

# Modelación Tabular: una alternativa sugerente para el análisis de los datos

Maité Torres Sánchez  
Yissel Espinosa Cervantes  
Alina Simón Cuevas  
Lucina García Hernández  
Alfredo J. Simón Cuevas

*Desde el siglo pasado se ha investigado en aras de incrementar la eficiencia en el almacenamiento y el acceso a las bases de datos analíticas, sobre cuyos resultados las grandes compañías han introducido productos comerciales. En este escenario, Microsoft SQL Server 2012 ofrece dos opciones independientes para la creación de los modelos analíticos, el modelo multidimensional y el reciente modelo tabular. En éste trabajo se profundiza en las características y potencialidades de cada uno, proponiendo los criterios más importantes que, a juicio de los autores, se deben tener en cuenta al emprender un nuevo proyecto. Se propone además una solución computacional que brinda a los especialistas y ejecutivos tanto visiones particulares como integradoras del estado del negocio, aprovechándose las facilidades recientes que proporciona la plataforma de Inteligencia de Negocios de Microsoft para la implementación de ambos modelos, multidimensional y tabular. El Grupo Empresarial CIMEX ha sido el escenario donde se ha desarrollado la propuesta de solución. Para comprobar la validez de la aproximación se llevó a cabo cuatro experimentos.*

*Palabras clave: inteligencia de negocios, modelo tabular, modelo multidimensional, bases de datos columnares, bases de datos en memoria.*

## RESUMEN

## ABSTRACT

*Since last century, multiple researches have been made with the goal of incrementing the efficiency in the storage and access to analytical databases, upon which results large companies have introduced commercial products. In this aspect, Microsoft SQL Server 2012 offers two independent options for the creation of analytical models, the multidimensional model and the more recent tabular model. This work looks deeply into the features and potentialities of each model, proposing the most important points that, in the authors' judgment, must be taken into account when undertaking a new project. A computational solution that provides both specialists and executives a particular as well as inclusive vision of the state of the business is proposed, taking advantage of the recent facilities provided by the Microsoft Business Intelligence platform for implementing the models, both multidimensional and tabular. CIMEX Enterprise Group has been the stage where the solution proposal has been developed. For checking the approximation validity four experiments were performed.*

*Keywords: business intelligence, tabular model, multidimensional model, columnar databases, in-memory databases*

## Introducción

Los avances tecnológicos de los últimos seis años han provocado una gran revolución informacional, incrementando la disponibilidad y las

posibilidades de acceso a la información.

A medida que aumenta la cantidad de datos acumulados, se incrementan también las necesidades de consultas más complejas para la toma de decisiones dentro de las

dentro de las organizaciones. La mayoría de los sistemas de gestión de bases de datos que ofrecen las herramientas para realizar el proceso de *data warehousing* se apoyan en la tecnología orientada a filas/registros, optimizada para el procesamiento

transaccional de los datos (Albano A. et al., 2006; Boateng, Singh, Greeshma, & Singh, 2012). Con el desarrollo del modelo multidimensional y diferentes alternativas de indexación no es posible eludir por completo el compromiso de almacenamiento por filas (*row-oriented*).

Durante los últimos 20 años, diferentes autores han defendido la contribución del almacenamiento columnar, que se basa esencialmente en la transposición de los ficheros para mejorar el desempeño de las consultas enfocadas hacia el análisis y la toma de decisiones (Albano A. et al., 2006). Se trata de beneficiar el procesamiento analítico de los datos, caracterizado por demandas que requieren el agrupamiento o la agregación de grandes cantidades de datos sobre unas pocas columnas, desde la perspectiva de los índices de proyección a través de las filas (*column-oriented*) (M., D., A, & X., 2005). Sin embargo, mientras el almacenamiento basado en discos constituyó la única opción para las bases de datos, no fue posible obtener soluciones comerciales (Albano A. et al., 2006).

Las tecnologías han ido evolucionado ostensiblemente gracias al desarrollo del hardware, por lo que actualmente es factible aprovechar las nuevas técnicas de bases de datos en memoria (*in-memory databases*) y el almacenamiento columnar para la optimización de las consultas en soluciones analíticas. De ahí que las grandes compañías también han invertido en investigaciones dirigidas a incrementar la eficiencia en el almacenamiento y el acceso a las bases de datos analíticas, cuyos resultados han derivado en productos comerciales. En este escenario, *Microsoft SQL Server 2012* ofrece dos opciones independientes para la creación de los modelos analíticos que representan la lógica del negocio, el clásico modelo multidimensional y el reciente modelo tabular (Russo, Ferrari, & Webb, 2012). El nuevo modelo tabular no es el remplazo del multidimensional, ellos no son más que dos técnicas de modelación sobre la misma tecnología, y ambas son el objeto de estudio de este trabajo (Russo et al., 2012).

