



Construcción de Cubo OLAP en Microsoft Analysis Services y Microsoft Excel

Building OLAP Cube in Microsoft Analysis Services and Microsoft Excel

Zully Kristel Guzmán Caraveo

División Académica de Ciencias y Tecnologías de la Información, Universidad Juárez Autónoma de Tabasco, México
zully_k1996@hotmail.com

Herman Aguilar Mayo

División Académica de Ciencias y Tecnologías de la Información, Universidad Juárez Autónoma de Tabasco, México
herman_aguilar1@hotmail.com

doi: <https://doi.org/10.36825/RITI.08.15.005>

Recibido: Noviembre 19, 2019

Aceptado: Febrero 29, 2020

Resumen: Este trabajo de investigación resulta de la construcción de un cubo OLAP (Procesamiento Analítico en Línea). El objetivo consiste en mostrar cómo es la construcción de un cubo OLAP, con el uso de las herramientas de Microsoft Analysis Services y Microsoft Excel, a partir de una base de datos obtenida de una encuesta realizada por el INEGI en el año 2016, la cual es tomada como ejemplo en esta investigación; el modelo desarrollado fue el modelo cascada, donde se implementaron las cinco etapas de este modelo, las cuales son la etapa de comunicación, planeación, modelado, construcción y despliegue; una de las principales conclusiones que se obtuvieron con la realización de esta investigación fue que las herramientas de Microsoft Analysis Services y Microsoft Excel, ayudan en la realización del procesamiento y análisis de datos de una base de datos por medio de dimensiones y medidas.

Palabras clave: *Cubos OLAP, Construcción, Analysis Services, Excel, Base de datos.*

Abstract: This research work results from the construction of an OLAP cube (Online Analytical Processing). The objective is to show how the construction of an OLAP cube is, using the tools of Microsoft Analysis Services and Microsoft Excel, from a database obtained from a survey conducted by INEGI in 2016, which It is taken as an example in this investigation; the model developed was the waterfall model, where the five stages of this model were implemented, which are the stage of communication, planning, modeling, construction and deployment; One of the main conclusions that were obtained with the accomplishment of this investigation was that the tools of Microsoft Analysis Services and Microsoft Excel, help in the realization of the processing and analysis of data of a database by means of dimensions and measures.

Keywords: *OLAP Cubes, Construction, Analysis Services, Excel, Database.*

1. Introducción

El Procesamiento Analítico en Línea (OLAP), es una categoría de software que permite el análisis multidimensional de datos; el análisis multidimensional es una técnica analítica que posibilita la visualización de los datos en un formato de cubo dimensional (o hiperdimensional), facilitando la selección y análisis [1]. De este modo, los productos OLAP proporcionan capacidades de modelización, análisis y visualización de grandes conjuntos de datos permitiendo la recuperación, manipulación y combinación de estos por medio de consultas o informes [2]. El término OLAP se presentó oficialmente en 1993 [3], publicado por Codd y asociados y apoyado por Arbor Software Corporation, Según Codd, OLAP con el análisis multidimensional, permite modelar la información en medidas, dimensiones y hechos [4] [5]:

- Medidas: son los valores de un dato.
- Dimensiones: son las descripciones de las características que definen dicho dato.
- Hechos: corresponden a la existencia de valores específicos de una o más medidas para una combinación particular de dimensiones.

La propuesta original, el modelo multidimensional no necesariamente se tiene que almacenar previamente en una base de datos multidimensional, sino que plantea que puede acceder directamente a múltiples fuentes de información, como base de datos (relacionales o multidimensionales), archivos planos, hojas de cálculo e incluso algunos datos pueden ser introducidos por usuarios finales [3]. Una vez adquiridos todos los datos, se consolida y organiza la información en el modelo lógico multidimensional, para luego presentarla al usuario [6]. En la Figura 1, describen todas las operaciones que se pueden realizar en las dimensiones y medidas dadas por el cubo.

