

EXPRESIONES MULTIDIMENSIONALES - MDX

OPTATIVA II: DATA WAREHOUSE

ANDRES SANCHEZ

LUIS FELIPE VELASCO TAO

JUAN DAVID GONZALEZ


UNIVERSIDAD DE SAN BUENAVENTURA

FACULTAD DE INGENIERIA

BOGOTÁ


22 DE SEPTIEMBRE

2021

 UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ	Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse	MDX – Expresiones multidimensionales	2021 - 2
---	---	--	----------

Contenido

MDX – Expresiones Multidimensionales	3
Definición	3
Historia (d)	3
Tipos de datos	3
Comparación entre SQL y MDX	5
Ejemplos de consultas	7
Herramientas	13
ETL.....	14
Extracción (E)	14
Transformación (T)	15
Carga (L).....	15
Herramientas ETL.....	16
Bibliografía	18
Tabla de ilustraciones	19

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>MDX – Expresiones multidimensionales</p>	<p>2021 - 2</p>
---	---	---	-----------------

MDX – EXPRESIONES MULTIDIMENSIONALES

Al momento de abordar el tema de cómo podemos manipular los datos en un modelo de datos MOLAP se debe obligatoriamente pensar en el concepto de expresiones multidimensionales, su historia, características y conocer como estas expresiones operan en las herramientas existentes de motores para cibos OLAP.

Definición

Es un lenguaje de consultas de datos multidimensionales que se encuentran almacenados en cubos OLAP. Utiliza un modelo de datos multidimensional para habilitar la navegación en múltiples dimensiones, niveles, y arriba y debajo de una jerarquía. Con MDX se puede acceder a agregados pre computados en posiciones específicas en una jerarquía. MDX es un estándar abierto (Microsoft, 2018).

MDX es implícitamente un paradigma basado en jerarquías. Todos los miembros de todas las dimensiones deben pertenecer a una jerarquía. MDX está basado en la especificación XML para análisis (XMLA), con extensiones específicas para los Servicios de Análisis de SQL Server. MDX utiliza expresiones compuestas de identificadores, valores, declaraciones, funciones y operadores que los servicios de análisis pueden evaluar para recuperar un objeto (como un conjunto o un miembro), o un valor escalar (como una cadena o un número) (Koo, 2014; Microsoft, 2018)


Historia (d)

MDX fue introducido como parte de la especificación OLE DB for OLAP (Vinculación de objetos e incrustación de bases de datos para Procesamiento Analítico en Línea) en 1997 desde Microsoft. La especificación fue seguida por el lanzamiento comercial de Microsoft OLAP Services 7.0 en 1998 y posteriormente por Microsoft Analysis Services. La última versión de OLE DB for OLAP fue emitida por Microsoft en 1999. La especificación XML para Análisis tomo como referencia al OLE DB for OLAP sobre los detalles del lenguaje de consultas MDX. A partir de la versión de Microsoft Excel 2007, se empezaron a añadir extensiones para el uso de este lenguaje (Microsoft, 2017; String Fixer, 2008).


Tipos de datos

A continuación, se expondrán los tipos de datos que se usan en las consultas MDX, esto en base a la información expuesta por (Microsoft, 2019; String Fixer, 2008):

- **Escalar:** dentro de los sistemas MOLAP, se considera un escalar al tipo de dato atómico que se manipulará, en resumidas cuentas, es un número o cadena de caracteres que podrá ser manipulada por funciones como *Aggregate*, *UniqueName*, *Value*, entre otras que permitirán realizar las operaciones necesarias para el análisis de los datos almacenados.

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>MDX – Expresiones multidimensionales</p>	<p>2021 - 2</p>
---	---	---	-----------------

