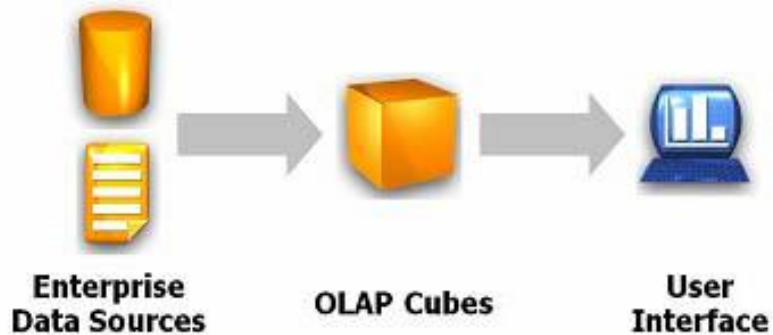


**Mondrian** funciona sobre les bases de dades estàndards de mercat: Oracle, DB2, SQL-Server, MySQL, Postgres... la qual cosa habilita i facilita el desenvolupament de negoci basat en la plataforma Pentaho.

**JPivot** està considerat com un "projecte germà" de Mondrian. És el client per defecte de Mondrian, però no l'únic. També serveix com a client per altres motors.

# Mondrian

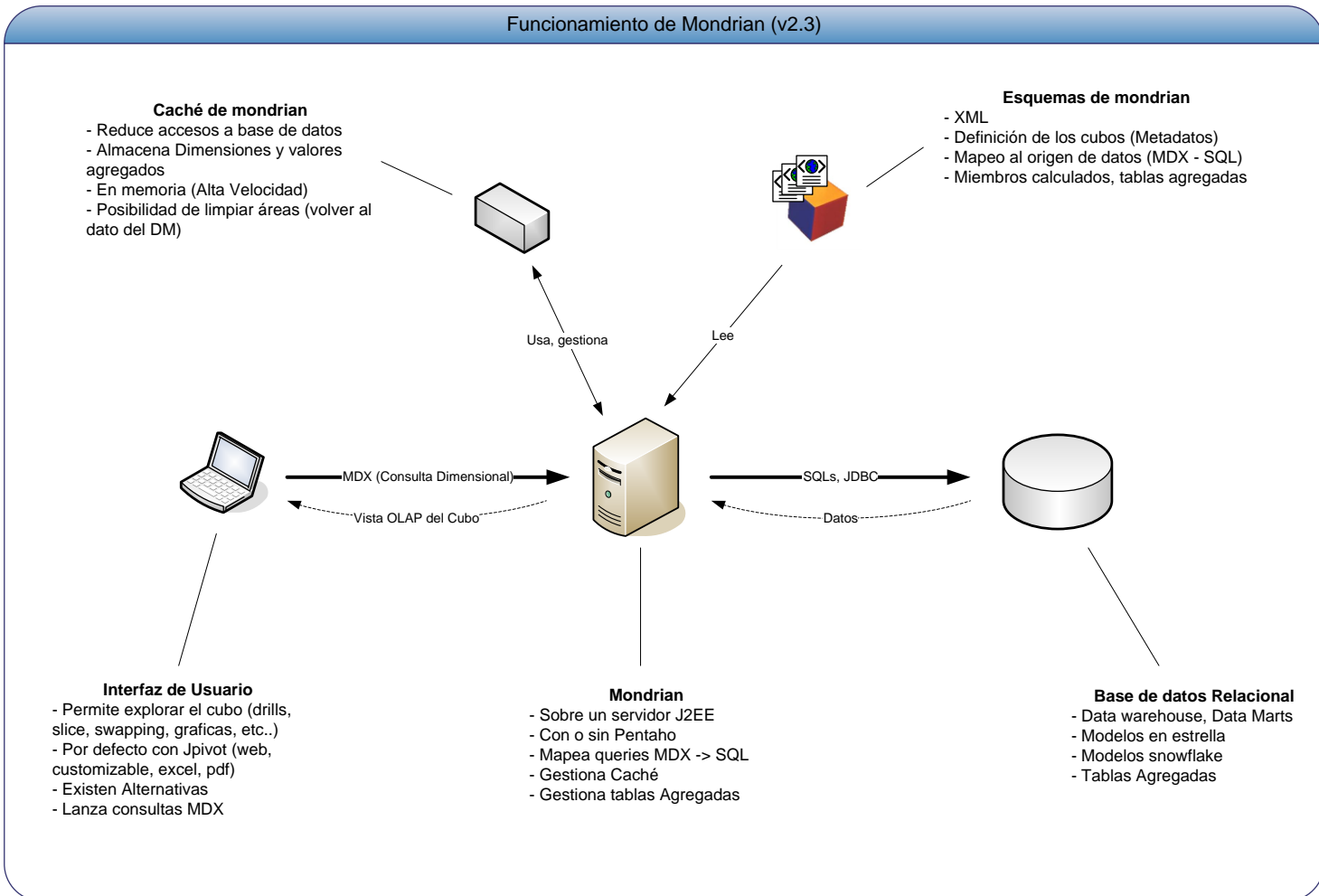
Ets un servidor OLAP Mondrian que està escrit en Java.  
Permet interactuar amb grans quantitats de dades emmagatzemades en una base de dades relacional, sense necessitat de sentències SQL complexes.