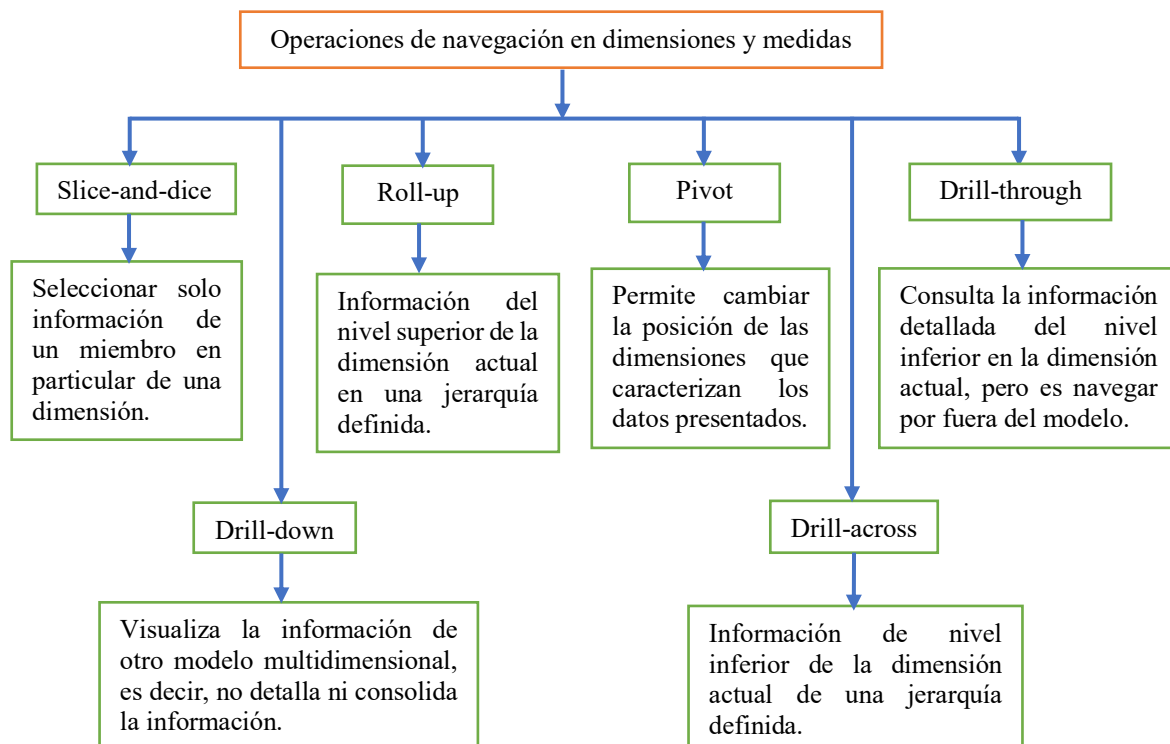


Figura 1. Operaciones de navegación en las dimensiones y agrupaciones de medidas en OLAP [7].

Las herramientas que se utilizaron para el desarrollo del cubo OLAP, pertenecen a la suite de Microsoft, las cuales se utilizan para explotar los datos que se encuentran en el gestor de base de datos y con ellos poder construir, analizar y tomar decisiones de acuerdo a los datos que obtengan con el análisis de estos datos.

La herramienta de Analysis Services se describe como un motor de datos analíticos en línea que se usa en soluciones de ayuda a la toma de decisiones y Business Intelligence (Inteligencia de Negocios), la cual proporciona los datos analíticos para informes empresariales y aplicaciones cliente como Excel, informes de Reporting Services y otras herramientas [8].

Microsoft Excel, es un tipo de hoja de cálculo, la cual se utiliza para calcular, analizar y gestionar datos; con este programa se puede realizar diversas operaciones desde las más sencillas hasta las más complejas, por medio de funciones y formulas [9]. Con Excel 2010 se ofrecieron nuevas características y componentes para la visualización y análisis de datos, estos son los componentes de Power Query y Power Pivot, con estos componentes se puede conectar Excel con datos externos desde una base de datos hasta servidores en línea, con estas conexiones se pueden utilizar tablas y gráficos dinámicos que brindan la posibilidad de analizar, explorar y presentar datos como otras herramientas que se utilizan para el Procesamiento Analítico en Línea.

Es por ello en esta investigación se presenta, el desarrollo de cómo es la construcción, implementación y visualización de los datos o consultas de un cubo OLAP tanto en Microsoft Analysis Services y Microsoft Excel, para que los usuarios tengan conocimiento de cómo se puede realizar un análisis de datos de N cantidad de registros de una base de datos en una herramienta especializada como lo es Microsoft Analysis Services, pero a su vez como también se realiza esto en Microsoft Excel, el cual es un programa que comúnmente los usuarios utilizan o pueden utilizar.