- **Dimensión:** será la forma en que se estructurarán los datos de un cubo del DW, en el cual se podrán observar elementos como medidas ordenadas por medio de jerarquías definidas por niveles. Estas dimensiones pueden ser invocadas por medio de su nombre, por ejemplo, *[Date]*, o por medio de la instrucción. *Dimensión*.
- **Jerarquía:** será la forma por medio de la cual se ordenarán los miembros de cada una de las dimensiones, de modo tal se presente una estructura que permita el análisis de los datos. Para acceder a la jerarquía de una dimensión se debe especificar su nombre, para luego, entre llaves cuadradas definir hasta que miembro de dicha jerarquía se quiere acceder, un ejemplo basado en la dimensión de fecha sería el que permita llegar al mes de agosto de un año, para esto deberías especificar de la siguiente forma la instrucción *[Date].[Year].[Month].[Day]* de modo tal, los valores de año, mes y día se reemplacen acorde a la información requerida, de la siguiente forma *[Date].[2016].[Agosto].[16]*
- **Nivel:** un nivel será la forma en la cual se le dará valor a los miembros de una jerarquía, por medio de la cual se definiría la forma en que se compone una dimensión, en base a la dimensión de fecha se puede dejar claramente definido el concepto de niveles, de modo tal el nivel más alto de la jerarquía de la dimensión de fecha podría ser Año, el cual se puede dividir en semestres, meses, semanas, días, días de la semana, horas, minutos y hasta segundos. Los niveles que componen a una jerarquía dependen de cuan específica se requiere la información que se ordenara en una dimensión.
- **Miembro:** cada jerarquía se compone de elementos nombrados miembros, los cuales son un dato en específico por medio del cual se les dará cierto orden a los datos dentro de una dimensión, basándose en la dimensión de fecha y en las definiciones de jerarquía y nivel, se puede definir que, un miembro de la dimensión *[Date]* es *[Year]*, siendo esta la de más alto nivel en la jerarquía de esta dimensión.
- **Tupla:** al definir los elementos que compone una dimensión y la forma en que estos se encuentran organizados, se puede definir el concepto de tupla como una colección ordenada de los miembros de una o varias dimensiones empleadas en una consulta, como ejemplo de este concepto se puede obtener el siguiente de una consulta compuesta por las dimensiones de fecha y lugar (*[Date].[2016]. , [Place].[Colombia].[Cundinamarca].[Bogota]*)
- **Set:** este concepto se define como la colección de varias tuplas en base a uno o varios miembros de una dimensión, de modo tal, con base al ejemplo dado en las tuplas, se puede dar como ejemplo un set de una consulta compuesta por las tuplas de la fecha y lugar, enfocados en los años 2016 y

 UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ	Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse	MDX Expresiones multidimensionales	– 2021 - 2
--	---	--	-------------------

2017 {([Date].[2016]. , [Place].[Colombia].[Cundinamarca].[Bogota]),
([Date].[2017]. , [Place].[Colombia].[Cundinamarca].[Bogota])}.

Reseller Sales Amount		Column Labels				Measure
Row Labels		CY 2005	CY 2006	CY 2007	CY 2008	
Europe			\$1,698,880.94	\$5,632,816.55	\$3,538,837.31	\$10,870,534.80
North America		\$8,065,435.31	\$22,445,548.71	\$25,722,421.91	\$11,752,320.88	\$67,985,726.81
Canada		\$1,513,359.46	\$4,822,999.20	\$5,651,305.43	\$2,390,261.51	\$14,377,925.60
United States		\$6,552,075.85	\$17,622,549.51	\$20,071,116.48	\$9,362,059.37	\$53,607,801.21
Central		\$951,240.65	\$2,625,639.72	\$3,005,591.43	\$1,323,536.38	\$7,906,008.18
Northeast		\$568,545.52	\$2,443,901.73	\$2,863,937.85	\$1,056,456.93	\$6,932,842.01
Northwest		\$1,689,790.14	\$3,471,099.54	\$4,640,535.06	\$2,633,651.25	\$12,435,076.00
Southeast		\$1,448,921.51	\$2,815,903.10	\$2,429,279.90	\$1,173,311.72	\$7,867,416.23
Southwest		\$1,893,578.02	\$6,266,005.43	\$7,131,772.25	\$3,175,103.09	\$18,466,458.79
Pacific				\$847,430.96	\$746,904.41	\$1,594,335.38
Grand Total		\$8,065,435.31	\$24,144,429.65	\$32,202,669.43	\$16,038,062.60	\$80,450,596.98

Dimensions and Attributes

Ilustración 1 demostración de elementos de un cubo consultado con MDX


Comparación entre SQL y MDX

El uso de lenguaje SQL puede parecer que es exclusivo para la manipulación de bases de datos relacionales y sus derivados, pero no así, por ejemplo, Microsoft OLAP permite la ejecución de consultas con SQL, pero lo que variara será el resultado de dicha consulta ya que será retornado en forma de tablas, rompiendo por completo la estructura multidimensional a la cual solo se podrá tener acceso por medio de la ejecución de consultas en MDX, pero no se debe ignorar la presencia de palabras reservadas para sentencias SQL en sentencias MDX, por lo que, con el conocimiento de lenguaje SQL se busca hacer un acercamiento al lenguaje MDX mediante la comparación de las funciones de cada una de las sentencias o clausulas existentes entre ambos lenguajes (Cruz, 2003):

- En el caso de SQL, recordemos que la cláusula SELECT se usaba para indicar las tablas y atributos de estas que iban a ser empleados en la consulta, en el caso de MDX se especifican las dimensiones, sus miembros y medidas, separados por medio de comas, como en el siguiente ejemplo

```
SELECT [Persona].[Sexo], [Grado].[Cursos],
[Medidas].[Num_deserciones]
```

- En la cláusula WHERE, en SQL se indicaban las tablas de donde se iban a extraer los atributos especificados en la sentencia SELECT, en el caso de

 UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ	Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse	MDX – Expresiones multidimensionales	2021 - 2
---	---	--	----------

MDX, se especificará únicamente el cubo del cual se extraerán las dimensiones, miembros y medidas definidos en la cláusula SELECT, como se muestra en el siguiente ejemplo

```
SELECT      [Persona].[Sexo],          [Grado].[Cursos],
[Medidas].[Num_decerciones] FROM [Deserciones]
```


- En el caso de las cláusulas WHERE y GROUP BY, recordemos que en el caso de la primera en sentencias SQL se especificaban condiciones por medio de las cuales se filtrara el resultado de la consulta y con la segunda cláusula se definía el o los atributos por medio de los cuales se agrupan los resultados de la consulta, en el caso de MDX, estas dos cláusulas son empleadas para definir las medidas de los cubos, tal cual como se muestra a continuación

```
SELECT      [Persona].[Sexo],          [Grado].[Cursos],
[Medidas].[Num_decerciones] FROM [Deserciones] WHERE
[Medidas].[Num_decerciones]
```

```
SELECT      [Persona].[Sexo],          [Grado].[Cursos],
[Medidas].[Num_decerciones] FROM [Deserciones] GROUP BY
[Persona].[Sexo], [Grado].[Cursos]
```

- El uso de la cláusula DISTINCT es exclusivo de las consultas SQL, pero puede ser usado para realizar consultas sobre cubos OLAP justo después de la cláusula SELECT, pero no se puede emplear en los siguientes casos:
 - o Si la cláusula SELECT contiene miembros, no se debe usar esta cláusula ni la cláusula GROUP BY.
 - o En el caso de que se requiera ver filas duplicadas, no se recomienda usar esta sentencia ya que eliminara todos los registros duplicados.

En resumen, al momento de realizar consultas sobre Cubos OLAP, se recomienda hacer uso enteramente del lenguaje MDX, el cual asegurara el acceso a la información de la forma en que esta se encuentra estructurada, y cabe destacar la importancia en que se emplean las cláusulas o sentencias acorde a las dimensiones, miembros y medidas definidas en cada uno de los cubos.

 UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ	Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse	MDX – Expresiones multidimensionales	2021 - 2
--	---	--	----------

TPC-H Query 4 (relational SQL)	Equivalent Query 4 (multidimensional MDX)
<pre>SELECT o_orderpriority, COUNT(*) AS order_count FROM orders WHERE o_orderdate >= '1993-07-01' AND o_orderdate < '1993-10-01' AND EXISTS (SELECT * FROM lineitem WHERE l_orderkey = o_orderkey AND l_commitdate < l_receiptdate) GROUP BY o_orderpriority ORDER BY o_orderpriority</pre>	<pre>SELECT {[Measures].[FACT Latecount]} ON COLUMNS, {[PriorityDim].[All PriorityDim].[1], [PriorityDim] PriorityDim].[2], [PriorityDim].[All PriorityDim]. [PriorityDim].[All PriorityDim].[4], [PriorityDim] PriorityDim].[5]} ON ROWS FROM Q4Cube WHERE ([TimeDim].[All TimeDim].[1993].[Qua</pre>

Ilustración 2 Cuadro comparativo de consultas SQL y MDX

Ejemplos de consultas


Al conocer los datos o elementos y sentencias básicas existentes para la ejecución de consultas en MDX, se puede definir una estructura básica para las consultas, todo esto se hará en base a la información expuesta por (Simon, 2006):

```
SELECT
    <especificación de eje> on columns,
    <especificación de eje> on rows
FROM
    <especificación de cubo>
WHERE
    <especificación Slicer (rebanador)>
```

De modo tal, se puede ver como en la sentencia SELECT se va a especificar las dimensiones que poblaran los ejes de las columnas y las filas, en la sentencia FROM se define el cubo del que se extraerá dichas dimensiones y finalmente, por medio de la sentencia WHERE se definiría filtros o limitantes que delimiten los datos a extraer, tal cual como en la operación de Slice, la cual concentra la consulta en las dimensiones de un cubo a ciertos parámetros. Para ver un cubo totalmente se puede abordar por medio de la siguiente instrucción:

```
SELECT
FROM
    <especificación de cubo>
```

A continuación, se expondrán ejemplos de consultas y se explicara en torno al ejercicio expuesto por (Bustos & Danysoft, 2019) en el cual se maneja un cubo de ventas de productos.

