



Identificación de los elementos necesarios para la elaboración de cubos OLAP utilizando Microsoft Analysis Services con SQL Server y Excel

Identification of the necessary elements for the elaboration of OLAP cubes using Microsoft Analysis Services with SQL Server and Excel

Zully Kristel Guzmán Caraveo¹ Gilberto Murillo González² Herman Aguilar Mayo³

Para citar este artículo: Z.K. Guzman Caraveo, G. Murillo González y H. Aguilar Mayo, "Identificación de los elementos necesarios para la elaboración de Cubos OLAP utilizando Microsoft Analysis Services con SQL SERVER y Excel". *Revista Vínculos*, vol. 17, n°. 1, pp. 32-41, enero-junio. 2020. DOI: <https://doi.org/10.14483/2322939X.15573>

Recibido: 20-11-2019 / Aprobado: 24-04-2020.

Resumen

El trabajo aquí presentado es resultado de un análisis comparativo entre dos herramientas para la creación de un cubo OLAP, el cual es un procesamiento analítico en línea que ejecuta un análisis de datos. Una de las herramientas es Microsoft Analysis Services con SQL Server y la otra Microsoft Excel; para la definición del universo de datos se utilizó una base de datos experimental obtenida de una encuesta realizada por el Instituto Nacional de Estadística, Geografía e Informática (INEGI) del 2016. Con estas herramientas existentes se busca proporcionar un análisis para el mejor uso en la explotación de datos, identificando los diversos factores necesarios para la creación, desarrollo e implementación de los cubos OLAP y entre esos se encuentran identificar algunos patrones de los riesgos o errores al ser creados por parte de estas dos soluciones existentes en el mercado, los cuales se describirán en el presente trabajo.

Palabras clave: base de datos, cubos OLAP, factores de riesgos, seguridad.

Abstract

The work presented here is the result of a comparative analysis in the creation of an OLAP cube, which is an Online Analytical Processing, which executes a data analysis; which was developed using two tools with one is Microsoft Analysis Services with SQL Server and the other using Microsoft Excel; For the definition of the data universe, an experimental database was used, obtained from a survey carried out by INEGI in 2016; With these tools, there is an attempt to provide an analysis for the best use in data exploitation, identifying the various factors necessary for its creation, development and implementation of an OLAP cube and among those are found to identify some patterns of risks or errors to be created by some of these two existing solutions in the market, which will be described in the present work.

Keywords: database, OLAP cubes, risk factors, security.

- 1 Licenciada en Sistemas Computacionales, Universidad Juárez Autónoma de Tabasco, México. Correo electrónico: zully_k1996@hotmail.com. ORCID: <https://orcid.org/0000-0001-6234-0981>
- 2 Doctor en Sistemas Computacionales, Universidad Juárez Autónoma de Tabasco, México. Correo electrónico: gmurillo76@gmail.com. ORCID: <https://orcid.org/0000-0003-1206-846X>
- 3 Doctor en Sistemas Computacionales: Universidad Juárez Autónoma de Tabasco, México. Correo electrónico: herman.aguilar@ujat.mx. ORCID: <https://orcid.org/0000-0003-1077-3210>

1. Introducción

El procesamiento analítico en línea (OLAP, por sus siglas en inglés), 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, además de permitir 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]. Al respecto, vale la pena considerar:

- Las medidas son los valores de un dato.
- Las dimensiones son las descripciones de las características que definen dicho dato.
- Los hechos corresponden a la existencia de valores específicos de una o más medidas para una combinación particular de dimensiones.

En 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 se muestra una estructura para la navegación en las dimensiones y medidas de los cubos OLAP, la

cual proporciona de manera general el uso de la dimensión y la acción resultante de su ejecución.

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

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 pueden 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; estos componentes permiten conectar Excel con datos externos, desde una base de datos hasta servidores en línea. Además de lo anterior, 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 lo mencionado que en el presente trabajo se presentará la construcción de un cubo OLAP utilizando Analysis Services y Excel para descubrir los factores más importantes en su creación y la identificación de algunos de los riesgos de seguridad que pueden existir en el momento de su ejecución o implementación.