2. Metodología

El enfoque de estudio de esta investigación es de tipo cualitativo, porque este enfoque según [10] se guía por áreas o temas significativos de investigación, además pueden desarrollar preguntas o supuestos antes, durante o después de la recolección y análisis de los datos; es por ello que se utiliza la recolección de datos sin medición numérica para descubrir o afinar preguntas de investigación en el proceso de interpretación.

El modelo de desarrollo para la investigación presente utilizado fue el modelo cascada; ya que con este modelo podemos identificar desde el principio todos los elementos o etapas que se desarrollaran durante la investigación (véase Figura 2); el modelo de cascada se describe como un ciclo de vida, el cual sugiere un enfoque sistemático y secuencial para el desarrollo del software, el cual comienza con la especificación de los requerimientos por parte del cliente (institución u organización), avanza a la planeación, modelado, construcción y despliegue [11].

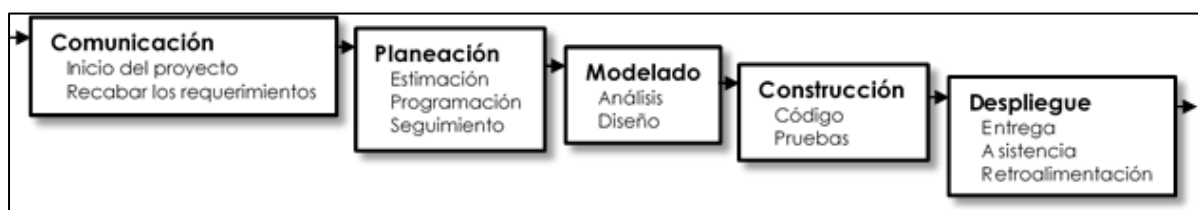


Figura 2. Modelo cascada [11].

2.1 Aplicación de la Metodología

Aplicando el modelo cascada, se desarrolló la primera etapa del modelo, la cual es Comunicación; en ella se recabaron los requerimientos que se necesitaba para la construcción del cubo, en ella se buscó una base de datos, la cual se utilizaría de forma de ejemplo para poder realizar la construcción del cubo OLAP y se eligió una base de datos del Instituto Nacional de Estadística Geografía (INEGI) del año 2016, para utilizarla como ejemplo, la cual es pública y se encuentra en la página oficial del INEGI, está conformada por 12 tablas y la doceava tabla es una tabla resumen con información a nivel hogares [12]. La base de datos se tendrá que migrar de un gestor de base de datos a otro gestor de base de datos, ya que se encuentra en Oracle y se migrará a Microsoft SQL Server.

La etapa de Planeación, tiene como fin desarrollar, identificar y plantear los problemas que se pueden presentar a lo largo de la construcción del cubo OLAP; se identificaron algunos datos que no serán necesarios tener para la construcción del cubo, estos datos se pueden eliminar o dejarlos y solo ignorarlos; en este caso algunas columnas de algunas tablas fueron eliminadas y otros datos como contienen información relacionada con otras tablas no se eliminaron solo se hizo una depuración lógica (ignorar los datos).

La etapa de Modelado se dividió en otras dos subetapas, las cuales son el Análisis de la base de datos y el Diseño; la subetapa del Análisis de la base de datos fue para analizar con más profundidad la base de datos que se utilizó para la construcción del cubo, la base de datos contiene información asociada a tres niveles o grupos de estudio y una tabla donde concentra información de hogares. En las Tablas 1, 2 y 3 describen los grupos e información [12]:

Tabla 1. Descripción del grupo información de las viviendas [12].

Tabla con información de las viviendas.	
VIVIENDAS	Contiene las características de las viviendas, el recuento de los residentes y los hogares de la vivienda, la ubicación y el diseño muestra correspondiente a esta.

Tabla 2. Descripción del grupo información de los hogares [12].

Tabla con información de los hogares.	
HOGARES	Se integra con información sobre los huéspedes y trabajadores domésticos del hogar, el acceso a la alimentación de los integrantes del hogar, el equipamiento del hogar, entre otros.
GASTOHOGAR	Contiene los gastos monetarios y no monetarios que realizó el hogar en el periodo de referencia, así como la estimulación del alquiler.
EROGACIONES	Erogaciones financieras y de capital realizadas por hogar.
GASTOTARJETAS	Gastos realizados por el hogar y que fueron cubiertos mediante alguna tarjeta de crédito bancaria y/o comercial.

Tabla 3. Descripción del grupo información de las personas [12].

