

6. Procesamiento analítico en línea (OLAP)

6.1. Introducción

Las bases de datos relacionales tienen como características principales:

1. Proporcionar métodos simples y confiables de almacenamiento y procesamiento de datos, y modos flexibles de acceso a los mismos.
2. Apoyar las funciones diarias de las empresas con aplicaciones de negocios que almacenan y analizan datos con precisión y confiabilidad.
3. Dar soporte a sistemas de información que también se conocen como de OLTP (On Line Transaction Processing).
4. Ser una fuente valiosa de datos para entender las tendencias y el modo en que la empresa funciona.

Ejemplos de sistemas OLTP son: registro de compras, registro de ventas, control de producción. Sin embargo, los sistemas OLTP no son muy apropiados para realizar análisis de datos.

Un sistema **OLAP (On Line Analytic Processing)** está hecho para llevar a cabo, principalmente, análisis de datos. El término OLAP se refiere a un conjunto de conceptos, lenguajes y productos cuyo objetivo central es facilitar el análisis de datos. Ejemplos de sistemas OLAP son: planeación de recursos, presupuestos, análisis de ventas.

Los requerimientos funcionales principales de los sistemas OLAP son:

A) Lógicos

1. Permitir una estructuración dimensional poderosa de la información con manejo de jerarquías. Ésta es una característica fundamental para poder modelar los conjuntos altamente multidimensionales del mundo real, con muchos niveles de datos, detalles o abstracciones.
2. Permitir una especificación eficiente de dimensiones y de cálculos dimensionales. Los análisis de datos realizados, además de incluir sumas o promedios de cantidades grandes de datos, involucran comparaciones inteligentes de valores e inferencias de tendencias sobre el tiempo y otras dimensiones, por lo que estas actividades deben efectuarse con gran eficiencia.
3. Brindar flexibilidad en las operaciones. Esto implica flexibilidad en la visualización de la información (gráficas, matrices, cartas, etc.), en las definiciones (formateo de datos, definiciones de fórmulas, etc.), en el análisis y en las interfaces (navegación en un modelo, ligas a fuentes externas, etc.).

B) Físicos

1. Separar la estructura interna de almacenamiento de los datos de su representación. Para brindar la posibilidad de hacer cambios en las vistas de los datos sin tener que cambiar la estructura interna de almacenamiento de los mismos.
2. Tener la velocidad suficiente para apoyar análisis ad hoc. Éste es un componente crucial de un sistema OLAP, ya que necesita soportar consultas analíticas ad hoc, algunas de las cuales pueden requerir cálculos realizados “al vuelo”. Para un análisis dado, puede ser necesario hacer una serie de consultas encadenadas las cuales deben responderse rápidamente para no perder el hilo del análisis.
3. Brindar soporte multi-usuario. Es deseable que un grupo de personas trabaje simultáneamente para que haya coincidencia hacia metas comunes; por ejemplo, una institución bancaria puede proyectar los ingresos anuales que espera obtener por la venta de sus productos (tarjeta, cheques, etc.), tomando como base las estimaciones de los gerentes y ajustando, después, con las expectativas de los agentes de ventas.

Comparación entre actividades OLTP vs. OLAP

La siguiente tabla muestra un comparativo entre las actividades principales en ambos tipos de sistemas:

Actividades operacionales (OLTP)	Actividades de análisis (OLAP)
Más frecuentes	Más frecuentes
Más predecibles	Menos predecibles
Cantidades de datos más pequeñas accedidas por consulta	Grandes cantidades de datos accedidas por consulta
Consultas mayoritarias sobre datos “sin procesar”	Consultas mayoritarias sobre datos derivados
Requieren datos actuales mayoritariamente	Requieren datos pasados, presentes y proyectados
Si hay derivaciones complejas, normalmente son pocas	Muchas derivaciones complejas

6.2. Modelo de datos multidimensional

Es un modelo para conceptualizar y visualizar los datos como un conjunto de variables (dimensiones) que son definidas por aspectos comunes del negocio. Es especialmente útil para resumir y reordenar los datos en diferentes vistas de los mismos con el fin de realizar su análisis. El análisis dimensional se enfoca principalmente en los datos numéricos como: valores, medidas, balances y ocurrencias.

6.2.1. Espacio multidimensional

Se usa el término **hipercubo** (o **cubo**, de manera abreviada) para describir un espacio de datos multidimensional. Hay que aclarar que no es un cubo en el sentido geométrico (el cual sólo tiene tres dimensiones), sino que el hipercubo puede tener cualquier cantidad de dimensiones,

cada una, posiblemente, con un tamaño distinto. El hipercubo contiene una cantidad discreta (esto es, no continua) de valores en cada dimensión.

6.2.2. Definiciones relacionadas con los cubos

A continuación se describen varias definiciones asociadas al manejo de los cubos:

- Una **dimensión** describe algún elemento en los datos que el negocio quiere analizar. Por ejemplo, las ventas de productos o el tiempo son dimensiones bastante comunes.
- Un **elemento (miembro)** corresponde a un punto dentro de una dimensión. Por ejemplo, en la dimensión del tiempo, el mes de *marzo* podría ser un elemento.
- Un **atributo** es una colección completa de elementos. Por ejemplo, todos los meses del año serían un atributo de la dimensión del tiempo. Los trimestres del año podrían ser otro atributo de esta dimensión. De los diversos atributos de una dimensión se escoge uno como *atributo clave*.
- El **tamaño, o cardinalidad**, de un atributo de una dimensión es la cantidad de elementos que contiene. Por ejemplo, un atributo de la dimensión del tiempo formado por los meses del año tendría una cardinalidad de 12.

La siguiente figura muestra un cubo con tres dimensiones: tiempo (en meses), productos (por su nombre) y clientes (por su nombre). El cubo define un espacio multidimensional de ventas de un producto específico a clientes específicos sobre un período específico de tiempo, medido en meses.

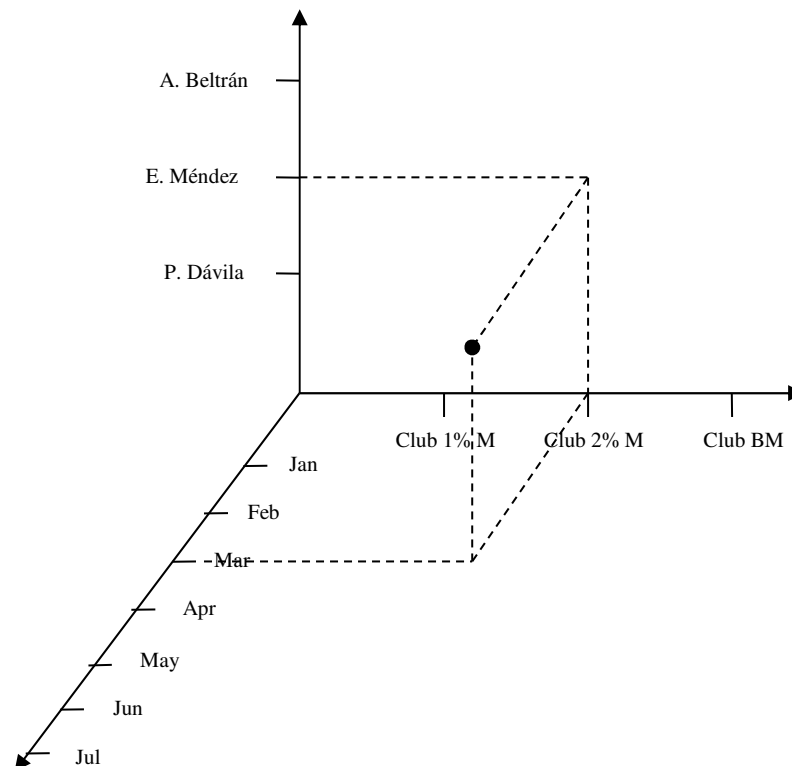


Fig. 6.1. Espacio de datos tri-dimensional que describe ventas de productos a clientes en un período de tiempo.