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>MDX Expresiones multidimensionales</p>	<p>– 2021 - 2</p>
---	---	---	-----------------------

1. **Consulta general al cubo:** si se quiere tener una vista general del cubo de ventas se ejecuta la siguiente consulta:

```
SELECT
FROM Sales
```

2. **Consulta de las medidas:** en caso de que se requieran ver las medidas del cubo, se especifican en la cláusula WHERE, como en este caso que se requiere ver las unidades vendidas, para esto se ejecuta la siguiente consulta:

```
SELECT
FROM Sales
WHERE
    [Measures].[Unit Sales]
```

3. **Consulta específica de un eje:** en el caso que se requiera ver, por ejemplo, una vista de los productos, se definirán por medio de la cláusula SELECT definiendo los miembros de la dimensión productos a los que requerimos tener acceso, tal cual como se define en la siguiente consulta.

```
SELECT
    {[Product].[All Products].[Drink].[Beverages],
    [Product].[All Products].[Drink].[Dairy],
    [Product].[All Products].[Food].[Eggs],
    [Product].[All Products].[Food].[Meat]} on columns
FROM Sales
```


Esta consulta retorna el siguiente resultado:

Beverages	Dairy	Eggs	Meat
13.573,00	4.186,00	4.132,00	1.714,00

Ilustración 3 Resultado de la consulta de un eje

4. **Consulta especificando de filas y columnas:** en el caso de que se requiera, por ejemplo, conocer las ventas de ciertos productos clasificados por departamento con relación al tiempo, se puede ejecutar la siguiente consulta.

```
SELECT
```


 UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ	Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse	MDX Expresiones multidimensionales	– 2021 - 2
--	---	--	-------------------

{[Product].[Product Department].Members} on columns,
 {[Yearly Income].[Yearly Income].Members} on rows
 FROM Sales

Esta consulta retornaría el siguiente resultado:


	Alcoholic Be	Beverages	Dairy	Baked Good	Baking Good	Breakfast Fc	Canned Foo	Canned Proc	Dairy
\$10K - \$30K	1.550,00	2.892,00	868,00	1.741,00	4.196,00	722,00	4.174,00	347,00	2.808,00
\$110K - \$13	278,00	655,00	187,00	375,00	839,00	141,00	825,00	95,00	616,00
\$130K - \$15	395,00	691,00	223,00	356,00	1.087,00	184,00	1.114,00	57,00	720,00
\$150K +	167,00	307,00	106,00	178,00	426,00	66,00	426,00	33,00	251,00
\$30K - \$50K	2.183,00	4.416,00	1.458,00	2.655,00	6.609,00	1.131,00	6.017,00	624,00	4.191,00
\$50K - \$70K	1.196,00	2.175,00	674,00	1.179,00	3.632,00	598,00	3.222,00	370,00	2.139,00
\$70K - \$90K	795,00	1.787,00	477,00	972,00	2.549,00	342,00	2.420,00	192,00	1.555,00
\$90K - \$110	274,00	650,00	193,00	414,00	907,00	133,00	828,00	94,00	605,00

Ilustración 4 Resultado de una consulta especificando filas y columnas

- Cambio de ejes e filtrado de las medidas:** en el caso que se requiera ver la información de ventas de productos con relación al tiempo solo se requerirá cambiar los valores que acompañan las sentencias *on columns* y *on rows*, además de puede ver la cantidad de unidades vendidas, las cuales están definidos por medio de la medida definida en el cubo, tal cual como se muestra en la siguiente consulta:

SELECT
 {[Yearly Income].[Yearly Income].Members} on columns,
 {[Product].[Product Department].Members} on rows
 FROM Sales
 WHERE [Measures].[Unit Sales]

Esta consulta retornaría algo parecido a la siguiente imagen:

 UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ	Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse	MDX Expresiones multidimensionales	2021 - 2
--	---	---	-----------------


	\$10K - \$30K	\$110K - \$13	\$130K - \$15	\$150K +	\$30K - \$50K	\$50K - \$70K	\$70K - \$90K	\$90K - \$110
Alcoholic Bev	1.550,00	278,00	395,00	167,00	2.183,00	1.196,00	795,00	274,00
Beverages	2.892,00	655,00	691,00	307,00	4.416,00	2.175,00	1.787,00	650,00
Dairy	868,00	187,00	223,00	106,00	1.458,00	674,00	477,00	193,00
Baked Good	1.741,00	375,00	356,00	178,00	2.655,00	1.179,00	972,00	414,00
Baking Good	4.196,00	839,00	1.087,00	426,00	6.609,00	3.632,00	2.549,00	907,00
Breakfast Fc	722,00	141,00	184,00	66,00	1.131,00	598,00	342,00	133,00
Canned Foo	4.174,00	825,00	1.114,00	426,00	6.017,00	3.222,00	2.420,00	828,00
Canned Proc	347,00	95,00	57,00	33,00	624,00	370,00	192,00	94,00
Dairy	2.808,00	616,00	720,00	251,00	4.191,00	2.139,00	1.555,00	605,00
Deli	2.700,00	471,00	672,00	202,00	4.046,00	1.957,00	1.479,00	510,00
Eggs	975,00	129,00	171,00	93,00	1.307,00	726,00	545,00	186,00
Frozen Foo	5.644,00	1.285,00	1.419,00	611,00	8.847,00	4.393,00	3.240,00	1.216,00
Meat	358,00	49,00	86,00	42,00	585,00	335,00	188,00	71,00
Produce	8.429,00	1.616,00	1.948,00	765,00	12.245,00	6.384,00	4.673,00	1.732,00