Tablas con información de las personas.	
POBLACION	Características sociodemográficas y ocupacionales de los integrantes del hogar.
INGRESOS	Ingresos y percepciones de capital de cada uno de los integrantes del hogar.
GASTOSPERSONAS	Contiene los gastos monetarios y no monetarios asociados a los integrantes del hogar en el periodo de referencia.
TRABAJOS	Condición de actividad de los integrantes del hogar de 12 o más años.
AGRO	Ingresos y gastos de los negocios del hogar dedicados tanto a las actividades agrícolas, forestales y de tala, como actividades de cría, explotación y productos derivados de la pesca y caza.
NOAGRO	Ingresos y gastos de los negocios del hogar dedicados a las actividades industriales, comerciales y de servicios, y sus características propias.

En la subetapa de Diseño en el Modelado, de acuerdo a los 3 grupos anteriormente descritos, se pueden construir tres cubos OLAP para tener una mejor visualización de los datos de cada tabla; pero en esta ocasión se decidió realizar el diseño y la construcción del cubo OLAP sobre la información de las Viviendas, en la Figura 3 se puede apreciar el diagrama del cubo de Viviendas.

La etapa de Construcción, se llevó a cabo la construcción del cubo con la información de Viviendas, ya explicado anteriormente, la construcción del cubo OLAP se realizó primero en Analysis Services, en Visual Studio ya que este entorno utiliza estos complementos para la creación de cubos OLAP, en Analysis Services se tiene que construir el origen de datos, vista de origen de datos, cubos y dimensiones (véase Figura 4), una vez construido estos puntos se procede a la implantación del cubo OLAP (véase Figura 5).

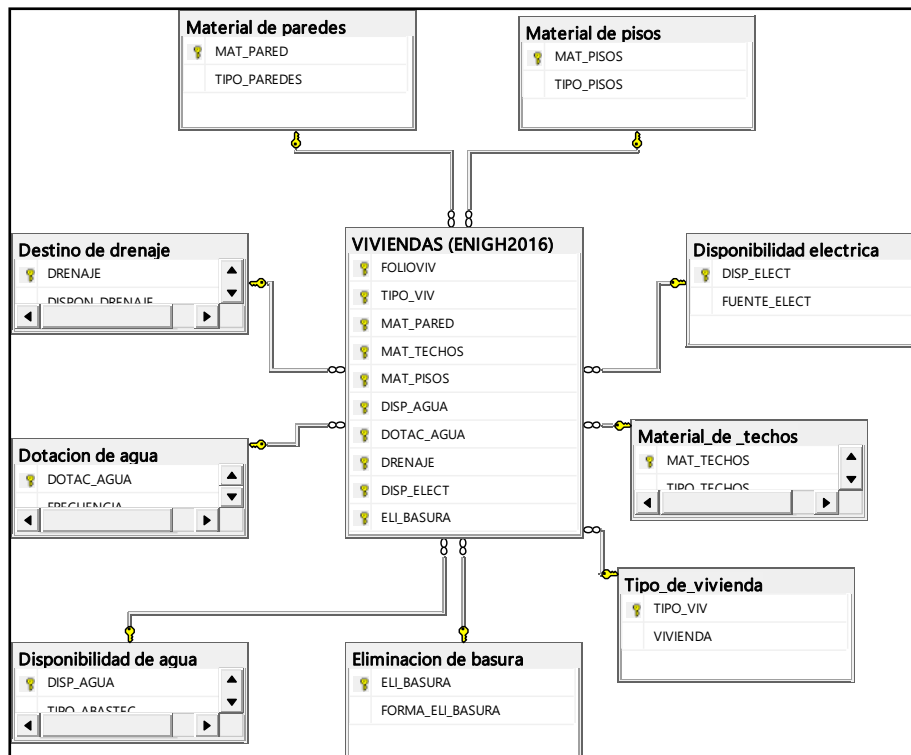


Figura 3. Diagrama del grupo vivienda

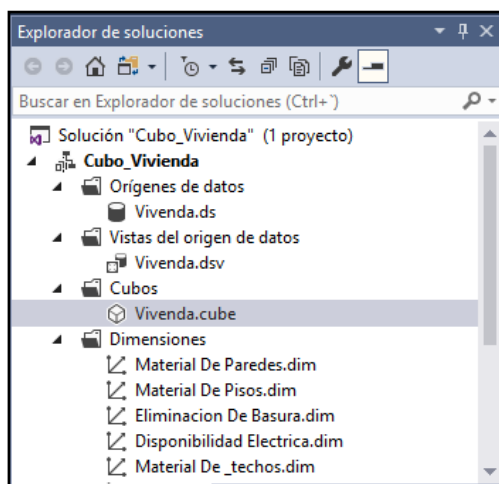


Figura 4. Creación de los elementos del cubo OLAP en Analysis Services.