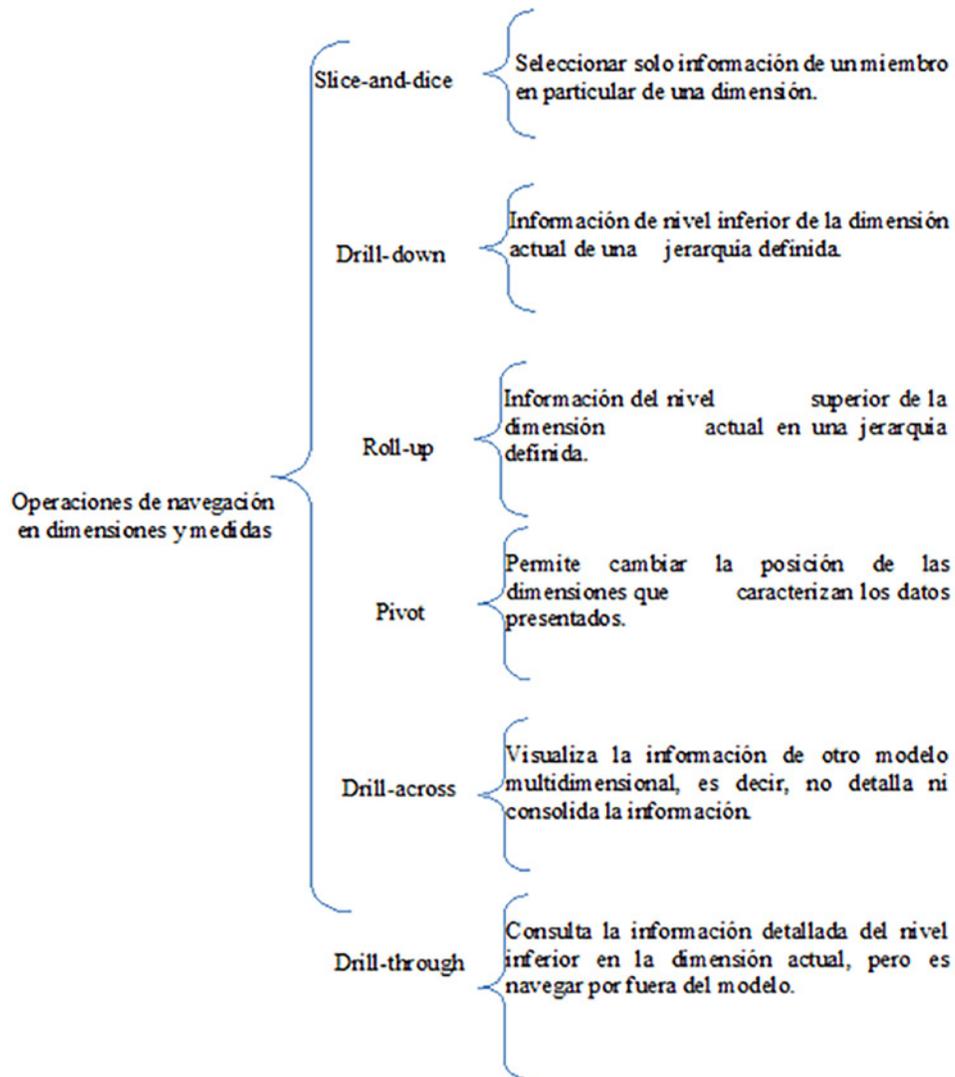


Figura 1. Operaciones de navegación en las dimensiones y agrupaciones de medidas en OLAP

Fuente: [7].

2. Metodología

El enfoque de estudio de esta investigación es de tipo cualitativo. Así, según [10], se guía por áreas o temas significativos de investigación, además pueden desarrollarse preguntas o supuestos antes, durante o después de la recolección y análisis de los datos; por ello, 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 utilizado para la presente investigación fue el modelo cascada, ya que con este modelo es posible identificar desde el principio todos los elementos o etapas que se desarrollarán durante la investigación. El modelo de cascada se describe como un ciclo de vida, el cual sugiere un enfoque sistemático y secuencial para el desarrollo del *software*. 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).

La etapa de modelado se dividió en otras dos: análisis de la base de datos y el diseño. En la etapa del análisis de la base de datos se fue realizando una revisión con más profundidad referente a la estructura de los datos y campos de la base de datos, la cual se utilizó para la construcción del cubo. Se encontró que las tablas contienen información asociada a tres niveles o grupos de estudio y la tabla

de Concentradohogar. En la Tablas 1, la Tabla 2 y la Tabla 3 se describen los grupos y la información de las tablas [12].