# Mondrian: Arquitectura

## Funcionamiento de Mondrian (v2.3)





Jpivot és un conjunt de llibreries personalitzades JSP que permet mostrar taules i gràfics, per mostrar la navegació típica dels entorns OLAP: drill-down, rotar eixos, drill-through, etc ...

Utilitza Mondrian com a servidor OLAP preferent, però també podria accedir als cubs OLAP de Microsoft Analysis Services.

La connexió amb les bases de dades es realitza via JDBC i realitza els càlculs en memòria, sense generar nous fitxers i bases de dades que mantenir i emmagatzemar.

## Barra de Navegació:





	Measures		
	Inicial		
	Tiempo		
Presupuestos	2004	2005	2006
-CAPITULO I - GASTOS DE PERSONAL	151,601,484.00	302,153,628.00	407,219,592.00
+ARTICULO 10. ALTOS CARGOS	21,402,816.00	46,059,948.00	54,823,044.00
+ARTICULO 11. PERSONAL EVENTUAL	45,437,520.00	90,504,864.00	118,141,764.00
+ARTICULO 12. FUNCIONARIOS	66,407,748.00	126,277,008.00	183,482,724.00
+ARTICULO 13. LABORALES	10,061,472.00	20,856,384.00	27,028,620.00
+ARTICULO 14. OTRO PERSONAL	142,968.00	321,240.00	415,992.00
+ARTICULO 15. INCENTIVOS AL RENDIMIENTO	7,202,736.00	15,708,504.00	20,629,416.00
+ARTICULO 16. CUOTAS, PRESTACIONES Y GASTOS SOC. A CARGO DEL EMPLEADOR	946,224.00	2,425,680.00	2,698,032.00

**Columns**

- Measures
- Tiempo

**Rows**

- Presupuestos

**Filter**

- Geografia ((All)=All Geografia)
- Organos Gestores ((All)=All Organos)

OK Cancel

**Tiempo**

☐ - All Geografia

- ☒ 2004
- ☒ 2005
- ☒ 2006

None Flat OK Cancel



	Measures							
	Diferencia							
	Geografia							
Presupuestos	Almeria	Cadiz	Cordoba	Granada	Huelva	Jaen	Malaga	Sevilla
+CAPITULO I - GASTOS DE PERSONAL	-189926856	122074656	-274446900	-1373480640	129441816	172278984	77199600	-189878220
+CAPITULO II - GASTOS CORRIENTES EN BIENES Y SERVICIOS	-137928888	87841776	47799000	-910705824	95912004	-281847468	-669525600	-55825560
+CAPITULO IV - TRANSFERENCIAS CORRIENTES	-10705104	-43802352	103639500	-376947648	-32825448	-308414604	-145035600	-636860700
+CAPITULO VI - INVERSIONES REALES	11297520	-26088480	-39047400	139537440	-10326564	-3045024	73116000	194363820
+CAPITULO VII - TRANSFERENCIAS DE CAPITAL	-6085728	-9328704	15119100	-124735968	-21491160	4358016	-30526800	-18908820



Slicer: (All)=All Organos, Año=2005

● Diferencia.Almeria. 
 ● Diferencia.Cadiz. 
 ● Diferencia.Cordoba. 
 ● Diferencia.Granada. 
 ● Diferencia.Huelva. 
 ● Diferencia.Jaen. 
 ● Diferencia.Malaga. 
 ● Diferencia.Sevilla.