Más definiciones:

- Un **espacio de hechos**, **datos de hechos** o, simplemente, **hechos**, es el conjunto de puntos en el espacio de datos. Por ejemplo, una venta realizada sería un hecho.
- Una **tupla** es una coordenada en el espacio multidimensional.
- Una **rebanada (slice)** es una sección del espacio multidimensional (ver Fig. 6.2). Un término más adecuado para rebanada sería **sub-cubo**.

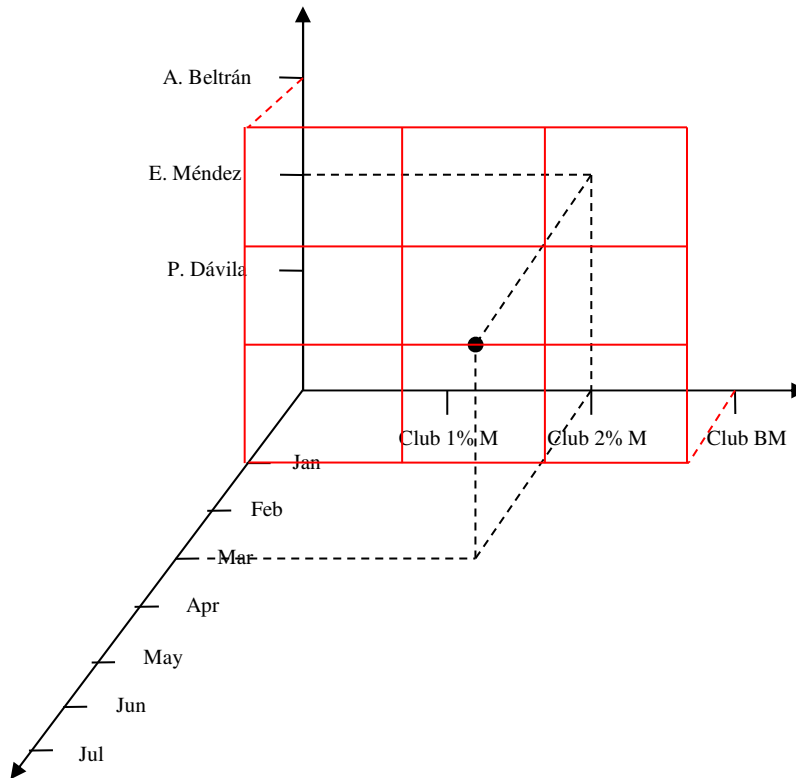


Fig. 6.2. Ejemplo de una *rebanada* de ventas en enero.

- Una **jerarquía** de una dimensión es un tipo de agrupación de sus atributos. Una jerarquía normalmente tiene varios niveles. Ejemplo:

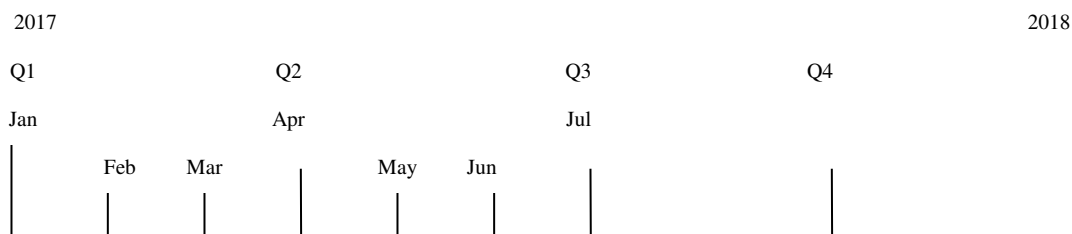


Fig. 6.3. Los atributos relacionados (año, trimestre) son calibrados con respecto al atributo clave (meses).

- Esta jerarquía presenta las siguientes características:
 - Atributo clave de la dimensión: meses.
 - Jerarquía: años, trimestres, meses.
 - Niveles de la jerarquía: 3, que son: años, trimestres, meses.

- Un nivel de la jerarquía es lo mismo que un atributo de la dimensión, aunque también se utiliza el término *atributo de la jerarquía*, como sinónimo de lo anterior.
- Una dimensión puede tener más de una jerarquía, aunque todas usar el mismo atributo clave (por ejemplo, los días o los meses del año).
- Con respecto a la figura anterior, si el atributo clave se cambiara a los días del año, se podrían tener las siguientes jerarquías:
 - Jerarquía 1: años, trimestres, meses, días (cuatro niveles).
 - Jerarquía 2: años, semanas, días (tres niveles).

Más definiciones:

- Cada nivel de una jerarquía define un conjunto de puntos en el espacio multidimensional.
- Los únicos puntos reales (esto es, que existen) son los del **espacio de hechos**.
- Los otros puntos forman un **espacio lógico de datos** y se obtienen sólo por medio de cálculos (por ejemplo, las ventas en un año o en un trimestre).
- El **espacio “completo”** de datos del cubo está formado por el de hechos más el lógico. Cada punto en este espacio se llama **celda**.

La siguiente figura ilustra estos conceptos:

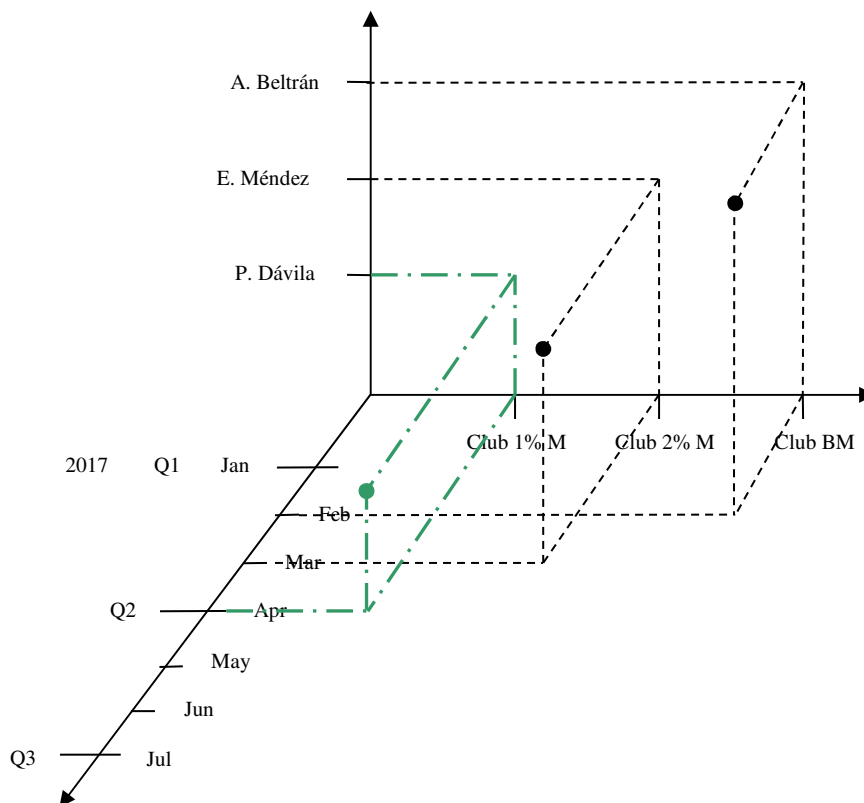


Fig. 6.4. Los atributos relacionados crean nuevos puntos en el espacio multi-dimensional.