En la etapa de diseño, de acuerdo con los tres grupos anteriormente descritos, se decidió realizar tres diseños de cubos para tener una mejor visualización de los datos de cada tabla (Figuras 4, Figura 5 y Figura 6).

Tabla 1. Descripción del grupo información de las viviendas

Tabla con información de las viviendas
<i>Viviendas.</i> 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 las viviendas.

Fuente: [12].

Tabla 2. Descripción del grupo información de los hogares

Tabla con información de los hogares
<i>Hogares.</i> 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.
<i>Gastos hogar.</i> Contiene los gastos monetarios y no monetarios que realizó el hogar en el periodo de referencia, así como la estimación del alquiler que fueron captados en el cuestionario de gastos del hogar, en el cuadernillo de gastos diarios y en el cuestionario de hogares y vivienda.
<i>Erogaciones.</i> Erogaciones financieras y de capital realizadas por los hogares.
<i>Gasto tarjetas.</i> Gastos realizados por el hogar y que fueron cubiertos mediante alguna tarjeta de crédito bancaria o comercial.

Fuente: [12].

Tabla 3. Descripción del grupo información de las personas

Tablas con información de las personas
<i>Población.</i> Características sociodemográficas y ocupacionales de los integrantes del hogar.
<i>Ingresos.</i> Ingresos y percepciones de capital de cada uno de los integrantes del hogar.
<i>Gastos personas.</i> Contiene los gastos monetarios y no monetarios asociados a los integrantes del hogar en el periodo de referencia.
<i>Trabajos.</i> Condición de actividad de los integrantes del hogar de 12 o más años.
<i>Agro.</i> 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.
<i>No agro.</i> Ingresos y gastos de los negocios del hogar dedicados a las actividades industriales, comerciales y de servicios, y sus características propias.

Fuente: [12].

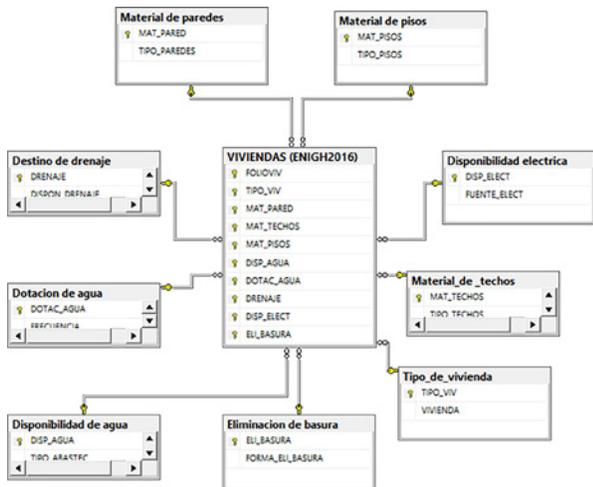


Figura 4. Diagrama del grupo vivienda
Fuente: elaboración propia.

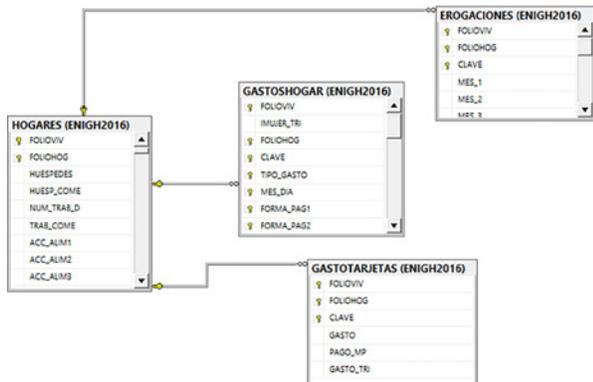


Figura 5. Diagrama del grupo de información de hogares
Fuente: elaboración propia.

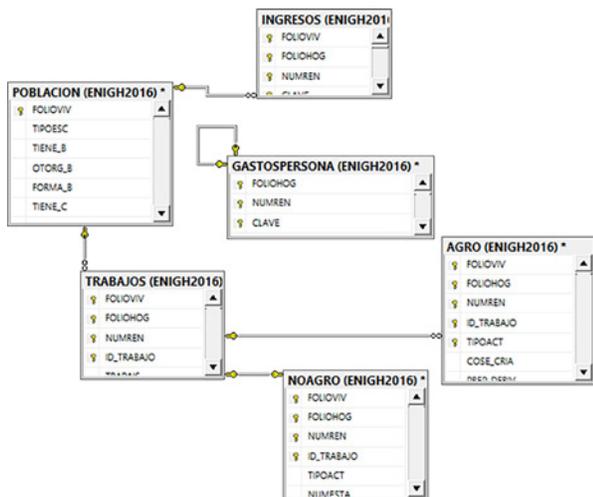


Figura 6. Diagrama del grupo de información por persona
Fuente: elaboración propia.