# Introducció al MDX

- 1. Introducció a l'anàlisi multidimensional i cubs OLAP**
- 2. Què és MDX?**
- 3. Principals aspectes de la sintaxi de MDX**
- 4. Dimensions de temps en MDX: funcions especials**
- 5. Membres calculats i fórmules complexes**

Abans de començar a parlar d'MDX hem d'entendre que significa l'anàlisi multidimensional i els cubs OLAP ja que el llenguatge MDX s'utilitza per realitzar consultes sobre aquests models de dades.

L'anàlisi multidimensional consisteix en organitzar la informació i les dades d'una organització de manera que els elements mesurables o mètriques del negoci (també anomenats fets) puguin ser qualificats utilitzant les diferents "Dimensions" o vistes del negoci.

L'objectiu d'aquest model és permetre analitzar una mètrica creuant amb diferents punts de vista del negoci. Aclarim-ho amb un exemple:

Posem que tenim un supermercat amb una targeta de fidelització que ens permeti saber qui compra què i quan.

Una anàlisi o organització multidimensional de les dades de vendes d'un supermercat ens podria permetre respondre a preguntes com:

Quin article és el més venut aquest mes?

Quina botiga ha venut més articles del tipus X?

Que edat tenen els clients que més gasten en el meu supermercat (de mitjana)?

Vegem el paral·lelisme entre aquestes consultes i el model multidimensional.

Quin article és el més venut aquest mes?

En aquesta pregunta volem analitzar el fet articles venuts i ho volem analitzar des de 2 dimensions diferents: temps i família de productes.

Quin supermercat ha venut més articles del tipus X?

El fet a comptar aquí és el mateix (quantitat d'articles) però les dimensions són aquí les diferents botigues i la família de productes

Quina edat tenen els clients que més gasten en el meu supermercat (de mitjana)?

En aquest cas el fet és l'import de les vendes i la dimensió són els clients, o bé l'edat dels clients, o bé el seu rang d'edat.



UNIVERSITAT DE BARCELONA

# Introducció al MDX

## Què és MDX?

MDX és un acrònim de MultiDimensional query eXpression. Aquest llenguatge va ser creat el 1997 per Microsoft. No és un llenguatge estàndard però diferents fabricants d'eines OLAP l'han adoptat com estàndard de fet.

El MDX és als sistemes OLAP l'equivalent al SQL per als sistemes gestors de bases de dades relacionals. Això vol dir que és el llenguatge a través del qual podem explotar la informació que resideix en els motors OLAP i satisfer les consultes analítiques de les que hem parlat.

És un llenguatge de consulta més proper al llenguatge natural que sql però té funcions i fórmules que el fan molt potent per a l'anàlisi de dades.

La principal diferència del món OLAP respecte al món relacional és que les estructures dimensionals estan jerarquitzades i es representen en forma d'arbre i per tant existeixen relacions entre els diferents membres de les dimensions.



UNIVERSITAT DE BARCELONA



# Introducció al MDX

## Què és MDX?

Aquest fet fa que el llenguatge MDX disposi de funcions i atributs especials que permeten referenciar els diferents elements de les dimensions amb expressions com MEMBRES-FILLS, MEMBRES-COSINS, MEMBRES-PARE, etc ... fent una analogia amb els arbres genealògics. Aquestes estructures jerarquitzades són especialment útils per poder visualitzar les dades a nivell temporal i fer comparacions entre períodes.



# Introducció al MDX

## Principals aspectes de la sintaxi de MDX

El llenguatge MDX permet realitzar consultes amb una sintaxi semblant al llenguatge natural.

Per exemplificar les consultes i la sintaxi treballarem amb un exemple d'un cub.