Más definiciones:

- Una **medida** es el valor de una celda. Una celda puede tener varias medidas, por ejemplo: el monto de una venta o la cantidad de unidades vendidas de un producto.
- Estas medidas pueden verse como una *dimensión de medidas*, cada medida con tipo de datos, unidad, etc.
- Las **funciones de agregación** son las que calculan los valores de las celdas del espacio lógico de datos, pudiendo ser simples o complejas.

La siguiente figura ilustra lo anterior:

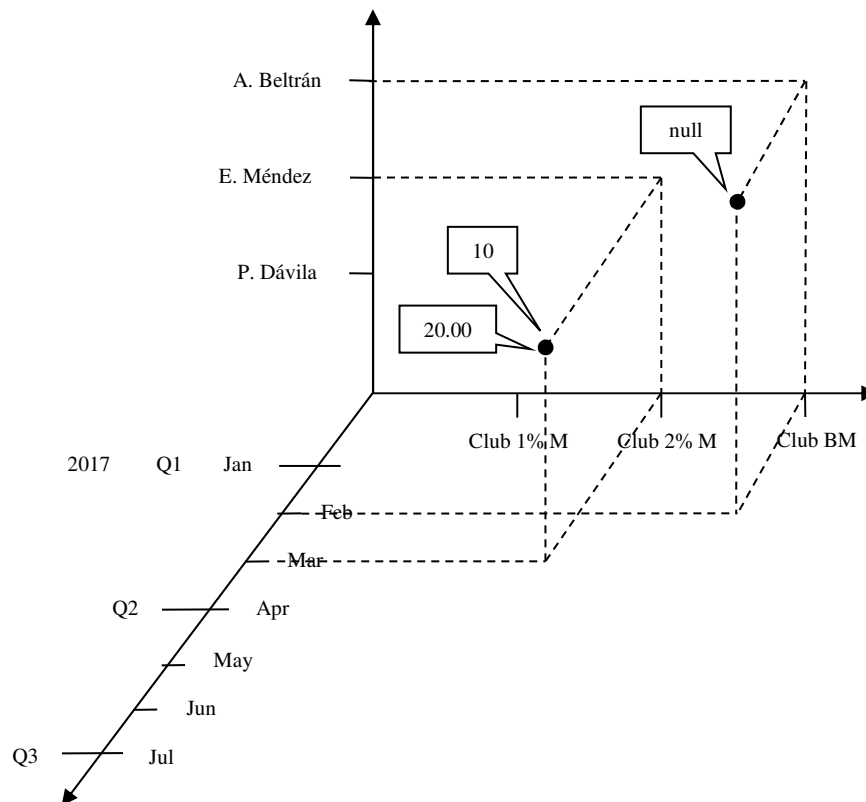


Fig. 6.5. Cada celda de hechos puede tener un valor real o uno potencial.

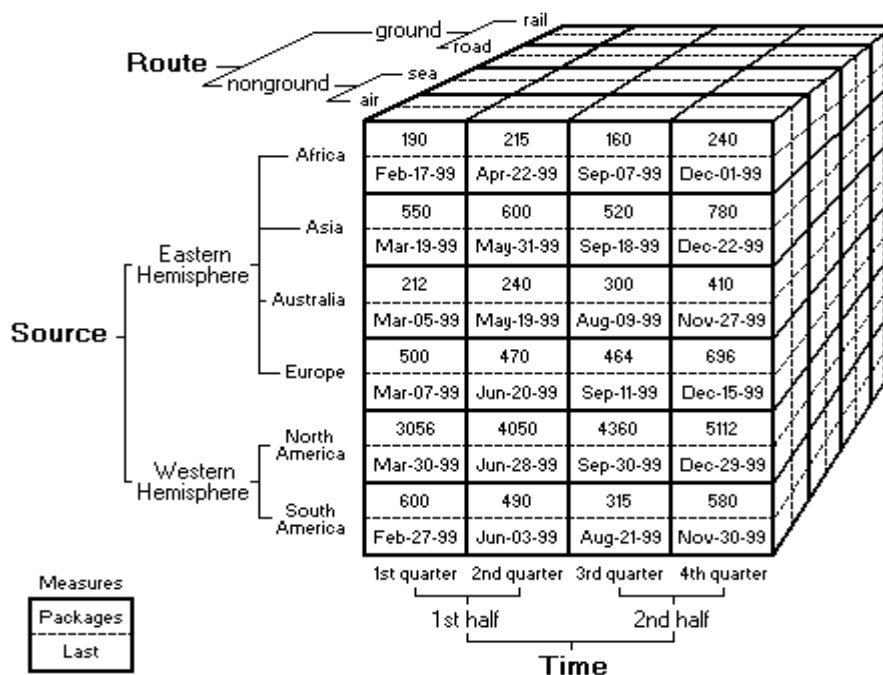
Cada celda de hechos en el cubo está asociada con una venta real o potencial de un producto a un cliente. Si es una venta potencial, la celda tiene un valor **null**; si es real, tiene un valor distinto de **null**.

Ejemplo de funciones de agregación (para calcular puntos del espacio lógico de datos):

- Simples: suma, cuenta, etc.
- Complejas: fórmulas y algoritmos.

6.2.3. Otros ejemplos de cubos

Si se tuviera el siguiente cubo (al cual llamaremos *Paquetes*):



La tupla:

```
(Source.[Eastern Hemisphere].Africa, Time.[2nd half].[4th quarter],
Route.Nonground.Air, Measures.Packages)
```

identificaría a la celda con valor 240.