La nueva propuesta de Microsoft resultó muy atractiva en la concepción y desarrollo de soluciones analíticas y constituyó una importante motivación durante la presente investigación que nos llevó a profundizar en por qué se propone un nuevo modelo de análisis de datos cuando ya existía el modelo

multidimensional con más de una década de explotación, cuáles son las ventajas que ofrece con respecto a su precedente, en qué contextos se debe utilizar uno u otro, o bien si ambos son necesarios. El modelo tabular ofrece un enfoque mucho más cercano a los usuarios finales, quienes generalmente están familiarizados con el modelo relacional y cuyos intereses fundamentales radican en encontrar soluciones a los problemas del negocio mediante herramientas amigables que les asegure evadir el uso de recursos matemático-computacionales complejos, considerando además que el modelo multidimensional siempre ha sido percibido como una alternativa compleja de comprender por parte de los usuarios finales. Dado que el trabajo con las tablas es más natural e intuitivo para los analistas, la concepción y la instrumentación del modelo tabular han estado estrechamente vinculadas con alcanzar la eficacia y la eficiencia en las facilidades de exploración y visualización personalizada de los datos, aprovechando los avances científicos y tecnológicos. De ahí, que el nuevo modelo se haya convertido en una alternativa de consideración en el marco de la toma de decisiones, especialmente en cuanto a la potenciación de las funcionalidades de «autoservicio», que han sido acogidas con gran aceptación por los usuarios finales, ya que les permite diversificar las tareas de análisis de los datos sin tener que depender frecuentemente de los profesionales de la computación (Russo et al., 2012). Esta ha sido precisamente una de las dificultades que en los últimos años ha limitado el desarrollo informático en las empresas, de ahí la necesidad de lograr una propuesta de solución que hoy responda al dinamismo en el análisis de la información que los usuarios requieren pero que, a la vez, se sustente en una elección fundamentada de los modelos de datos a utilizar.

El mundo empresarial en Cuba no está exento de esta búsqueda, por el contrario, los sistemas de información cobran relevancia para una acertada conducción de cualquier negocio y el Grupo Empresarial CIMEX es un ejemplo de ello. Esta organización, líder en el mercado comercial, tiene como principal objetivo la adquisición y la comercialización de productos y servicios. El desarrollo informático de la empresa no solo ha facilitado dar respuestas oportunas y pertinentes al procesamiento de los datos operacionales, sino también ha posibilitado la evolución de los procesos

de dirección con el empleo de herramientas y ambientes computacionales propios, los cuales constituyen antecedentes de esta investigación.

En el contexto de la investigación desarrollada se concibió y diseñó una solución informática que implementa los modelos multidimensional y tabular sobre *Microsoft SQL Server 2012 Analysis Services (SSAS)*, la cual sirvió de entorno práctico para el análisis de ambos modelos. A partir del estudio realizado, de la experiencia adquirida en el desarrollo del sistema y del comportamiento del resultado de los experimentos, se incurrió en la comparación práctica de los mismos y se esbozaron consideraciones generales en cuanto a sus fortalezas y debilidades, como contribución al trabajo ulterior en esta área del conocimiento. Esta propuesta constituye una continuación de la investigación reportada en las memorias del Congreso Internacional de Matemática y Computación COMPUMAT 2013 (García, Simón, Torres, & Espinosa, 2013), en la cual se logró un primer acercamiento.

La propuesta de solución matemático-computacional está sustentada sobre el paradigma de la Inteligencia de Negocios (B. & N., 2008), lo que ha permitido garantizar el análisis de los principales indicadores comerciales y económico-financieros de la empresa. Asimismo, proporciona un ambiente de consultas dinámicas con funcionalidades de autoservicio, siguiendo una de las más importantes tendencias de la industria de la Inteligencia de Negocios en los últimos años, de manera que sea posible mantener experiencias interactivas con los datos en forma independiente (Russo et al., 2012).