La etapa de construcción se llevó a cabo estableciendo el desarrollo arquitectónico del cubo para los tres diseños ya explicados anteriormente. El desarrollo del cubo OLAP se realizó primero en Analysis Services utilizando el entorno de desarrollo de Visual Studio, ya que este utiliza estos complementos para la creación de cubos OLAP (Figura 7). Una vez realizado el cubo en Analysis Services, se procedió a realizar las estructuras necesarias para crear los cubos en Excel, esto fue por medio de la conexión al gestor de base de datos Microsoft SQL Server y la conexión a Analysis Services (Figura 8).

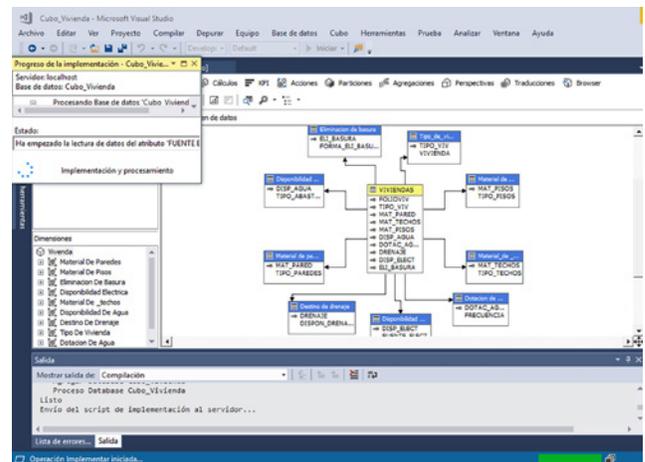


Figura 7. Implementación del cubo en Analysis Services y Business Intelligence Development Studio
Fuente: elaboración propia.

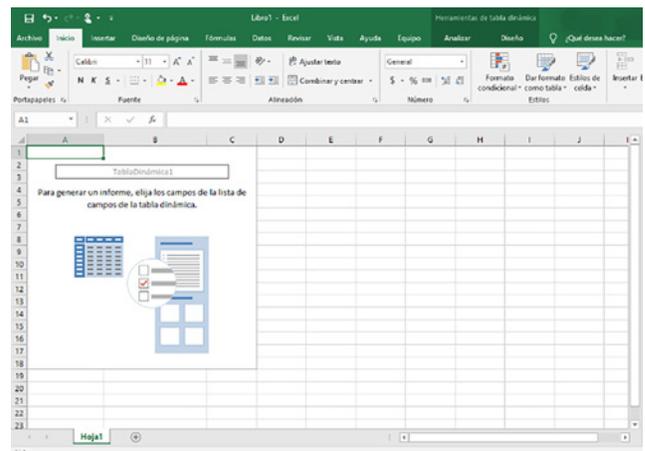


Figura 8. Creación de la tabla dinámica para construir la consulta del cubo en Excel
Fuente: elaboración propia.

La última etapa fue el despliegue. En esta se pudo visualizar los datos, medidas y dimensiones del cubo para así poder ver las diversas combinaciones posibles que se pueden obtener con los cubos OLAP y la base de datos utilizada.