La tupla:

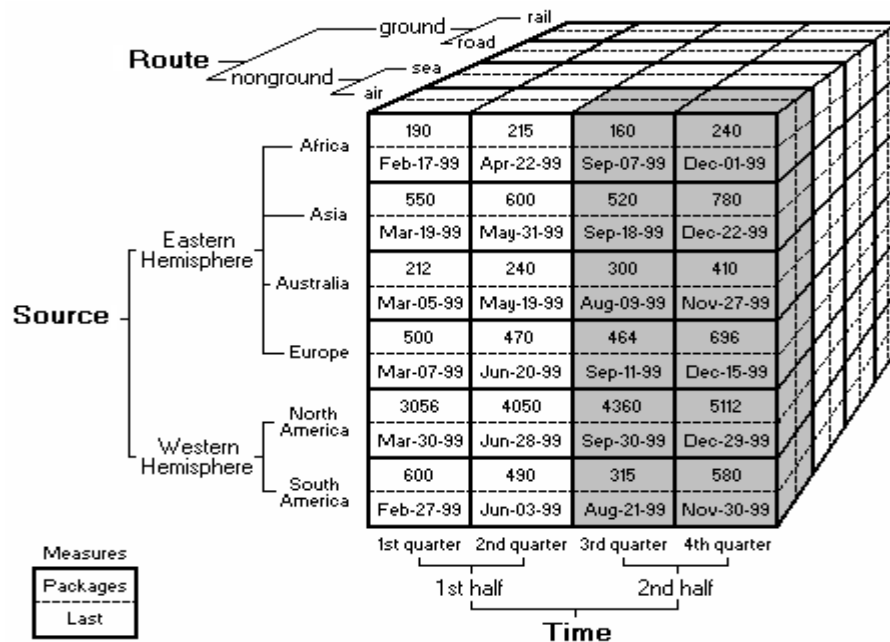
```
(Time.[2nd half].[3rd quarter])
```

identificaría a todos los paquetes y fechas del tercer trimestre, sin importar la ruta ni la fuente.

La tupla:

(Time.[2nd half], Route.nonground.air)

identificaría a la parte sombreada del cubo:



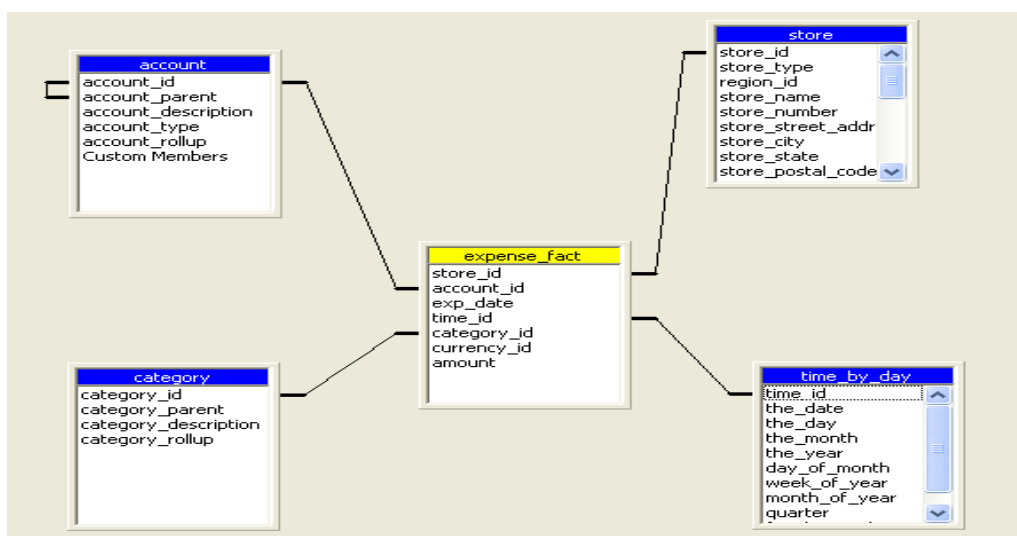
6.2.4. Modelos de diseño para los cubos

Existen dos modelos básicos para el diseño de los cubos:

A) Modelo de Estrella (Star)

Es la estructura básica para un cubo. Se compone de una (o varias) tabla(s) central(es) (llamada(s) tabla(s) de hechos) y de un conjunto de tablas más pequeñas (las tablas de las dimensiones) concentradas alrededor de la(s) tabla(s) de hechos.

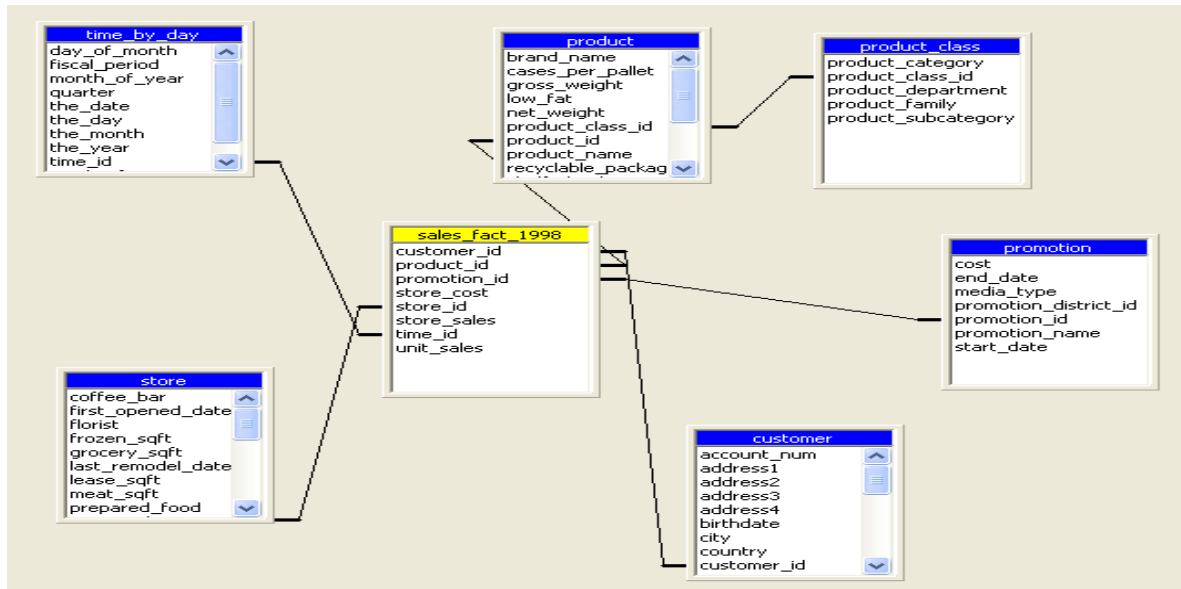
La siguiente figura muestra este diseño:



B) Modelo de Copo de nieve (Snowflake)

Es el resultado de descomponer una jerarquía de una dimensión en una o más tablas. Los vínculos entre estas tablas normalmente serán uno a muchos.

La siguiente figura ilustra el concepto:



6.2.5. Operaciones básicas con los cubos

Son cuatro principalmente:

1. Operación de *Slice*: consiste en mostrar una sección (sub-cubo) del cubo.
2. Operación de *Dice*: consiste en mostrar los datos del cubo desde otra dimensión (también se usa el término “rotar” para esta operación).
3. Operación de *Drill down*: consiste en navegar hacia los niveles inferiores (más detallados) de una jerarquía.
4. Operación de *Drill up*: consiste en navegar hacia los niveles superiores (más agregados) de una jerarquía.

6.3. Creación de cubos con Analysis Services

A continuación se describen, a grandes rasgos, los pasos a seguir para crear un cubo por medio de los asistentes del software de **SQL Server Analysis Services (AS)**. Algunos de los pasos están particularizados para el ejemplo que se desarrollará en clase.