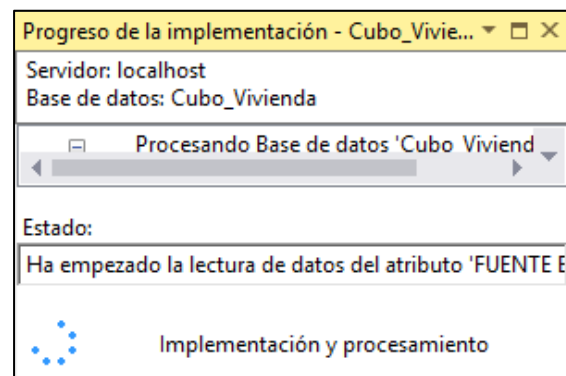


Figura 5. Implementación del cubo OLAP en Analysis Services.

En la Figura 6 se muestra el cubo OLAP realizado en Analysis Services, implementado y procesado, con las dimensiones y medidas que conforman el cubo; la medida de este cubo es la tabla que se encuentra en el centro y esta de color amarillo el nombre de la tabla, y las dimensiones son las que se encuentran alrededor de la tabla medida y esta representadas con azul el nombre de cada una de las dimensiones.

Una vez realizado el cubo en Analysis Services, se procedió a crear el cubo en Excel, en Excel hay dos formas de crear un cubo OLAP y visualizar los datos desde una conexión a un gestor de base de datos como lo es Microsoft SQL Server, importando los datos de tabla(s) o realizando una conexión a un cubo realizado en Analysis Services

(véase Figura 7); en este caso la construcción del cubo en Excel se realizó por medio de la conexión al gestor de base de datos Microsoft SQL Server y se crearon las tablas dinámicas para poder realizar la visualización de los datos (véase Figura 8).

La última etapa fue el Despliegue, en donde se pudo visualizar los datos, medidas y dimensiones del cubo, para así poder ver todas combinaciones posibles que se pueden obtener con los cubos OLAP y la base de datos utilizada.

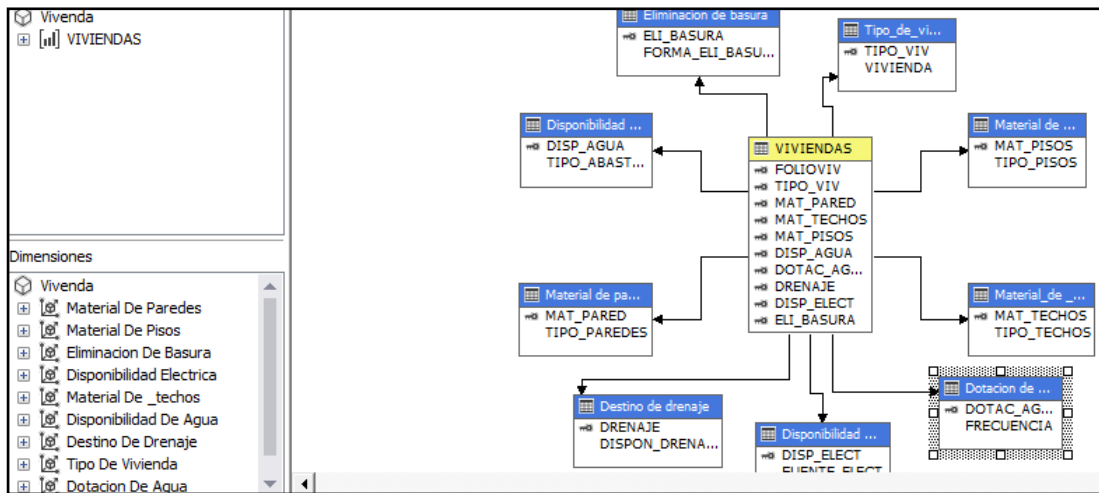


Figura 6. Representación del cubo OLAP en Analysis Services.