## Materiales y métodos

Los encargados de tomar decisiones reconocen que hoy día es imposible actuar basándose solo en la intuición para hacer crecer su negocio o para permanecer exitosamente en el mercado. Vinculado a esta premisa, ha evolucionado un conjunto de conceptos, modelos y tecnologías con el transcurso de los años cuya interacción facilita el éxito.

La Inteligencia de Negocios (BI, del inglés *Business Intelligence*) se puede definir como un conjunto de tecnologías, metodologías,

arquitecturas y procesos, que transforman los datos en información útil e importante que posibilita ideas estratégicas, tácticas y operativas más eficaces para la toma de decisiones (B. & N., 2008). Se materializa cuando se ofrecen herramientas y políticas organizacionales a nivel empresarial que permiten a los directivos transformar la información estratégica de su empresa en acciones concretas que se traduzcan en beneficios palpables. Se ha convertido actualmente, en un modelo de control y crecimiento corporativo en pos de lograr competitividad. Una frase popular acerca de la Inteligencia de Negocios plantea: «Inteligencia de Negocios es el proceso de convertir datos en conocimiento y el conocimiento en acción para la toma de decisiones» (Bernabeu, 2010). Resulta más pertinente hablar de soluciones o sistemas de Inteligencia de Negocios como aproximaciones sucesivas, puesto que no existe un modelo único para su desarrollo, dado el alcance y complejidad del proceso (García Hernández, Oliva Santos, Prendes Arencibia, Velázquez Vidal, & Véliz Montegudo, 2008). En esta línea, muchas compañías de software, han ofrecido a las empresas un producto completo al producir plataformas que integran varias herramientas, que respondan a las diferentes etapas del proceso de BI, a partir del cual los equipos de desarrollo pueden generar con mayor holgura y productividad las aproximaciones de soluciones BI propias. La plataforma de Inteligencia de Negocios de Microsoft ha sido seleccionada para la presente investigación por las facilidades que posee, su utilización en innumerables soluciones computacionales a nivel mundial y en CIMEX como caso particular, donde se cuenta con más de 8 años de experiencias.

En función del procesamiento analítico son múltiples las propuestas existentes en la representación de los datos, que requieren estructuras que posibiliten agilizar el tratamiento de grandes volúmenes de datos en términos de operaciones de resúmenes y/o comparaciones. El clásico Modelo Multidimensional es un ejemplo de estos modelos, el cual propone una estructura multidimensional de la información, afín con los sujetos de análisis del negocio, empleando dos componentes básicos: los hechos y las dimensiones. Al trabajar con la multidimensionalidad es importante analizar los datos desde perspectivas diferentes, con el objetivo de lograr una

permita fundamentar las decisiones en diferentes circunstancias, tanto objetivas y subjetivas, con una incidencia significativa de la temporalidad. Sin embargo, la eficacia del análisis multidimensional depende de la manera en que los datos se representen y se almacenen (Velázquez-Vidal, 2009).

En *SQL Server 2008 R2* se incorporó una nueva herramienta llamada *Power Pivot*, también conocido como tablas dinámicas, como un complemento (o *plug-in*) de *Microsoft Excel 2010*. *Power Pivot* está dirigido al análisis de información, del cual su principal particularidad radica en la posibilidad de trabajar con cantidades masivas de datos utilizando *Excel* como interfaz de usuario, por lo que se convierte en una herramienta atractiva para los analistas, dada la popularidad de este último («Power Pivot Add-in,» 2010). *Power Pivot* utiliza un motor analítico conocido como *VertiPaq*, que posibilita procesar millones de registros con bajos tiempos de respuesta al usuario desde su máquina de escritorio.

En *SQL Server 2012 Analysis Services* se introduce el Modelo Semántico de Inteligencia de Negocios (BISM, del inglés *Business Intelligence Semantic Model*), común a todas las formas de presentación de la información (reportes, monitores, tableros de mando, etc.), para todos los tipos de usuarios y además único para las aplicaciones clientes (P., 2014). BISM es una plataforma de Inteligencia de Negocios unificada que tiene capacidades para presentar la información en dos tipos de modelos: multidimensional (utilizando los conceptos de cubos y dimensiones) y tabular (utilizando los conceptos de tablas y relaciones) (Myers, 2012; Vitt & Cameron, 2012). Esta dualidad motivó de forma significativa, considerar ambas alternativas durante el desarrollo de la nueva solución, interiorizando cada uno de estos enfoques, de modo que no solo respondiera a los desafíos del escenario actual de CIMEX, sino también aprovechara las bondades de ambos y se valorara las insuficiencias respectivas a ser consideradas en futuras investigaciones.