1. Cuestiones importantes que se deben considerar antes de iniciar el trabajo con SQL Server y Analysis Services 2014:

- a. Verificar que SQL Server y SQL Server Analysis Services estén ejecutándose: <Iniciar (con botón derecho)>, elegir Administración de equipos, expandir Servicios y aplicaciones, expandir SQL Server Configuration Manager, clic a SQL Server Services, verificar en el panel central que el SQL Server (MSSQLSERVER) y el SQL Server Analysis Services (MSSQLSERVER) están activados (los íconos de ambos servidores deben tener una flecha verde); si no: colocarse en la entrada SQL Server, <botón derecho>, clic a Start. Repetir lo mismo para el Analysis Services. Cerrar la ventana de la Administración de equipos.
- b. Ambos paquetes trabajan con las bases de datos colocadas en el **directorio default** de SQL Server: C:\Programa Files\Microsoft SQL Server\MSSQL12.MSSQLSERVER\MSSQL\DATA.

2. Definición de la base de datos a usar con AS

- a. Si se va a usar la base de datos que brinda SQL Server (Adventure Works DW) y aún no está en el directorio default:
 - Copiar desde el servidor (carpeta BD/Desarrollos/OLAP) la carpeta awdwdb al **directorio default** de SQL Server.
 - Ejecutar SQL Server. En la pantalla de registro:
 - Tipo de servidor: motor de base de datos
 - Nombre del servidor: localhost o el número de máquina que se usa
 - Autenticación: Autenticación de SQL Server
 - Inicio de sesión: sa
 - Contraseña: sqladmin.
 - Seleccionar en SQL Server la carpeta Bases de datos, abrir el archivo instawdwdb.sql (script de creación de la base de datos), que está en la carpeta awdwdb, y ejecutarlo. Cerrar la ventana del script.
 - Puede ser que al final del script marque un error con una tabla (ProspectiveBuyer), pero como ésta se usa poco, no hay problema. La BD de Adventure Works DW queda guardada en el directorio default y está lista para usarse.
- b. Si se va a usar una base de datos Access (FoodMart, por ejemplo) colocarla en el directorio raíz del proyecto tipo AS.

3. Creación del proyecto tipo Analysis Services

- a. Ejecutar Visual Studio 2015.
- b. **Definir la carpeta** donde se guardará la solución (menú Herramientas, Opciones..., Proyectos y soluciones, caja de texto Ubicación de proyectos, seleccionar la carpeta donde quedarán los proyectos, p. ej., C:\BD\Programas).
- c. Crear un proyecto nuevo de tipo Business Intelligence, entrada Analysis Services. Elegir la plantilla Proyecto multidimensional y de minería de datos, dar el nombre del proyecto, quitar la marca a la opción Crear directorio para la solución, <Aceptar>.

4. Definir una nueva fuente de datos

- a. En el Explorador de soluciones, abrir el menú breve en la carpeta Orígenes de datos. Elegir Nuevo origen de datos...
- b. Seguir los pasos indicados por el asistente. Observar las siguientes cuestiones:
 - En la ventana Seleccionar cómo definir la conexión, dar clic a <Nueva...>.
 - En la ventana Administrador de conexiones:
 - En Proveedor, elegir:
 - Para SQL Server 2014: OLE DB nativo\SQL Native Client 11.0.
 - Para Access 2016: Microsoft Office 16.0 Access Database...
 - En Nombre del servidor, elegir: el nombre de la máquina local (por ejemplo, CC201-08) o localhost.
 - Marcar la opción de autenticación que corresponda: Windows o SQL Server; de preferencia usar la de SQL Server, con usuario: **sa**, y contraseña: **sqladmin** (marcar Guardar mi contraseña).
 - Si la base de datos a usar (por ejemplo, *AdventureWorksDW*) ya está registrada, elegirla en el combo Seleccionar o escribir el nombre de la base de datos; si no, marcar la opción Adjuntar un archivo de base de datos, dar clic a <Examinar...> y seleccionar la base de datos a usar en el directorio que corresponda.
 - Para el ejemplo, elegir la base de datos: *AdventureWorksDW* (la cual ya debe estar colocada en el **directorio default**).
 - Si se requiere, dar un nombre lógico a la base de datos (por ejemplo, *Adventure_BD*).
 - Dar clic en <Probar conexión>. Si no hay error, dar clic en <Aceptar> y <Siguiente>.
 - En la ventana Información de suplantación, marcar la opción Utilizar la cuenta de servicio (es la cuenta de servicio de AS con los permisos necesarios para acceder a la BD *AdventureWorksDW*), dar clic a <Siguiente>.
 - En la ventana Finalización del asistente, dar el Nombre del origen de datos o dejar el default, <Finalizar>.

5. Definir una nueva vista de la fuente de datos

- a. En el Explorador de soluciones, abrir el menú breve en la carpeta Vistas del origen de datos. Elegir Nueva vista del origen de datos...
- b. Seguir los pasos indicados por el asistente. Observar las siguientes cuestiones:
 - En la ventana Seleccionar un origen de datos, elegir el origen creado en el punto anterior, <Siguiente>.
 - En la ventana Seleccionar tablas y vistas, elegir las tablas que se usarán para construir la vista de datos desde la lista de Objetos disponibles. Esta vista se empleará para construir el cubo. Para el ejemplo elegir:
 - *DimCustomer*, *DimGeography*, *DimProduct*, *DimTime* y *FactInternetSales*.
 - <Siguiente>, dar nombre a la vista y <Finalizar>.
- c. Aparece una ventana con varias secciones de la vista.
- d. En la sección del diagrama de la vista cambiar el nombre de cada tabla (a uno más legible). Para esto, elegir a la tabla y en su propiedad FriendlyName dar el nuevo nombre; por ejemplo, en lugar de *FactInternetSales*, dar *InternetSales*; en lugar de *DimTime*, dar *Time*, etc.
- e. Cerrar la vista y guardar lo realizado.

6. Definición del cubo de datos

- a. En el Explorador de soluciones, abrir el menú breve de Cubos. Elegir Nuevo cubo...
- b. Seguir los pasos indicados por el asistente. Observar las siguientes cuestiones:
 - En la ventana Seleccionar método de creación marcar la opción: Usar tablas existentes.
 - En la ventana Seleccionar tablas de grupo de medida: elegir la vista a utilizar y marcar las tablas de hechos.
 - Por default, AS selecciona como medidas, en las tablas de hechos, a cada columna de tipo numérico que no esté ligada a una tabla de dimensión. Para el ejemplo, en la página Seleccionar medidas, quitar las marcas a: *Promotion key*, *Currency key* y *Revision Number*.
 - En la ventana Seleccionar nuevas dimensiones, marcar a las tablas de las dimensiones (normalmente las que no son de hechos).
 - En la ventana Finalización del asistente dar nombre al cubo y <Finalizar>.
 - Se genera la definición del cubo y de sus dimensiones.