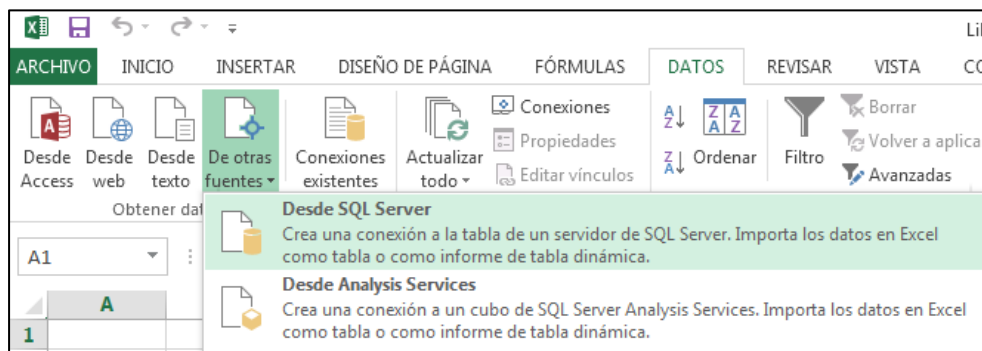


Figura 7. Creación de la conexión desde SQL Server a Excel para la creación del cubo.

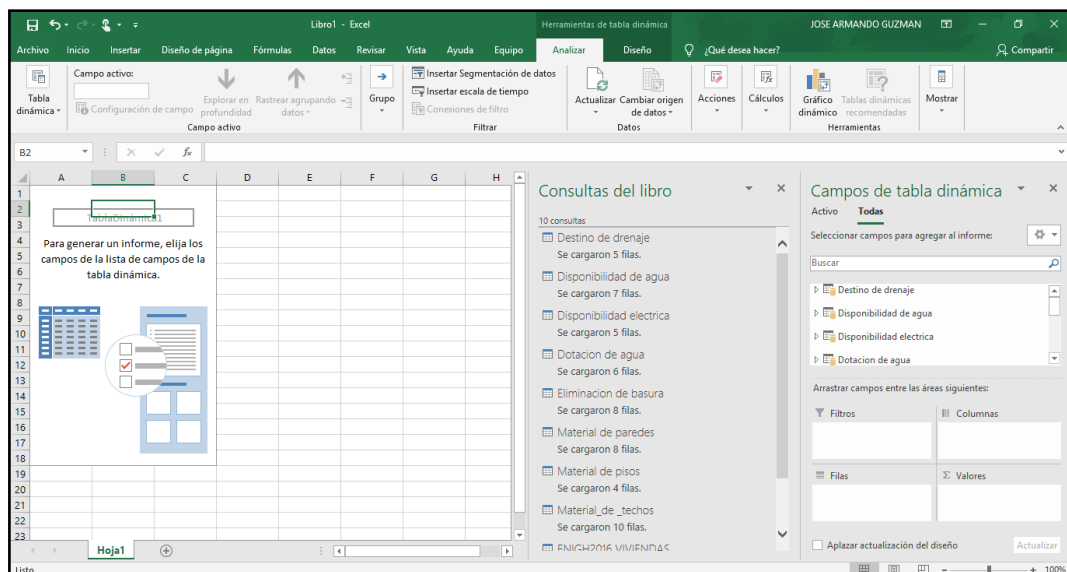


Figura 8. Creación de las tablas dinámicas para la visualización de datos en Excel.

3. Resultados

Con la creación e implementación del cubo, se pueden crear muchas combinaciones entre cada dimensión y medida, ya que la base de datos contiene gran cantidad de registros y eso hace que las posibles combinaciones sean muchas; tanto con Analysis Services y Excel se pueden analizar y tener los mismos resultados que se obtiene en una herramienta como en otra, ya que se creó el mismo diseño del cubo para utilizarlo en ambas herramientas; en Analysis Services se hace las consultas en el mismo programa en el navegador que proporciona y en cambio en Excel se inserta una tabla dinámica para poder crear las consultas.

A modo de ejemplo, se muestran algunos de los resultados obtenidos con estos cubos creados, utilizando el grupo de medida y las dimensiones a utilizar en cuestión, en cada consulta y además de poder realizar otras consultas con las dimensiones y medida que se encuentra en el cubo OLAP, con solo arrastrar o seleccionar los campos necesarios. En las Figuras 9 y 10, se obtuvo información acerca del número de viviendas que tienen un piso de tipo cemento o firme, madera, mosaico u otro recubrimiento y de tierra, pero a la vez se encontraron con datos perdidos los cuales se clasificaron como no especificado. Se obtuvo que 37,986 viviendas cuentan con un piso de cemento o firme y 28,958 viviendas con un piso de madera, mosaico u otro recubrimiento, pero también se obtiene 4 datos que no especificaron el tipo de pisos que tienen en sus viviendas.