Ilustración 5 Resultado del cambio de ejes y filtrado de medidas

6. **Consulta de medidas sobre una dimensión:** en el caso que se requiera saber la relación entre las medidas y una dimensión con en el caso de los productos, se definirán la dimensión, los miembros de esta a listar y las medidas a visualizar en la cláusula SELECT, tal cual como se muestra a continuación:

```
SELECT
    {[Measures].[MeasuresLevel].Members} on columns,
    {[Product].[Product Department].Members} on rows
FROM Sales
```

Esta consulta retornaría el siguiente resultado:

 UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ	Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse	MDX Expresiones multidimensionales	2021 - 2
--	---	---	-----------------

	Unit Sales	Store Cost	Store Sales	Sales Count	Store Sales I
Alcoholic Bev	6.838,00	5.576,79	14.029 €	2219	8.452,29
Beverages	13.573,00	11.069,53	27.749 €	4392	16.679,00
Dairy	4.186,00	2.830,92	7.059 €	1367	4.227,68
Baked Good	7.870,00	6.564,09	16.455 €	2561	9.891,34
Baking Good	20.245,00	15.370,61	38.670 €	6588	23.299,80
Breakfast Fc	3.317,00	2.756,80	6.941 €	1078	4.184,66
Canned Foo	19.026,00	15.894,53	39.774 €	6182	23.879,81
Canned Proc	1.812,00	1.317,13	3.315 €	590	1.997,39
Dairy	12.885,00	12.228,85	30.509 €	4189	18.280,00
Deli	12.037,00	10.108,87	25.319 €	3908	15.210,06
Eggs	4.132,00	3.684,90	9.201 €	1339	5.515,86
Frozen Foo	26.655,00	22.030,66	55.207 €	8684	33.176,84
Meat	1.714,00	1.465,42	3.670 €	555	2.204,47
Produce	37.792,00	32.831,33	82.248 €	12310	49.417,09

Ilustración 6 Resultado de consulta de medidas sobre una dimensión


7. **Consulta con varios slicers/rebanadores:** cuando se requiere filtrar la información bajo mas de una condición, solo se requiere definir los miembros de las dimensiones y medidas en la clausula WHERE, como el el siguiente ejemplo el cual busca conocer los productos y la promoción que ha recibido por los medios almacenados, filtrando los datos con base a los datos de los clientes, de modo tal se obtiene la siguiente consulta:

```

SELECT
    {[Product].[All Products].[Drink].[Alcoholic
    Beverages].[Beer and Wine],
    [Product].[All Products].[Drink].[Beverages].[Carbonated
    Beverages],
    [Product].[All Products].[Drink].[Beverages].[Pure Juice
    Beverages]} on columns,
    {[Promotion Media].[All Media].[TV]} on rows
FROM Sales
WHERE
    ([Customers].[All Customers].[USA].[CA],
    [Time].[1997].[Q2],
    [Marital Status].[All Marital Status].[M],
    [Measures].[Profit])

```

Esta consulta retorna los siguientes valores:

 UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ	Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse	MDX Expresiones multidimensionales	2021 - 2
--	---	---	-----------------

	Beer and Wine	Carbonated	Pure Juice B
TV	26,00	15,06	
Radio			

Ilustración 7 Resultado de consulta con dos slicers

En base a esta misma consulta y teniendo en cuenta que se retornan celdas vacías, se puede controlar que solo se obtengan resultados con algún valor, por medio de la clausula **NON EMPTY** dentro de la especificación de ejes, tal cual como se muestra a continuación:

```

SELECT
    NON EMPTY
    {[Product].[All Products].[Drink].[Alcoholic
    Beverages].[Beer and Wine],
    [Product].[All Products].[Drink].[Beverages].[Carbonated
    Beverages],
    [Product].[All Products].[Drink].[Beverages].[Pure Juice
    Beverages]} on columns,
    {[Promotion Media].[All Media].[TV]} on rows
FROM Sales
WHERE
    ([Customers].[All Customers].[USA].[CA],
    [Time].[1997].[Q2],
    [Marital Status].[All Marital Status].[M],
    [Measures].[Profit])

```


Obteniendo el siguiente resultado:

	Beer and Wine	Carbonated
TV	26,00	15,06

Ilustración 8 Resultado de consulta eliminando espacios vacíos





A parte de los ejemplos anteriormente expuestos, cabe mencionar otras acciones que se pueden realizar por medio de MDX:


- Combinación y clasificación entre combinaciones
- Cálculos entre medidas y celdas
- Ordenamiento y filtrado basándose en los valores de las medidas
- Consultas con métodos de agregaciones (sumas, medias, máximos, etc.)