7. Inspección de las propiedades del cubo y de las dimensiones

- a. En este punto, se puede hacer una inspección de algunos elementos importantes del cubo (abrirlo y elegir la pestaña Estructura de cubo), como son su estructura y sus dimensiones. En el primer caso se pueden ver además sus medidas. En el segundo, en el panel de Dimensiones, se pueden ver algunas características de las mismas.
- b. Al elegir una dimensión y con menú breve, Editar dimensión, se pueden ver sus atributos, jerarquías (cuando ya existan) y una vista parcial de las mismas.
En este caso, y estando en la pestaña Estructura de dimensión, hay que arrastrar los atributos que se quieran visualizar posteriormente en el cubo, desde el panel Vista del origen de datos al panel Atributos.
Para el ejemplo, arrastrar los siguientes atributos:
 - *Customer: nombres* (todos), *City*, *State Province Name* y *Spanish Country Region Name*.
 - *Product: Product Subcategory Key*, *Spanish Product Name*, *Product Line*, *Model Name*.
 - *Order Date: Spanish Month Name*, *Calendar Quarter*, *Calendar Year*, *Calendar Semester*.
- c. En la pestaña Uso de dimensiones, en Estructura del cubo, se puede ver un poco más de detalle de las dimensiones. Seleccionando el asistente de alguna de ellas, se abre una ventana que contiene más información acerca de las mismas.

8. Ejecución del proyecto en la instancia local del AS

- a. En el Explorador de soluciones, abrir el menú breve del proyecto y elegir Proceso... (si aparece una ventana pidiendo generar e implementar el proyecto, dar <Sí>). El software construye y ejecuta el proyecto en la instancia especificada del AS (en este caso, en localhost o en la máquina elegida). El avance del proceso se muestra en el panel Progreso de la implementación, y en la ventana Procesar base de datos. En el panel debe verse el mensaje: La implementación finalizó correctamente; en la ventana, se debe dar clic a <Ejecutar...>.
- b. En el segundo caso, se abre la ventana Progreso del proceso. Si no hay errores, aparece el mensaje Proceso finalizado correctamente, en cuyo caso se pueden abrir los diferentes nodos que aparecen en la ventana, para ver las diversas características con las que se construyó el proyecto. Dar clic a <Cerrar>, <Cerrar>.

9. Visualización del cubo construido

- Se pueden ver, jerárquicamente, los valores de las diferentes dimensiones que tiene el cubo (doble clic a la dimensión en el Explorador de soluciones, se abre la dimensión, pestaña Explorador).
- Al seleccionar la pestaña Explorador, del diseñador del cubo, aparece una tabla dinámica, similar a las de Excel, con la cual se pueden probar las diversas alternativas de visualización del cubo.

10. Modificación de medidas, atributos y jerarquías

Una vez construido el cubo se pueden hacer varias modificaciones sobre el mismo para mejorar su presentación, modificar las jerarquías de las dimensiones, hacer ciertos cálculos, etc. A continuación se describen los principales cambios que se pueden hacer:

- Presentación de las medidas:** en la pestaña Estructura del cubo, del diseñador del cubo, se pueden seleccionar las medidas y trabajar con sus propiedades para mejorar la presentación de los datos; por ejemplo, se pueden usar las **propiedades FormatString y Name** para dar formato a los valores y para dar un título a una medida.

Nota: Para poder ver los cambios, en general, es necesario volver a procesar al cubo (con el ícono respectivo en la barra de herramientas del diseñador correspondiente) para que se reconozcan, y reconectarlo o actualizarlo (con el ícono respectivo en la barra de herramientas del diseñador correspondiente) para visualizarlos.

- Modificación de dimensiones:** AS permite hacer diversos cambios a las dimensiones de un cubo. Las principales son las siguientes:
 - Modificación de propiedades de elementos:** en la sección de Atributos se puede cambiar el nombre de los elementos (para el ejemplo, en *Customer: Spanish Country Region Name* a *Country-Region*, *State Province Name* a *State-Province*).
 - Creación de una jerarquía:** arrastrar un elemento de la sección de Atributos a la sección de Jerarquías colocándolo en el lugar adecuado (para el ejemplo, en *Customer*: arrastrar *Country-Region*, *State-Province*, *City* y *CustomerKey*, en este orden). Darle el nombre: *Country-State-City*.
 - Agregar un cálculo con nombre a una tabla:** consiste en agregar un cálculo a una tabla de tal manera que se comporte como si fuera una nueva columna de la misma (se ve como una pseudo-columna). **Proceso:** abrir la vista definida en el punto 5, elegir la tabla (*Customer*, en este caso), menú breve, Nuevo cálculo con nombre. En el cuadro de dialogo dar el nombre de la pseudo-columna (*FullName*, en este caso) y la expresión que representa al cálculo (al final, la pseudo-columna aparece en los campos de la tabla):

```
CASE
  WHEN MiddleName IS NULL THEN
    FirstName + ' ' + LastName
  ELSE
    FirstName + ' ' + MiddleName + '. ' + LastName
END
```

- iv. Uso del cálculo con nombre para nombres de elementos:** el cálculo con nombre puede usarse como una propiedad de un atributo de la dimensión. **Proceso:** Explorador de soluciones, editar la dimensión *Customer*, pestaña Estructura de dimensión, seleccionar el atributo *Customer Key* en la sección Atributos, mostrar sus propiedades, cambiarle el nombre (en la propiedad *Name* dar “*Full Name*”), abrir el asistente de su propiedad *NameColumn*, elegir *FullName* en Columna de origen, <Aceptar>, cambiar (a *Full Name*) el nombre de *Customer Key* en la sección de jerarquías.

11. Agregación de tablas de hechos y de dimensiones a un cubo ya existente

Una vez que se tiene un cubo se pueden agregar más tablas al mismo, tanto de hechos como de dimensiones. A continuación se describen los principales pasos a seguir.

- a. **Modificación de la vista.** Abrir la vista a usar, en la ventana central abrir el menú breve y elegir Agregar o quitar tablas..., seleccionar (en Objetos disponibles) las tablas a agregar y pasarlas a Objetos incluidos (*FactResellerSales* y *DimEmployee*, para el ejemplo), oprimir <Aceptar>, renombrar a las tablas (a *ResellerSales* y *Employee*, respectivamente), cerrar la ventana y guardar los cambios.
- b. **Creación de las nuevas dimensiones.** Indicar qué tablas, de las agregadas, se usarán como dimensiones, para lo cual: abrir el menú Proyecto y elegir Nueva dimensión...; seguir los pasos indicados por el asistente en una manera similar a como se construyó el cubo, tal como está descrito en el punto 6 anterior. Aspectos a considerar:
 - En la ventana Seleccionar método de creación, dejar la marca en Usar una tabla existente.
 - En la ventana Especificar información de origen, especificar la tabla y la clave de la nueva dimensión (*Employee* y *EmployeeKey*, para el ejemplo).
 - En la ventana Seleccionar las tablas relacionadas, marcar las que correspondan (ninguna para *Employee*).
 - En la ventana Seleccionar los atributos de la dimensión, marcar, adicionalmente, los que correspondan (para el ejemplo: *First Name*, *Last Name*, *Middle Name*, *Title*, *Department Name*).
- c. **Creación de jerarquías en la dimensión.** Repetir el proceso descrito en el punto 10.b.ii (para el ejemplo, crear la jerarquía: *Dept-Title: Department Name, Title, Employee Key*).
- d. **Registrar las nuevas dimensiones.** En la misma pestaña Estructura del cubo, en la sección Dimensiones, elegir el ícono del cubo, menú breve, Agregar dimensión de cubo..., elegir las tablas de las nuevas dimensiones, <Aceptar>.
- e. **Registrar las nuevas tablas de hechos.** En AS las nuevas tablas de hechos se conocen como **Nuevos grupos de medidas** y hay que registrarlas dentro del cubo: abrir el cubo en su pestaña Estructura del cubo, sección Medidas, menú breve, Nuevo grupo de medidas..., en la ventana abierta elegir las nuevas tablas de hechos, <Aceptar>. **Eliminar** atributos innecesarios de *Reseller Sales* (en la sección de medidas): *Reseller Key*, *Sales Territory Key*, *Promotion Key*, *Currency Key* y *Revision Number*.
- f. **Visualizar los nuevos hechos y dimensiones.** Procesar y reconectar. **En el Explorador del cubo, ya deben poder visualizarse los nuevos datos** en la tabla dinámica.
- g. **Modificación de la visualización de una dimensión con vínculo recursivo (Parent-Child).** Abrir la dimensión, elegir el atributo “padre” (*Parent Employee Key*, para el ejemplo), dar clic en el asistente de la propiedad NamingTemplate; en la ventana que se abre, teclear *Employee Level ** en la columna Nombre de la segunda fila (la del *), <Aceptar>, elegir el valor *NonLeafDataHidden* en la propiedad MembersWithData del atributo “padre”, procesar.
- h. Una vez que ya se agregaron las nuevas tablas de hechos y de dimensiones al cubo, se pueden volver a hacer las actividades descritas en los puntos 9 y 10 anteriores.
- i. **Guardar todo.**