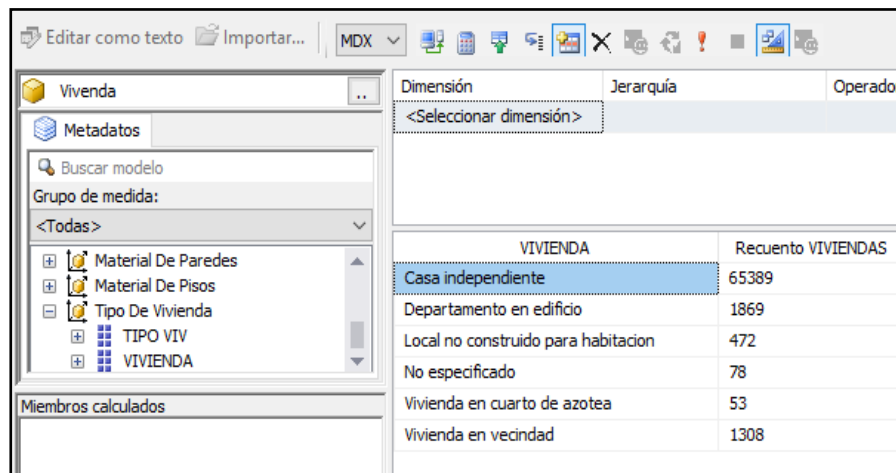
TIPO PISOS	Recuento VIVIENDAS
Cemento o firme	37986
Madera, mosaico u otro recubrimiento	28958
No especificado	4
Tierra	2221

Figura 9. Consulta sobre el tipo de pisos en las viviendas en Analysis Services.

Etiquetas de fila	Recuento de FOLIOVIV
Cemento o firme	37986
Madera, mosaico u otro recubrimiento	28958
No especificado	4
Tierra	2221
Total general	69169

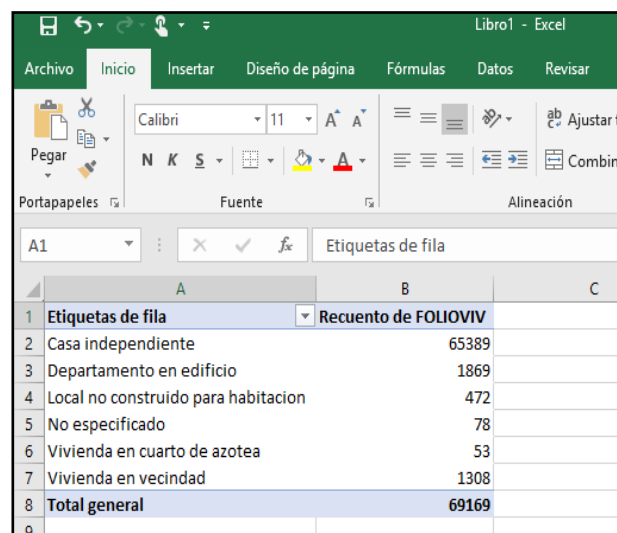
Figura 10. Consulta sobre el tipo de pisos en las viviendas en Excel.

Otra consulta realizada fue utilizando la medida de la tabla vivienda y la dimensión del tipo de vivienda, las Figuras 11 y 12, respectivamente en Analysis Services y Excel, muestra como hay seis tipos de viviendas en la base de datos y el número de viviendas que existen de cada tipo. De lo anterior se obtuvo que 65,389 viviendas son del tipo casa independiente, es decir, es la población con casa propia; pero también se puede observar que el tipo de vivienda con menor índice de la población que habitan en ella son las viviendas en cuarto de azotea y otras 78 viviendas no se especificó al tipo de vivienda que son.



VIVIENDA	Recuento VIVIENDAS
Casa independiente	65389
Departamento en edificio	1869
Local no construido para habitacion	472
No especificado	78
Vivienda en cuarto de azotea	53
Vivienda en vecindad	1308

Figura 11. Consulta sobre el tipo y número de vivienda que existen en Analysis Services.