Imaginem un cub de vendes amb les següents dimensions:

- Dimensió temporal de les vendes amb nivells d'any i mes.
- Dimensió productes venuts amb nivells de família de productes i productes.
- Dimensió de mesures amb import de les vendes i unitats venudes

Si volem obtenir per exemple l'import de les vendes per a l'any 2008 per a la família de productes làctics la consulta seria la següent:

**SELECT**

```
{[Mesures]. [Import vendes]}
```

**on columns,**

```
{[Tems]. [2008]}
```

**on rows**

**FROM** [cub vendes]

**WHERE** ([Família]. [Lactics])

# Introducció al MDX

## Principals aspectes de la sintaxi de MDX

Analitzant la consulta d'exemple podem establir una analogia amb el sql per explicar millor l'estructura del llenguatge. En primer lloc tenim l'estructura idèntica general de la consulta amb les clàusules **SELECT** ... **FROM** ... **WHERE** que en MDX s'usa idènticament al sql especificant en el select un conjunt d'elements que volem visualitzar, al `from` indiquem el cub del qual volem treure la informació i al where les condicions de filtrat.

Una de les particularitats de l'MDX és que la informació la torna en forma tabular normalment i ha de ser processada. Si us fixeu en la clàusula select s'usa la clàusula on columns. Aquesta clàusula serveix per separar els elements que volem visualitzar en les files dels que volem a les columnes. A la consulta de exemple veurem l'import en les columnes i el temps en les files. Els {} són per permetre llistes d'elements en les seleccions i els [] encapsulen elements de les dimensions i nivells.

A nivell de funcionalitats i potència a l'hora de realitzar consultes MDX és potent com l'SQL encara que el seu objectiu està orientat a temes de comparacions, relacions jeràrquiques entre elements, etc ...

Una de les funcionalitats que distingeixen el MDX és el accedir als elements utilitzant estructura d'arbre. Així per a un determinat nivell d'una dimensió tenim comandaments com:

[Família]. [Lactics]. CurrentMember: **Permet accedir al membre actual**

[Família]. [Lactics]. Children: **Permet accedir als fills de la família dels lactics**

[Família]. [Lactics]. PrevMember: **Permet accedir al membre anterior de la dimensió**

D'altra banda existeixen diferents funcions que permeten realitzar càlculs i complementar les consultes com per exemple:

- `CrossJoin(conjunt_a, conjunt_b)`: Obté el producte cartesià de dos conjunts de dades.
- `BottomCount(conjunt, N)`: Obtenir un nombre determinat d'elements d'un conjunt, començant per baix, opcionalment ordenat.
- `BottomSum(conjunt, N, S)`: Obtenir d'un conjunt ordenat els N elements el total és com a mínim el especificat (S).
- `Except(conjunt_a, conjunt_b)`: Obtenir la diferència entre dos conjunts.



UNIVERSITAT DE BARCELONA



# Introducció al MDX

## Principals aspectes de la sintaxi de MDX

Finalment comentar que en MDX també estan implementades moltes funcions matemàtiques i estadístiques que permeten enriquir les anàlisis. Trobem entre moltes altres funcions com:

AVG

COUNT

Variance

VARIANCEn

i totes les funcions trigonomètriques (Sinus, Cosinus, tangent, etc ...)

Com ja hem comentat una de les funcionalitats destacades del llenguatge MDX són les funcions especials de tractament del temps. En l'anàlisi multidimensional és molt freqüent comparar els valors de les mètriques amb els valors de períodes anteriors. Per exemple comparar les vendes de juny de 2009 amb les vendes de juny de 2008, 2007, 2006, etc .... També és molt comú fer operacions acumulades amb períodes mòbils: per exemple les vendes dels últims 6 mesos o des de gener fins al mes actual, etc ...

Per tot això MDX permet identificar i tractar de forma especial les dimensions temporals, per poder aplicar les fórmules que permeten obtenir els càlculs que hem esmentat en el paràgraf anterior. A continuació anomenem algunes de les principals fórmules relacionades amb les dimensions temporals i la seva sintaxi.

### PeriodsToDate

Funcionalitat: Retorna un conjunt de membres del mateix nivell que un membre determinat, començant pel primer membre del mateix nivell i acabant amb el membre en qüestió, d'acord amb la restricció del nivell especificat en la dimensió de temps.