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>MDX – Expresiones multidimensionales</p>	<p>2021 - 2</p>
---	---	---	-----------------

Herramientas

A continuación, se expondrán algunas herramientas en las cuales se puede emplear sentencias MDX:

Icono	Nombre	Descripción
	<p><i>Microsoft SQL Server Analysis Services</i></p>	<p>Motor de datos analíticos creado y distribuido por Microsoft que permite el análisis empresarial, teniendo presencia en plataformas como Power BI y brindando conectividad con herramientas de Business Inteligente y visualizadoras de datos. (Microsoft, 2021)</p>
	<p><i>Excel</i></p>	<p>Herramienta para la gestión de tablas, el cual se emplea para la visualización de consultas por medio de tablas dinámicas generadas por motores de análisis de datos. (Microsoft, 2019)</p>
 <p>Power BI</p>	<p><i>Power BI</i></p>	<p>Servidor OLAP que trabaja con el motor Vertipaq de Microsoft SQL Server Analysis Services permitiendo la conexión con aplicaciones en arquitectura cliente servidor y la integración de bibliotecas para la manipulación de cubos OLAP. (Microsoft, 2021)</p>
 <p>MONDRIAN</p>	<p><i>Mondrian</i></p>	<p>Servidor OLAP open source que permite gestionar la comunicación entre aplicaciones OLAP desarrolladas con Java, operando como una especie de JDBC para cubos OLAP relacionales, permitiendo la consulta de datos por medio de sentencias MDX. (Universidad de Granada & Samos Jiménez, 2021)</p>

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>MDX Expresiones multidimensionales</p>	<p>2021 - 2</p>
---	---	---	-----------------

ETL

En los procesos de consolidación de datos se debe contemplar procesos por medio de los cuales se realice las modificaciones necesarias a los datos almacenados en las fuentes identificadas para ser volcados o cargados al Data Warehouse, por lo que se requiere de realizar copias de los datos desde las fuentes identificadas, realizar modificaciones a los datos copiados y finalmente cargarlos, estos procesos no se realizan de forma manual, sino por medio de herramientas de software que permitan realizar los tres subprocesos de extracción, transformación y carga de datos, antes de ver que herramientas o métodos existen para realizar el proceso de consolidación de datos, se deben ver las etapas de estos procesos (Abella et al., 1999; Grupo Atico34, 2021; Gutiérrez, 2011; López, 2002; Sierra, 2020).

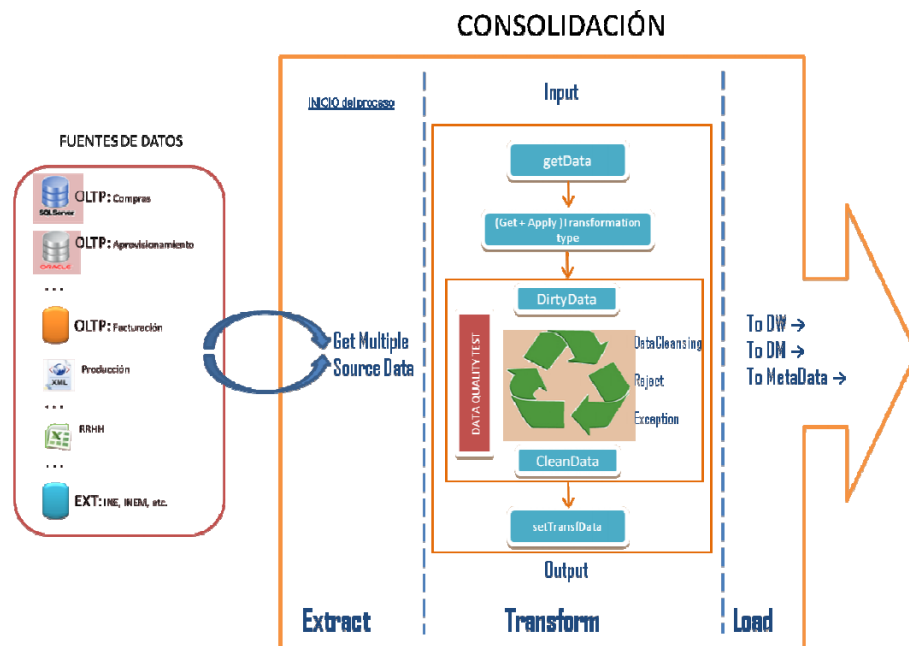



Ilustración 9 Procesos de consolidación

Extracción (E)