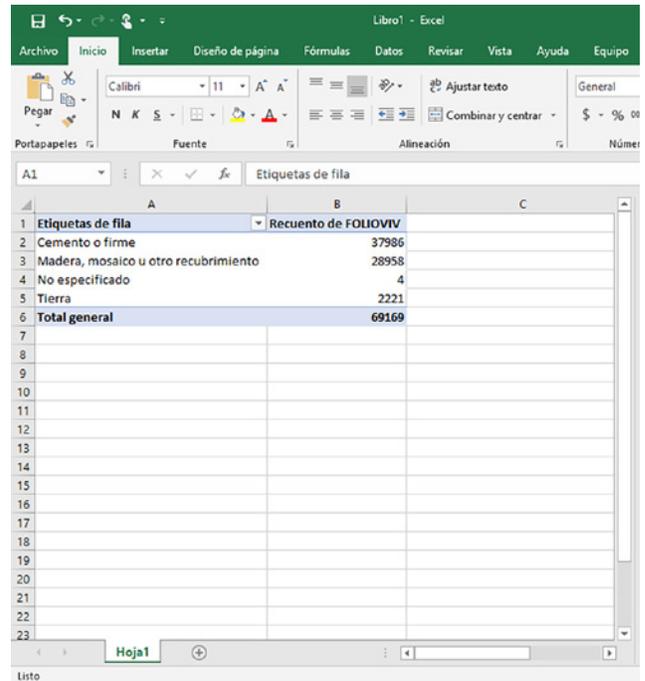
3. Resultados

Con la creación e implementación del cubo se pueden elaborar 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 numerosas. Tanto con Analysis Services como con Excel se pueden analizar y tener los mismos resultados, ya que se creó el mismo diseño del cubo para utilizarlo en ambas herramientas. En Analysis Services se hacen las consultas en el navegador que el mismo programa proporciona, mientras que en Excel se inserta una tabla dinámica para poder crear las consultas. Algunos de los resultados obtenidos con la creación de los cubos en ambas herramientas se mencionan a continuación.

En la Figuras 9 y la Figura 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 especificados.

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 obtienen cuatro datos que no especificaron el tipo de pisos que tienen en sus viviendas.

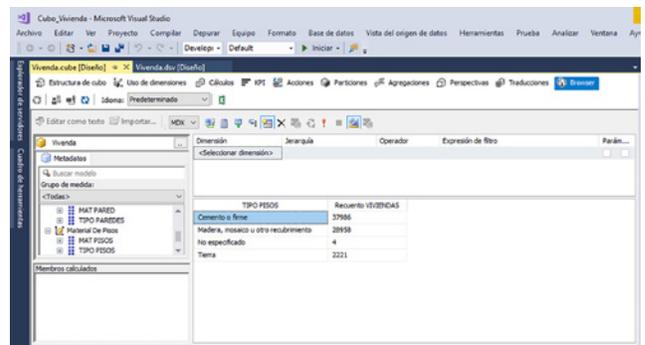
Otra consulta que se realizó fue en el segundo diseño de cubo. En este se utilizaron cuatro medidas y atributos, los cuales son: celular, conexión a internet, teléfono y TV de paga. En la dimensión hogares se aprecia que los datos que contienen estos atributos son 1 y 2, los cuales se describen de la siguiente



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

Fuente: elaboración propia.



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

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

Fuente: elaboración propia.

forma, 1 indica una respuesta “Sí” y 2 indica una respuesta “No”.

La Figuras 11 y la Figura 12 muestran los resultados de esta consulta, donde se aprecia que en 643 754 hogares existen gastos de celular, conexión a internet, teléfono fijo y TV de paga; a su vez, en 9548 hogares hubo erogaciones, es decir, se realizó un gasto por los servicios con los que cuenta el hogar

y 4841 hogares pagaron con tarjeta estos gastos. Sin embargo, así como hay hogares que cuentan con estos servicios, hay otros que solo cuentan con uno, con dos o con tres, pero también hogares que no cuentan con ningún servicio de estos, es el caso de 290 045 hogares que no gastan y ni cuentan con estos servicios, también 1853 hogares en los que no hubo erogaciones (un gasto), así como 38 hogares que no gastan en pagar estos servicios con tarjeta.

	CELULAR	CONEX INTE	TELEFONO	TV PAGA	Recuento GASTOSHOGAR	Recuento EROGACIONES	Recuento GASTOTARJETAS
1	1	1	1	1	442794	9548	4841
1	1	1	2	2	225726	2981	1103
1	1	2	1	1	300171	4099	1086
1	1	2	2	2	98507	1195	361
1	2	1	1	1	107383	1170	169
1	2	1	2	2	102874	973	175
1	2	2	1	1	916571	9301	954
1	2	2	2	2	1202387	10749	1004
2	1	1	1	1	45288	435	98
2	1	1	2	2	27890	219	31

Figura 11. Hogares donde se cuenta con servicio de telefonía, internet y TV de paga en Analysis Services

Fuente: elaboración propia.