Sintaxi: `<Conjunt> PeriodsToDate (<Nivell>, <Membre>)`

Exemple: `PeriodsToDate ([Temps]. [Mensual]. [Semestre], [Temps]. [Mensual]. [Mes]. [Agost 2009])`

El següent exemple tornaria el conjunt de mesos des d'agost del 2009 fins al final del semestre és a dir juliol i agost de 2009.

Existeixen funcions especials per simplificar l'ús de PeriodsToDate en les seves versions més comuns i són les **següents**:

**WTD (<Membre>)**: Retorna els membres de la mateixa **setmana** del membre especificat.

**QTY (< Membre >)**: Retorna els membres del mateix **trimestre** del membre especificat.

**YTD (< Membre >)**: Retorna els membres d'aquest **any** del membre especificat.

### **ParallelPeriod**

Funcionalitat: Retorna un membre d'un període anterior a la mateixa posició relativa que el membre especificat.

Sintaxi: <Membre> ParallelPeriod (<Nivell>, <Expressió numèrica>, <Membre>)

Exemple: *ParallelPeriod ([Temps]. [Mes]. [Semestre], 3, [Temps]. [Mensual]. [Mes]. [Juny 2009])*

El següent exemple tornaria desembre 2007 que és l'equivalent a l'últim mes del semestre retrocedint 3 semestres en el temps.

Hi ha més funcions de temps però amb aquestes només pretenc il·lustrar el concepte i la utilitat d'aquestes funcions.



# Introducció al MDX

## Membres calculats i fórmules complexes

Una de les funcionalitats més potents que ofereix el llenguatge MDX és la possibilitat de realitzar càlculs complexos tant dinàmics (en funció de les dades que s'estan analitzant en aquest moment) com estàtics. Els cubs multidimensionals treballen amb mesures (de l'anglès measures) i amb membres calculats (calculated members). Les mesures són les mètriques de la taula de fets a les quals s'aplica una funció d'agregació (count, DISTINCT count, sum, màx, avg, etc ...), en altres paraules són les mètriques del negoci en estat pur. Els membres calculats són tractats com mètriques a l'hora d'explotar els cubs, però són molt més potents que aquestes.

Un membre calculat és una mètrica que té com a valor el resultat de l'aplicació d'una fórmula que pot utilitzar tots els elements disponibles en una galleda, així com altres funcions d'MDX disponibles. Vegem un exemple senzill:



# Introducció al MDX

## Membres calculats i fórmules complexes

Imaginem que tenim dos mètriques en el nostre cub que són:

`[Mesures]. [Import vendes]`

`[Mesures]. [Nombre d'articles venuts]`

Podríem obtenir el preu mitjà d'article utilitzant un membre calculat amb la següent formula:

$$[Mesures]. [Preu mitjà article] = [Mesures]. [Import vendes] / [Mesures]. [Nombre d'articles venuts]$$

Els membres calculats són realment molt potents ja podem utilitzar expressions condicionals per mostrar diferent informació en funció del nivell o les dimensions que estem visualitzant. Un segon exemple una mica més complex:

```
IIF ([Tiempo.mensual]. CurrentMember.Level IS  
[Tiempo.mensual]. [Mes], [Measures]. [Mitjana mensual], null)
```

En l'exemple anterior utilitzem una funció condicional d'MDX anomenada *IIF* que ens permet assignar valor a un membre calculat només quan estem en un nivell determinat d'una dimensió. Concretament aquest membre calculat només té valor quan estem en el nivell *Mes* de la dimensió temporal, en qualsevol altre cas el valor és nul.

```
sum ({ParallelPeriod ([temps]. [període],1,  
[Tems]. CurrentMember).Lag (5):  
ParallelPeriod ([temps]. [Període], 1,[Tems]. CurrentMember)  
}, [Import Vendes])
```

Aquí podem veure un exemple de com es defineix una fórmula més complexa i com es dona un format especial en funció del valor a aquesta formula.

La fórmula està utilitzant una funció d'agregació per sumar la mètrica [Import Vendes] fent l'acumulat dels darrers 5 valors del nivell (5 anys, 5 mesos o 5 setmanes) utilitzant la funció de temps ParallelPeriod amb l'interval dels últims 5 membres (amb ajuda de la funció Lag (5))