El modelo tabular es una nueva forma de estructurar las bases de datos que, aunque tiene como fundamento las fuentes de datos relaciones y se expresa mediante tablas e interrelaciones, permite instrumentar la

modelo dimensional con mayor efectividad y eficiencia en ciertos casos, como por ejemplo las jerarquías *parent-child* (Russo et al., 2012). La concepción de este modelo, introducida inicialmente con las tablas dinámicas de *Excel (Power Pivot)*, busca mayor acercamiento a los usuarios avanzados brindándoles capacidades de autoservicio, básicamente en la interacción y exploración de los datos, visualización personalizada, entre otros; mientras por otro lado, el modelo multidimensional está más reservado para profesionales informáticos. Es incorrecto afirmar que el modelo tabular responda a limitaciones del multidimensional sino más bien que su concepción se acerca más a los usuarios finales familiarizados con el enfoque relacional. En modelación tabular es posible definir medidas sobre las tablas implicadas e indicadores de desempeño (KPI, del inglés *Key Performance Indicator*). Igualmente se pueden definir perspectivas para crear vistas resumidas del negocio (Lachev, 2012; Vitt & Cameron, 2012).

A partir de la versión *SQL Server 2012 Analysis Services (Tabular)*, el motor de búsquedas *VertiPaq* fue renombrado como el motor de búsqueda analítico en memoria *xVelocity* (del inglés, *xVelocity in-memory analytics engine*), el cual logró un cambio radical en el rendimiento de las consultas analíticas, por a la utilización de técnicas tales como almacenamiento por columnas, compresión de datos, caché en memoria y algoritmos de escaneo y agregación de datos en paralelo (Russo et al., 2012). Con el almacenamiento por columnas cada página de datos contiene valores de una sola columna, en el proceso de indización se conservan los valores repetidos solo una vez y se sustituyen las cadenas de texto y fechas por números enteros, todo lo cual favorece la compresión de los datos (Russo et al., 2012). Por otra parte, las bases de datos *in-memory* utilizan la memoria principal de la máquina para el almacenamiento de los datos. Desde el punto de vista del usuario final, *xVelocity* posibilita rápidos accesos a los datos almacenados en las bases de datos tabulares utilizando las aplicaciones clientes como *Excel* y *Power View*, lo cual se considera una mejora en el rendimiento de las consultas de entre 10 y 100 veces (Leland, 2012). *Power View* constituye una intuitiva herramienta de reportes en la que los usuarios pueden interactuar con las vistas de su negocio

modelo analítico («Power View: Explore, visualize, and present your data,» 2013).

## Los modelos multidimensional y tabular en la solución BI

La actividad comercial de CIMEX con alcance nacional y su extensa red minoristas con más de 1000 tiendas constituye uno de los baluartes de este grupo empresarial y genera diariamente un gran volumen de datos. Por tal motivo, es imprescindible mantener el control de los principales procesos que tienen lugar en cada uno de los establecimientos con el objetivo de brindar información actualizada a los analistas y directivos de la corporación, así como a otras entidades del país. En el escenario comercial se realizan varias operaciones que provocan movimientos de entrada y salida en el inventario relacionadas con los conceptos compra y venta de mercancías, transferencias y ajustes, cuyo comportamiento se analiza a partir de un conjunto de indicadores comerciales. Actualmente se cuenta con un sistema de gestión de información que incluye almacenes de datos operacionales (ODS, del inglés *Operational Data Store*) como repositorio de datos, con detalle diario y frecuente actualización, mediante el cual se logran tener los datos de manera centralizada y consolidada, brindando a los usuarios nuevas funcionalidades y el acceso web a la misma información desde cualquier establecimiento. Los reportes existentes aún están sujetos a esquemas predefinidos con posibilidades limitadas de navegación. Además, se cuenta con un portal web para el apoyo a la toma de decisiones, denominado Sistema de Administración de Negocios (SAN), desarrollado desde el año 2008.