3.1. Identificación de algunos factores de riesgos, errores y seguridad

Uno de los principales problemas que se presenta durante la utilización de Microsoft SQL Server es el Error de conexión con el servidor, el cual se define como una problemática intermedia debido a que su posible origen puede estar establecido por diversos factores y entornos (configuración del tipo de sistema operativo de Microsoft en sus distintas versiones donde sea factible la instalación). Uno de esos factores se identifica en la ejecución del servicio principal del SQL Server, el cual se encuentra en un estado no iniciado.

Otro de los problemas identificados en la creación de los cubos se encuentra relacionado con los elementos de seguridad de las herramientas utilizadas. Para el caso específico con el sistema manejador de base de datos (SMBD), los elementos de vulnerabilidad se encuentran en el uso de las credenciales de acceso; así, de acuerdo con la naturaleza del SQL Server y el sistema operativo Microsoft Windows, en muchas ocasiones no se logra identificar el modo de autenticación entre estos dos elementos, lo cual será necesario para ingresar a los servicios del Analysis Server.

	CELULAR	CONEX INTE	TELEFONO	TV PAGA	Recuento GASTOSHOGAR	Recuento EROGACIONES	Recuento GASTOTARJETAS
1	1	1	1	1	443754	9548	4841
1	1	1	2	2	225726	2981	1103
1	1	2	1	1	300171	4099	1086
1	1	2	2	2	98507	1195	361
1	2	1	1	1	107383	1170	169
1	2	1	2	2	102874	973	175
1	2	2	1	1	916571	9301	954
1	2	2	2	2	1202387	10749	1004
2	1	1	1	1	45288	435	98
2	1	1	2	2	27890	219	31
2	1	2	1	1	4960	53	14

Figura 12. Hogares donde se cuenta con servicio de telefonía, internet y TV de paga en Excel

Fuente: elaboración propia.

En lo relacionado con Microsoft Excel, uno de los factores de error identificado es el de la comunicación con el SMBD, en este caso SQL Server derivado por el desconocimiento del tipo de credenciales a utilizar.

Al momento de realizar un cubo OLAP hay que tener en cuenta diferentes factores que pueden ocurrir durante el diseño, construcción, implementación y visualización de los datos; en cada caso los problemas de seguridad, riesgo o errores pueden ser diferentes. Realizar este tipo de procesamiento es muy útil, ya que se está realizando una explotación de datos sobre los que la institución u organización no están enterada, permitiéndoles establecer

nuevas estrategias a partir de su información bajo herramientas muy profesionales o a través de aplicaciones básicas existentes dentro de los entornos de organizaciones como lo es la Suite Ofimática.

4. Conclusiones

Es importante destacar que, con el crecimiento acelerado de la tecnología, hoy en día la utilización de técnicas, herramientas y diversos recursos informáticos resulta de alto impacto en esta sociedad de consumo, pues da mayor flexibilidad y acercamiento para el usuario final. En este sentido, el uso de los cubos OLAP también se han flexibilizado, lo que viene a establecer una mejora considerable a las dos dimensiones de una hoja de cálculo y, de igual forma, vislumbra las limitantes existentes en las bases de datos relacionales, las cuales no son adecuadas a los análisis instantáneos de grandes volúmenes de datos.

Partiendo de este crecimiento tecnológico, en la actualidad crear e implementar un cubo OLAP es más “fácil” desde la óptica de la actualización y mejoras del *software*, así como la creación de diversas herramientas que permitan realizar las actividades de explotación de la información. La flexibilidad en las nuevas herramientas tecnológicas creadas para establecer análisis de crecimiento y explotación de datos permiten utilizar grandes volúmenes, potencializando su uso en los diversos ámbitos y, al mismo tiempo, permite a cada universo de datos ser personalizado y extensible a través de herramientas tales como: Microsoft Analysis Services con SQL Server y Excel.

Con el uso de Microsoft Analysis Services con SQL Server y Excel en su versión 2003, 2007, 2010, 2013 y 2016, se pueden realizar análisis a la información debido a que la base de datos que las instituciones utilizan en su gran mayoría utiliza Microsoft Excel para almacenar sus transacciones. Suelen tener registros con valores unitarios, los cuales no están preparados para realizar un correcto y pleno análisis

de su información, generando con ello un posible descontrol en las estrategias planteadas por las organizaciones, permitiendo a los cubos OLAP ser un verdadero aliado para realizar esta tarea. Esta razón se ve reflejada en el desarrollo de las actividades en la presente investigación, en la cual se identifica las bondades en el uso del *software* de reciente generación, la utilización de herramientas como Microsoft Analysis Service con SQL Server y Excel 2013, mismas que permiten brindarle al usuario final y a las organizaciones pequeñas, medianas y grandes el uso de estas técnicas que fueron siempre conceptualizadas para empresas grandes y hoy en día están más cercas para su manipulación, uso y explotación a segmentos empresariales de diversos tamaños.