Previamente se realizó el proceso de selección de fuentes de datos, las cuales pueden proveernos de datos, sean estructurados o no estructurados, tales como las bases de datos transaccionales de la empresa, en donde los datos tienen cierta estructura y permiten la extracción de información, y por otro lado nos podemos encontrar con una variedad de archivos los cuales también pueden proveer ciertos datos pero conllevando procesos específicos para la obtención de los datos y su posterior estructuración. En esta etapa se deben tener en cuenta las necesidades o requerimientos expuestos previamente por la empresa y de esta forma se podrá

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>MDX – Expresiones multidimensionales</p>	<p>2021 - 2</p>
---	---	---	-----------------

identificar, seleccionar y unir todos aquellos datos que cumplan con dichas necesidades, se puede definir que en esta etapa se filtran las fuentes de datos seleccionadas con el fin de dejar en un área de almacenamiento temporal a convertir en datos históricos. (Gutiérrez, 2011; Vega Fajardo & Chiriguaya Freire, 2014)

Transformación (T)

Al seleccionar los datos que cumplen con las necesidades y requerimientos debemos contemplar que estos datos se encuentran estructurados todos de una forma diferente entre sí, esto se puede expresar bajo un ejemplo muy simple, contemplando el resultado de una consulta de una base de datos como SQL, en donde los datos son de cierto tipo y por otro lado tenemos los datos que se encuentran que una hoja de cálculo de Excel, ambos tipos de datos son necesarios y se van a almacenar dentro del almacén de datos pero para esto se deben aplicar actividades por medio de las cuales se realice una corrección de errores de tipado, eliminación de campos innecesarios y demás procesos que permitan la normalización y estructuración de los datos obtenidos. (Vega Fajardo & Chiriguaya Freire, 2014)

Carga (L)

Finalmente, al realizar los procesos anteriores, nos encontraremos con datos de calidad, los cuales ya han sido tratados con el fin de ser almacenados en el DW, en esta etapa se implementaran métodos de carga, los cuales garanticen que los datos se almacenen dentro del almacén, bajo cierta estructura y contemplando que los datos que se cargan son de gran volumen, dichos métodos de carga deben garantizar los tiempos más óptimos, además de contemplar la carga de datos históricos o actualizados y el control de las transformaciones aplicadas en el proceso de transformación. (Gutiérrez, 2011; Vega Fajardo & Chiriguaya Freire, 2014)




 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>MDX Expresiones multidimensionales</p>	<p>2021 - 2</p>
---	---	---	-----------------





Ilustración 10 ETL y actividades a realizar en cada etapa


Herramientas ETL

Al momento de realizar los procesos anteriormente indicados, nos encontraremos con dos caminos que se pueden seguir: el primero es el uso de herramientas incluidas en servicios de Business Intelligence que nos permiten realizar los procesos ETL de forma optima y rápida, y segundo, la creación de programas los cuales este hecho acorde a cada una de las fuentes de datos definidas, contemplando la extracción, manipulación y carga de los datos al almacén de datos. En resumidas cuentas: uso de herramientas de terceros o creación de software propio, por lo que, a continuación, nos enfocaremos en la definición de algunas herramientas creadas por terceros que prestan las funcionalidades de ETL:

Icono	Nombre	Descripción
	<p><i>Microsoft SQL Server Analysis Services</i></p>	<p>Dentro de las funcionalidades que presenta este motor de análisis se encuentra el de acceder a una variedad de fuentes de datos y realizando los procesos de transformación y carga de datos directamente al Data Warehouse (Microsoft, 2021)</p>
	<p><i>Oracle Data Integrator</i></p>	<p>Esta es una plataforma que permite la integración de los datos desde las fuentes que tenga la organización, de modo tal que permite la integración</p>


 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>MDX – Expresiones multidimensionales</p>	<p>2021 - 2</p>
---	---	---	-----------------

		directamente con las demás herramientas para Data Warehouse prestadas por Oracle (Oracle, 2021).
 <p>MONDRIAN</p>	<p><i>Pentaho Mondrian</i></p>	<p>Siendo un servidor OLAP open source, esta herramienta contiene un apartado relacionado con la integración de los datos desde las fuentes, en las cuales se puede obtener datos desde bases de datos, documentos, archivos y demás elementos que pueden ser procesados por las herramientas de Data Integrator (Universidad de Granada & Samos Jiménez, 2021)</p>

 <p>UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ</p>	<p>Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse</p>	<p>MDX – Expresiones multidimensionales</p>	<p>2021 - 2</p>
---	---	---	-----------------