La gran cantidad de las aplicaciones que se han desarrollado responde, directamente a los procesos del negocio y no a los sujetos de análisis. Hasta el momento no había sido posible integrar las informaciones comerciales y contables, así como de otras áreas, ni comprobar la magnitud de la correspondencia entre ellas con el fin de evaluar el funcionamiento dentro de la organización. Tampoco se garantizaba la información histórica que permitiera realizar los análisis retrospectivos ni los prospectivos que contemplaran los cambios efectivos y posibles en el transcurso del tiempo. Esta problemática se estudió en la investigación desarrollada, la cual se orientó

a la concepción, diseño e implementación de una nueva aproximación de solución de Inteligencia de Negocios que permita el análisis informacional integrando los datos de diferentes áreas de CIMEX, valorando contribuciones e inconvenientes del uso de los modelos multidimensional y tabular al respecto. Una de las principales tareas en el desarrollo de esta solución fue identificar los requerimientos informacionales, a partir de entrevistas e intercambios con los usuarios analistas.

Desde el punto de vista informacional, la propuesta de solución de Inteligencia de Negocios se fundamenta en el diseño e implementación de un almacén de datos orientado al análisis, que contiene la información comercial y contable de CIMEX. El ambiente web desarrollado sobre *SharePoint* existente en la empresa se usa para la presentación de los resultados, proporcionando además la navegación por los escenarios de análisis, la confección dinámica de consultas y el enriquecimiento de los efectos visuales. En la Fig. 1 se muestra el modelo general de la solución propuesta.

El almacén de datos se basa en la arquitectura de datos de tres capas propuesta por Devlin (Devlin, 1997) y también conocida como *Enterprise data warehouse* (Ballard, Farrell, Gupta, Mazuela, & Vohnik, 2006; Kimball & Ross, 2002; Véliz Monteagudo, 2009). Se identifican como componentes fundamentales: el *data warehouse* empresarial (DWE), el *data warehouse* informacional (DWI) y la presentación de la información. Aun cuando las fuentes constituyen almacenes de datos operacionales con sus procesos de carga respectivos, el diseño y la instrumentación del proceso de población del *data warehouse* se han caracterizado por un examen minucioso de los datos disponibles en función de la calidad de la información suministrada para la toma de decisiones.

La primera capa de datos corresponde a las fuentes de datos que poseen información de los procesos contables y comerciales relacionados con el comercio minorista en CIMEX, que constituyen almacenes de datos operacionales (ODS-) provenientes de los sistemas transaccionales. La segunda capa corresponde al DWE, que constituye un repositorio único que concilia información contable y comercial disponiendo de los datos para el análisis. El DWE es una base de datos relacional en Tercera Forma Normal preparada para almacenar la información histórica. La tercera capa de datos derivados corresponde al *warehouse* informacional (WI), que posee un diseño orientado a apoyar la toma de decisiones de modo que los datos previamente conciliados se denormalizan y agrupan con el fin de garantizar buenos tiempos de respuestas durante la navegación y las consultas informacionales. Ya en la capa final de presentación de la información se brinda mayor dinamismo a partir de la experiencia interactiva con los datos, en ella se usan herramientas como *Power View* sobre *SharePoint* y las tablas dinámicas de *Excel*, poniendo a disposición de los usuarios funcionalidades de autoservicio, tanto para la navegación como para la creación de nuevas consultas.

El diseño informacional del repositorio de datos concibió la creación de estructuras multidimensionales que responden a los requerimientos generales. Los sujetos del negocio modelados dentro del escenario comercial son: Ventas, Compras, Inventario, Transferencias, Ajustes y Vales. En el escenario contable se modeló el Mayor General, las Cuentas por Cobrar y las Cuentas por Pagar. La integración de estos procesos de negocios se lleva a cabo a partir del diseño del esquema dimensional «Validación de Ventas» y se concibió el cubo virtual «Análisis Comercial y Contable». Se modelaron las jerarquías entre los atributos