12. Definición de cálculos

Se pueden definir cálculos dentro de un cubo. Para esto es necesario hacer las definiciones dentro del proyecto tipo Analysis Services para que queden como parte del cubo en cuestión.

Nota: hay que posicionarse siempre en el último elemento de Cálculos para que el nuevo elemento creado se coloque al final de todos los elementos calculados.

- a. **Miembros calculados.** Abrir el diseñador del cubo, elegir la pestaña Cálculos, en su barra de herramientas: clic al botón Vista de formulario, clic al botón Nuevo miembro calculado. En la ventana que aparece especificar lo siguiente:
 - Dar el nombre del nuevo elemento (para el ejemplo: [Total Sales Amount]).
 - En el combo de las dimensiones se debe elegir aquella que se usará para el nuevo elemento (para el ejemplo, Measures).
 - En el cuadro Expresión, dar la expresión que define al nuevo elemento. Para el ejemplo:
[Measures].[Sales Amount] + [Measures].[Sales Amount - Reseller Sales]
 - Elegir el formato ("\$,##0.00...", para el ejemplo). En el combo Comportamiento, si no está vacío, se marcan los mismos elementos de la expresión anterior.
 - Dar clic en el ícono Vista de script (barra de herramientas del cubo). Agregar comentarios para el nuevo miembro calculado.

Una vez que se hayan definido todos los nuevos elementos, hay que volver a procesar el cubo para poder visualizarlos.

13. Acceso al cubo desde SQL Server

Se puede visualizar el cubo desde SQL Server para lo cual se debe hacer lo siguiente, después de ejecutar SQL Server Management Studio (situarse en él, menú breve, Ejecutar como administrador): elegir Analysis Services en el combo Tipo de servidor, elegir el servidor en su combo (p. ej., CC201-10 o localhost), <Conectar>. En la carpeta Bases de datos debe aparecer el proyecto con el cubo formado (si no, menú breve, Actualizar), en la carpeta Cubos se puede elegir el cubo para procesarlo/examinarlo (menú breve, Procesar/Examinar). Se abre una tabla dinámica y ya se puede visualizar el cubo.

14. Acceso al cubo desde Excel

Una vez que el cubo ya está creado también se puede visualizar desde Excel. Esto es particularmente útil para el caso de graficación de los datos, ya que Excel posee un subsistema, simple, que permite graficar de manera fácil. El procedimiento a seguir es el siguiente:

- a. Situarse en la pestaña Explorador del cubo (en Visual Studio), en la barra de herramientas dar clic en el ícono de Excel (si aparece una ventana indicando un posible problema de seguridad, dar clic en <Habilitar>).
- b. Al abrirse Excel, se puede crear una tabla dinámica en la cual ya se pueden mostrar los datos, con sus diferentes variantes, tal como en el explorador de cubos de Analysis Services. Para esto, en el panel derecho de la hoja de Excel hay cuatro áreas donde se deben ir colocando los diferentes campos del cubo (arrastrándolos) que quieren visualizarse en la tabla dinámica, tanto para medidas como para elementos o jerarquías de las dimensiones.
- c. En la barra de herramientas, grupo Herramientas, botón Gráfico dinámico, se puede elegir al asistente de gráficas para graficar los datos anteriores.

6.4. Lenguaje MDX (MultiDimensional eXpressions)

Es un lenguaje usado para definir y manipular cubos por medio de Analysis Services. Tiene cierto parecido con SQL, aunque tiene una sintaxis y elementos propios para trabajar con los cubos. MDX posee las siguientes características:

- Instrucciones de definición de datos para crear, eliminar y trabajar con objetos multidimensionales.
- Instrucciones de manipulación de datos para recuperar datos desde los objetos multidimensionales.
- Un conjunto robusto de operadores y funciones para la manipulación de datos de los objetos multidimensionales.
- Posibilidad de elaborar scripts (guiones).

MDX trabaja con los conceptos propios de los cubos, por lo que también maneja los términos de: **dimensiones, jerarquías, niveles, elementos (miembros) y medidas**. También maneja el concepto de **celda** (la intersección de elementos específicos de las dimensiones de un cubo) y de **tupla** (el concepto usado para identificar a una celda, a una sección o a un subcubo del cubo).

Se introduce un nuevo término: **conjunto**, como una colección ordenada de cero, una o más tuplas. El siguiente ejemplo (con respecto al cubo de Paquetes de la sección 6.2.3):

```
{ (Time.[1st half], Route.nonground.air), (Time.[2nd half],  
Route.nonground.sea) }
```

muestra un conjunto de dos tuplas.

Elementos básicos

Expresiones. Son combinaciones de identificadores, valores y operadores. Entre los elementos que se pueden usar para una expresión están: constantes (las cadenas se delimitan con apóstrofo), funciones escalares (regresan un solo valor), identificadores, variables y cualquier combinación de éstos unidos por operadores.

Los operadores permitidos son:

- Aritméticos: +, /, *, -, ^ (potencia)
- Lógicos: and, or, not, xor
- De comparación: =, <>, >, >=, <, <=
- De concatenación (de cadenas): +
- De conjuntos: - (diferencia), * (producto cruz), : (rango- devuelve todos los elementos en el rango), + (unión)

Ejemplos:

```
with  
    member Measures.[Special Discount] as  
        Measures.[Discount Amount] * 1.5  
select  
    Measures.[Special Discount] on columns,  
    non empty Product.[Product Line].members on rows  
from [Clientes-Productos]
```



```

with
  member Measures.[Total Sales Amount] as
    Measures.[Sales Amount] +
    Measures.[Sales Amount - Reseller Sales]
select
  Measures.[Total Sales Amount] on columns
from [Clientes-Productos]

```

Funciones. MDX proporciona una gran cantidad de funciones para aplicar sobre los objetos multidimensionales. Existen de varios tipos, agrupadas en categorías: de dimensiones, de jerarquías, lógicas, de conjuntos, financieras, etc. Se remite a la ayuda tanto de SQL Server (consultas MDX), como a la de Analysis Services (pestaña **Cálculos** del diseñador del cubo en Visual Studio) para ver la lista completa de estas funciones.