Los cubos OLAP buscan reducir el tiempo y el esfuerzo en la obtención de respuestas debido a que su naturaleza es procesar grandes volúmenes de datos existentes en los bancos de datos de las organizaciones. Entre las características más claras en su uso se logran visualizar las ventajas sobre los modelos relacionales ordinarios, sobre todo al momento de hablar de reportes rápidos y seguros, los cuales permiten establecer acciones como segmentar, profundizar, sintetizar, rotar y filtrar los datos con el fin de obtener la información necesaria para su análisis pertinente.

Referencias

- [1] E. Turban, R. Sharda y D. Delen, *Decision support and business intelligence systems*. Nueva Jersey: Prentice Hall, 2011.
- [2] J. L. Roldán, G. Cepeda y J. L. Galán, “Los sistemas de inteligencia de negocios como soporte a los procesos de toma de decisiones en las organizaciones”, *Papeles de Economía Española*, n.º 132, pp. 239-260, 2012. <https://hdl.handle.net/11441/76099>
- [3] E. F. Codd, S. B. Codd y C. T. Salley, “Providing OLAP to user-analysts: An IT mandate”, E. F. Codd and Associates, p. 1003, 1993.

- [4] R. Kimball, L. Reeves, M. Ross y W. Thornthwaite, *The Data Warehouse Lifecycle Toolkit: expert Methods for designing, developing and deploying data warehouses*. Nueva Jersey: John Wiley and Sons, 1998.
- [5] K. Koperski, J. Han y N. Stefanovic, "An Efficient two-step method for classification of Spatial Data", en *Symposium on spatial Data Handling*, agosto 1998.
- [6] D. Orlando, A. Frade y J. Pérez, "Estado actual de las tecnologías de bodega de datos y OLAP aplicadas a base de datos espaciales", *Ingeniería e Investigación*, vol. 27, n.º 1, pp. 58-67, 2007. <https://repositorio.unal.edu.co/handle/unal/28796>
- [7] Y. Bédard, T. Merrett y J. Han, "Fundamentals of spatial data warehousing for geographic knowledge Discovery", en H. Miller and J. Han (Editors), *Geographic data mining and Knowledge discovery*. Londres: CRC Press, 2001.
- [8] Microsoft, "Analysis Services". [En línea]. Disponible en: <https://docs.microsoft.com/es-es/sql/analysis-services/analysis-services?view=sql-server-2014>
- [9] G. Carballeiro, "Excel 2013: guía práctica para el usuario". [En línea]. Disponible en: <https://www.tesuva.edu.co/phocadownloadpap/Guia%20basica%20excel%202013.pdf>
- [10] R. Hernández, C. Fernández y P. Baptista, *Metodología de la investigación*. Ciudad de México: McGraw Hill, 2014.
- [11] R. Pressman, *Ingeniería del software: un enfoque práctico*. Ciudad de México: McGraw Hill, 2010.
- [12] Instituto Nacional de Estadística, Geografía e Informática [INEGI], "Encuesta Nacional de Ingresos y Gastos de los Hogares 2016: ENIGH: nueva serie: descripción de la base de datos / Instituto Nacional de Estadística y Geografía". [En línea]. Disponible en: <https://www.inegi.org.mx/programas/enigh/nc/2016/>
- [13] Trucos y Cursos, "El Cubo de OLAP en Excel". [En línea]. Disponible en: <http://trucosycursos.es/el-cubo-de-olap-en-excel/>
- [14] Juan2510, "OLAP en Excel". [En línea]. Disponible en: <https://juan2510.wordpress.com/2010/05/10/olap-en-excel/>
- [15] Microsoft, "Analizar los datos del cubo OLAP con Excel". [En línea]. Disponible en: <https://docs.microsoft.com/es-es/system-center/scsm/olap-cube-excel?view=sc-sm-2019>
- [16] S. Simon, "Procedimientos Recomendados para SQL Server Reporting Services". [En línea]. Disponible en: <https://www.sqlshack.com/es/procedimientos-recomendados-para-sql-server-reporting-services/>