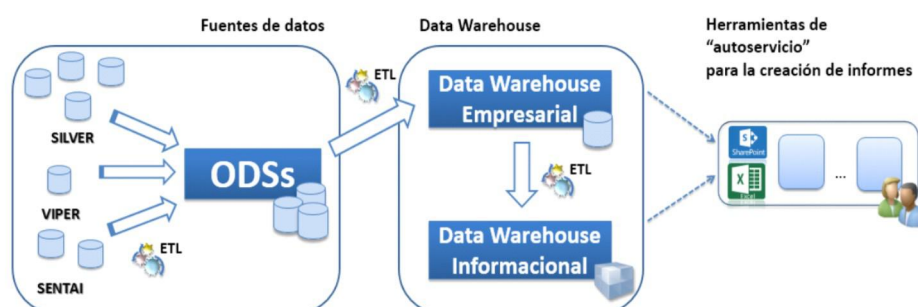


Figura 1: Modelo general de la solución

presentes en cada una de las dimensiones. En particular, la dimensión Entidad representa los establecimientos de CIMEX e incluye tres jerarquías: Entidades, que representa la estructura organizacional y es una jerarquía no balanceada o *parent-child*; Comercio Minorista: que representa el vínculo entre los establecimientos que son las entidades del comercio minorista y es una jerarquía balanceada; y Clasificación Comercial: que categoriza establecimientos, y es también balanceada. Con vistas a almacenar la historia de los cambios que han ocurrido en las dimensiones, se decidió definir una llave sustituta en la dimensión y utilizar el método de añadir un nuevo registro con los cambios efectuados (Véliz Monteagudo, 2009). Una característica importante que ha de definirse en modelos multidimensionales es la granularidad, que representa el nivel de detalle al que se almacenan los registros condicionando las posibilidades analíticas en el modelo resultante. En la propuesta, se decidió que la granularidad corresponda al nivel mensual en la dimensión Período y al nivel de detalle en la dimensión Producto, tomando en consideración que el análisis contable se realiza sobre la información mensual.

La integración de los datos en las soluciones BI (ETCL, del inglés *Extract, Transform, Cleaning and Load*) agrupa un conjunto de concepciones, técnicas y subprocesos que se encargan de llevar a cabo las tareas relacionadas con la obtención, depuración, manipulación y actualización del almacén de datos en función de los requerimientos informacionales para ulteriormente derivar la información necesaria (Cano, 2007). Para la población del DWE, estos procesos se implementaron sobre la herramienta *SQL Server 2012 Integration Services (SSIS)* de *Microsoft*. También se contempló especialmente la depuración de los datos en términos de la detección de incongruencias o información omitida. El proceso ETCL implementado está compuesto por tres paquetes principales que tienen a cargo el procesamiento de las dimensiones, los hechos comerciales y los hechos contables. Estos paquetes son ejecutados de forma periódica desde el paquete central, de modo que se actualice el repositorio de datos a partir del estado resultante de la ejecución anterior. En el flujo de ejecución se implementó la población completa de las dimensiones utilizando técnicas de concurrencia con el objetivo de

minimizar su costo computacional (A., P., & A., 2013; K. & S., 2012). No obstante, atendiendo al nivel de complejidad, el proceso ETCL se hacía muy costoso al procesar simultáneamente el volumen de datos de diferentes tareas. Concretamente, para algunas dimensiones, como Entidad y DPA (División Político-Administrativa), la ejecución de las tareas se realizan de manera secuencial, mientras que para otras dimensiones, como Producto, la población en las jerarquías involucradas se realiza de forma concurrente.

La población del *warehouse* informacional corresponde a la implementación de las bases de datos analíticas en *SQL Server 2012 Analysis Services*, el cual propone varias alternativas para hacerlo, independientes entre sí. En la solución propuesta se implementaron dos proyectos, uno que utiliza el modelo multidimensional y otro, el modelo tabular. Asimismo, se preparó un conjunto de experimentos prácticos que validan y evalúan las variantes de solución.

En la herramienta *SQL Server Data Tools* se definieron estructuras tabulares y multidimensionales que responden a los requerimientos informacionales realizados al inicio. La fuente de datos en ambos casos está constituida por el DWE. Algunas modificaciones fueron aplicadas al origen de datos como la creación de columnas calculadas, para lo que utiliza el lenguaje MDX en modo multidimensional y el lenguaje DAX para modo tabular. DAX (del inglés *Data Analysis Expression*) es el lenguaje de expresión de fórmulas analíticas el cual es utilizado para definir cálculos personalizados en los modelos tabulares y en las tablas dinámicas de *Excel* a través de *Power Pivot*. Las fórmulas DAX integran funciones, operadores y valores para llevar a cabo cálculos avanzados en tablas y columnas relacionales. Estos dos lenguajes de consultas atienden a los diferentes conceptos de modelación, debido a que MDX tiene una semántica basada en dimensiones, atributos, jerarquías y medidas, mientras que DAX está basado en tablas y columnas.