Consultas básicas en MDX

La consulta básica en MDX tiene la siguiente sintaxis:

```

SELECT [<axis_specification>
      [, <axis_specification>...]]
FROM  [<cube_specification>]
[WHERE [< slicer_specification>]]

```

Esta frase *select* es usada para especificar un conjunto de datos que es un subconjunto del cubo. La consulta debe contener la siguiente información:

- Los elementos de cada dimensión a ser incluidos en cada eje de la consulta.
- El nombre del cubo.
- Los elementos de una dimensión segmentadora (*slicer*) sobre la cual los datos son segmentados para los miembros de las dimensiones eje (opcional). De cierta manera funciona como una restricción sobre las dimensiones eje.

Dimensión eje (axis dimension): se le da este nombre a las dimensiones que pueden ser especificadas en la cláusula *select*.

Dimensión segmentadora (slicer dimension): se le da este nombre a la dimensión que puede ser especificada en la cláusula *where*.

Ejemplo 1:

```

-- 1. Muestra las ventas (monto y unidades) de los años
--      antepasado y pasado de la provincia de Alberta, Canada.
select
  { Measures.[Sales Amount],
    Measures.[Order Quantity] } on columns,
  { [Order Date].[2016], [Order Date].[2017] } on rows
from [Clientes-Productos]
where Customer.Canada.Alberta

```

Ejemplo 2:

```

-- 2. Muestra los montos de las ventas de cada modelo de
--      los productos.
select
  Measures.[Sales Amount] on columns,
  non empty Product.[Model Name].members on rows
from [Clientes-Productos]

```

Ejemplo 3:

```
-- 3. Cantidades vendidas de las líneas de productos en el año
--    actual. La medida se especifica en el where.
select
    Product.[Product Line].members on columns,
    [Order Date].[2018] on rows
from [Clientes-Productos]
where Measures.[Order Quantity]
```

Ejemplo 4:

```
-- 4. Muestra las ventas (montos y unidades) del año actual en la
--    ciudad de Ballard, de dos modelos de productos.
select
    { Measures.[Sales Amount],
      Measures.[Order Quantity] } on columns,
    { ([HL Road Tire],[2018]),
      ([LL Mountain Tire],[2018]) } on rows
from [Clientes-Productos]
where City.Ballard
```

Ejemplo 5:

```
-- 5.a. Una consulta puede tener sólo la dimensión de columnas.
select [Country-Region].members on columns
from [Clientes-Productos]

-- 5.b. Ejemplo de los hijos de un elemento de una dimensión.
--      No es necesario dar todos los niveles de la jerarquía.
select
    non empty [Sales Amount] on columns,
    [Product Model Lines].[Touring-1000].children on rows
from [Clientes-Productos]

-- 5.c. Un ejemplo similar, pero usando la función Descendants.
--      También se pueden "saltar" niveles de la jerarquía.
select non empty [Sales Amount] on columns,
       descendants([Product Model Lines].[Touring-1000]) on rows
from [Clientes-Productos]
```

Ejemplo 6:

```
-- 6. Ejemplo de la función Filter. Sirve para aplicar una
--    condición.
select [Sales Amount] on columns,
       filter(Product.[Model Name].members, [Sales Amount]>100000) on rows
from [Clientes-Productos]
```

Observaciones y restricciones del select

En esta sección se comentan algunas cuestiones que deben considerarse al usar MDX.

1. Antes de dar cualquier instrucción, hay que elegir el cubo a usar, ya sea en el combo Bases de datos disponibles, que se encuentra en la barra de herramientas del ambiente de MDX, o directamente en el Explorador de objetos, en la carpeta Cubos del proyecto en cuestión. Después dar clic al botón Nueva consulta (en la barra de herramientas).
2. Las dimensiones en una tupla siempre deben ser diferentes; los tipos de las tuplas en conjunto siempre deben ser iguales.

3. Un elemento de una dimensión puede darse directamente, sin antecederlo con su jerarquía, siempre y cuando su nombre sea único en todas las dimensiones; de forma similar, también puede darse el nombre de la dimensión y el elemento (ésta sería la forma recomendada). En cambio, cuando se da una jerarquía completa, normalmente no puede omitirse nivel alguno.
4. La frase **non empty** sirve para eliminar tuplas vacías del resultado de una consulta.
5. Cuando se escribe el nombre concreto de un elemento de una dimensión (por ejemplo, `Customer.France`), no se puede usar `Members`, aunque sí se puede usar `Children`. Cuando se quiere usar `Members`, entonces hay que usar los nombres de los niveles de la jerarquía (por ejemplo, `Customer.[Country-Region].members`).
6. Cuando se usa la dimensión del tiempo en una consulta, **siempre** hay que encerrar entre `[]` a cada elemento que se especifica en la jerarquía. Con cualquier otra dimensión no es necesario hacerlo, salvo que el nombre de uno de sus elementos usados contenga espacios o símbolos que podrían confundirse con operadores (como el guión), en cuyo caso también hay que encerrarlo entre `[]`.

Algunas instrucciones adicionales de MDX

Case. Sirve para definir expresiones en base al cumplimiento de condiciones. Su sintaxis es:

```
CASE [Numeric_expression]
WHEN {Numeric_expression | Boolean_expression} THEN
    Result_expression
...
[ELSE Result_expression]
END
```

Ejemplo:

```
WITH MEMBER Measures.StringSales AS
CASE
    WHEN Measures.[Sales Amount] >= 50000 THEN "50,000+"
    WHEN Measures.[Sales Amount] >= 30000 THEN "30,000-50,000"
    WHEN Measures.[Sales Amount] >= 10000 THEN "10,000-30,000"
    WHEN Measures.[Sales Amount] < 10000 THEN "< 10,000"
END
SELECT [Order Date].[Calendar Year] on 0,
    Customer.[State-Province].Members on 1
FROM [Clientes-Productos]
WHERE Measures.StringSales
```

Create member. Sirve para definir un nuevo elemento dentro de un cubo para usarlo posteriormente dentro de otras expresiones MDX. El elemento se puede definir para usarlo localmente dentro de una consulta, como en el ejemplo anterior, o puede definirse para usarse dentro de toda una sesión de trabajo como en el siguiente ejemplo:

```
create member [Clientes-Productos].[Ganancia Internet]
as [Sales Amount] - [Total Product Cost]
```

Una vez definido el elemento puede usarse en cualquier lugar como en el siguiente ejemplo:

```
select [Ganancia Internet] on 0,
    Customer.[Country-Region].members on 1
from [Clientes-Productos]
```

Create set. También pueden definirse nombres para etiquetar a conjuntos que son usados comúnmente; la definición puede ser local a un **select** o global para toda una sesión, como en el siguiente ejemplo:

```
create session set [Clientes-Productos].[Ventas Internet] as
{ [Sales Amount], [Order Quantity] }
```

Después este nombre puede usarse tal como se muestra a continuación:

```
select [Ventas Internet] on 0,
       Customer.[Country-Region].members on 1
from [Clientes-Productos]
```