BIBLIOGRAFÍA

- Abella, R., Cópola, L., & Olave, D. (1999). *Sistema DataWarehousing*.
- Bustos, J., & Danysoft. (2019). *Expresiones MDX en Analysis Services*.
https://www.danysoft.com/estaticos/free/mdx_as.pdf
- Cruz, R. (2003). *Utilización de la tecnología Data Warehouse en instituciones educativas* [Universidad Autónoma del estado de Hidalgo].
[http://dgsa.uaeh.edu.mx:8080/bibliotecadigital/bitstream/handle/123456789/47/Utilizacion de la tecnología.pdf?sequence=1](http://dgsa.uaeh.edu.mx:8080/bibliotecadigital/bitstream/handle/123456789/47/Utilizacion%20de%20la%20tecnologia.pdf?sequence=1)
- Grupo Atico34. (2021, February 11). *Data warehouse. Definición y funciones*. Grupo Atico34.
https://protecciondatos-lopd.com/empresas/data-warehouse/#Funciones_de_los_data_warehouses
- Gutiérrez, P. (2011). *DATA WAREHOUSE: MARCO DE CALIDAD*. UNIVERSIDAD CARLOS III.
- Koo, A. (2014). *Introducción a Multidimensional Expressions (MDX)* .
<https://docplayer.es/24345270-Introduccion-a-multidimensional-expressions-mdx.html>
- López, B. (2002). *GUIAN PARA LA CONSTRUCCION DE UN DATA WAREHOUSE* (01 ed., Vol. 01). Universidad Autónoma de Nuevo León.
<http://eprints.uanl.mx/1131/1/1020147975.pdf>
- Microsoft. (2017, May 3). *Introduction to OLE DB for OLAP* . Microsoft Docs.
[https://docs.microsoft.com/es-mx/previous-versions/windows/desktop/ms725355\(v=vs.85\)](https://docs.microsoft.com/es-mx/previous-versions/windows/desktop/ms725355(v=vs.85))
- Microsoft. (2018, February 5). *Querying Multidimensional Data with MDX* . Microsoft Docs.
<https://docs.microsoft.com/en-us/analysis-services/multidimensional-models/mdx/querying-multidimensional-data-with-mdx?view=asallproducts-allversions>
- Microsoft. (2019, April 19). *Conceptos clave de MDX (Analysis Services)*. Microsoft Docs.
<https://docs.microsoft.com/es-es/analysis-services/multidimensional-models/mdx/key-concepts-in-mdx-analysis-services?view=asallproducts-allversions>
- Microsoft. (2021, April 22). *¿Qué es Analysis Services?* . Microsoft Docs.
<https://docs.microsoft.com/es-es/analysis-services/analysis-services-overview?view=asallproducts-allversions>
- Oracle. (2021, January 15). *Oracle Data Integrator*. Oracle Colombia.
<https://www.oracle.com/co/middleware/technologies/data-integrator.html>

 UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ	Universidad de San Buenaventura Facultad de ingeniería Optativa II: Data Warehouse	MDX – Expresiones multidimensionales	2021 - 2
--	---	--	----------

Sierra, Y. (2020, July 20). *DWH: ejemplos, características y arquitectura del Data Warehouse - ADN Cloud*. Media Cloud. https://blog.mdcloud.es/dwh-ejemplos-arquitectura-data-warehouse/#Ejemplos_del_uso_de_DWH_en_distintos_sectores

Simon, A. (2006). *Propuesta de herramientas para la extracción y visualización de datos multidimensionales* [UNIVERSIDAD DE LA HABANA]. https://www.researchgate.net/publication/280255735_Propuesta_de_herramientas_para_la_extraccion_y_visualizacion_de_datos_multidimensionales

String Fixer. (2008, March 5). *Expresiones multidimensionales*. String Fixer. https://stringfixer.com/es/MultiDimensional_eXpressions

Universidad de Granada, & Samos Jiménez, J. (2021, January 15). *Capítulo 11 Mondrian | Sistemas Multidimensionales*. Sistemas Multidimensionales Prácticas Con Power BI y Power Query, Tableau y Tableau Prep, SSIS, SSAS, PDI, Mondrian y R. <https://lsi2.ugr.es/jsamos/sm2019/olap-mondrian.html>

Vega Fajardo, M., & Chiriguaya Freire, M. (2014). *Estudio y Diseño de un Data center alternativo para el Centro de Cómputo de la Universidad de Guayaquil* [Universidad de Guayaquil]. <http://repositorio.ug.edu.ec/bitstream/redug/17295/1/BFILO-PIN-0504.pdf>

TABLA DE ILUSTRACIONES

Ilustración 1 demostración de elementos de un cubo consultado con MDX	5
Ilustración 2 Cuadro comparativo de consultas SQL y MDX	7
Ilustración 3 Resultado de la consulta de un eje	8
Ilustración 4 Resultado de una consulta especificando filas y columnas	9
Ilustración 5 Resultado del cambio de ejes y filtrado de medidas	10
Ilustración 6 Resultado de consulta de medidas sobre una dimensión	11
Ilustración 7 Resultado de consulta con dos slicers	12
Ilustración 8 Resultado de consulta eliminando espacios vacíos	12
Ilustración 9 Procesos de consolidación	14
Ilustración 10 ETL y actividades a realizar en cada etapa	16