Etiquetas de fila	Recuento de FOLIOVIV
Casa independiente	65389
Departamento en edificio	1869
Local no construido para habitacion	472
No especificado	78
Vivienda en cuarto de azotea	53
Vivienda en vecindad	1308
Total general	69169

Figura 12. Consulta sobre el tipo y número de vivienda que existen en Excel.

4. Conclusiones y Trabajos futuros

Se concluye que las herramientas utilizadas en esta investigación para realizar la construcción de cubos OLAP en Microsoft Analysis Services y Microsoft Excel son herramientas de gran ayuda para la realización del Procesamiento Analítico en Línea (OLAP); ya que se puede visualizar los datos desde diferentes perspectivas con las dimensiones y medidas que se crean con el cubo OLAP, porque se crean diferentes combinaciones de consultas con la información que contiene la base de datos y con ayuda de esta información extraída por medio de las consultas realizadas al cubo OLAP se pueden tomar decisiones en una organización o institución educativa o empresarial, así como en cualquier ámbito que sea necesario utilizar el procesamiento OLAP, ya que hay datos o información que a simple vista o con una consulta sencilla no se podrán visualizar, en cambio con este tipo de tratamiento de los datos si se puede realizar. Además estas herramientas son compatibles, se puede hacer una conexión de datos desde Excel a Analysis Services para trabajar directamente con el cubo creado en esta herramienta y en Excel realizar las consultas a través de tablas dinámicas o gráficos dinámicos y las dos herramientas pueden crear una conexión con el gestor de base de datos Microsoft SQL Server.

En trabajos futuros, se realizará una comparativa entre las herramientas de Analysis Services y Excel para construir e implementar cubos OLAP. También se realizarán comparativas entre otras herramientas de suite de Oracle y Microsoft u otro gestor de base de datos para la creación de cubos OLAP.

5. Referencias

- [1] Turban, E.; Sharda, R., Delen, D. (2011). *Decision support and business intelligence systems* (9na. ed.). Upper Saddle River, NJ: Prentice Hall.
- [2] Roldán Salgueiro, J. L., Cepeda Carrión, G., Galán González, J. L. (2012). Los sistemas de inteligencia de negocios como soporte a los procesos de toma de decisiones en las organizaciones. *Papeles de Economía Española*, 132, 239-260. Recuperado de: <https://idus.us.es/xmlui/handle/11441/76099>
- [3] Codd, E. F., Codd, S. B., Salley, C. T. (1993). *Providing OLAP to user-analysts: An IT mandate*. Toronto, Canada: E. F. Codd & Associates.
- [4] Kimball R., Reeves L., Ross M., Thornthwaite W. (1998). *The Data Warehouse Lifecycle Toolkit: expert Methods for designing, developing and deploying data warehouses*. New York, NY: John Wiley & Sons, Inc.
- [5] Koperski K., Han J., Stefanovic N. (2001). *An Efficient two-step method for classification of Spatial Data*. Recuperado de: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.12.2505>
- [6] Abril Frade, D. O., Pérez Castillo, J. N. (2007). Estado actual de las tecnologías de bodega de datos y OLAP aplicadas a base de datos espaciales. *Revista Ingeniería e Investigación*, 27 (1), 58-67. Recuperado de http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-56092007000100008
- [7] Bédard, Y., Merrett, T., Han, J. (2001). Fundamentals of spatial data warehousing for geographic knowledge discovery. En H. Miller and J. Han (Eds.), *Geographic data mining and knowledge discovery* (pp. 53-73). London: CRC Press.
- [8] Microsoft. (2017). *Analysis Services*. Recuperado de: <https://docs.microsoft.com/es-es/sql/analysis-services/analysis-services?view=sql-server-2014>
- [9] Carballeiro, G. (2013). *Excel 2013: guía práctica para el usuario*. Recuperado de: <https://www.tesuva.edu.co/phocadownloadpap/Guia%20basica%20excel%202013.pdf>
- [10] Hernández, R., Fernández, C., Baptista, P. (2014). *Metodología de la investigación* (6ta. ed.). México: McGraw-Hill.
- [11] Pressman, R. (2010). *Ingeniería del software: Un enfoque práctico* (7ma. Ed.). México: McGraw-Hill.
- [12] Instituto Nacional de Estadística y Geografía [INEGI]. (2017). Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH), 2016 nueva serie. Recuperado de: <https://www.inegi.org.mx/programas/enigh/nc/2016/>