Entre las funcionalidades más atractivas empleadas en el DWI se encuentran las medidas semiaditivas y las jerarquías *parent-child*. Un ejemplo de uso de las medidas semiaditivas es el cálculo del Saldo Final. En el modo multidimensional

el conjunto de funciones de agregación *built-in* incluye también facilidades para agregaciones semiaditivas, por lo que la implementación del Saldo Final y otras medidas similares se realizó de manera directa. En el modo tabular las funciones *built-in* no incluyen facilidades para las medidas semiaditivas, por lo que fue necesario instrumentarlas con fórmulas DAX. Un ejemplo de empleo de las jerarquías *parent-child* es la instrumentación del organigrama de entidades de CIMEX. Es conveniente destacar que las jerarquías *parent-child* del modelo multidimensional en ocasiones son lentas, aunque utilizando las funciones DAX sobre jerarquías balanceadas en el modelo tabular pueden resultar más eficientes (Russo, 2012). Una vez creadas las estructuras tabulares y multidimensionales, se implementaron dos procesos ETCL para la población de las bases de datos informacionales respectivas empleando una tarea del tipo *Analysis Services Processing Task*.

En el modelo dimensional se emplea el almacenamiento por filas, requiriéndose más recursos de lectura de disco y menos de procesamiento de CPU. Por su parte, el modelo tabular utiliza el almacenamiento por columnas, de modo que el procesamiento de consultas requiere más de la utilización de CPU que de lectura a disco (J., G., C., & K., 2013; J. & H., 2013). Ambas soluciones utilizan compresión de datos dado que reducen el tamaño de las bases de datos de *Analysis Services*. Ahora bien, resulta crucial considerar que si los requerimientos de tamaño están en el orden de los *terabytes*, la solución tabular puede no ser conveniente si se tiene poca disponibilidad en memoria (RAM). Existen opciones de paginado para las soluciones tabulares, pero las grandes cantidades de datos se manejan mejor en soluciones multidimensionales. La Tabla 1 resume las características esenciales de los servidores de *Analysis Services* que se deben tener en cuenta para la selección del modelo según los recursos disponibles (Russo, 2012).

La solución propuesta también incluye la presentación de la información, aprovechando la riqueza visual de *Power View* y las facilidades que se ofrecen con las tablas dinámicas de *Excel* a través de *Power Pivot*. Los informes diseñados, que responden a un conjunto de consultas frecuentes, se publican en un sitio de

**Tabla 1: Características a tener en cuenta en los servidores de *Analysis Services*.**

| Característica                              | Multidimensional        | Tabular             |
|---|-------------------------|---------------------|
| RAM   | Menos (16/32 Gb)        | Bastante (64/128Gb) |
| Velocidad de RAM                            | Es importante           | Es crucial          |
| Número de CPU                               | 4 / 8 / 16              | 4 / 8 / 16          |
| Velocidad de CPU                            | Menos importante        | Es crucial          |
| Utilización de disco de estado sólido (SSD) | Fuertemente recomendado | No se utiliza       |
| Velocidad de la red                         | Importante              | Importante          |

*SharePoint* y pueden ser editados por los propios directivos según sus intereses.

## Resultados y discusión

Para validar la solución propuesta, se diseñaron cuatro experimentos que permiten corroborar algunos de los supuestos teóricos a los cuales se ha arribado en la presente investigación. Estos experimentos responden a las fases principales en la implementación y la presentación de los resultados. Debido al gran volumen de datos existente en los sistemas fuentes, correspondientes a los últimos tres años, en el escenario comercial se utilizaron datos de tres de las principales sucursales de CIMEX en el país, a saber: Pinar del Río, Holguín y La Habana. En el escenario contable se emplearon los datos de todas las entidades de CIMEX. Los experimentos diseñados se ejecutaron en un servidor con sistema operativo *Windows Server 2008 R2 Standard (Service Pack 1)*, un procesador *Dual-Core AMD Opteron(tm) Processor 2218* de velocidad 2.60 GHz, con memoria RAM de 3.00 GB y una arquitectura de 64 bit.

El primer experimento corresponde a la población del DWE y se concibió con el objetivo de comprobar la capacidad de los procesos ETCL implementados para poblar la capa de datos conciliados, destacando que es independiente de los modelos multidimensional y tabular. Para analizar los tiempos de ejecución de la carga inicial y posterior actualización, se dividió el experimento en tres fases. La primera corresponde a la población de los criterios de análisis; luego se cargaron los datos de las Compras de los últimos tres años, brindando una idea del tiempo de la carga inicial para un sujeto de análisis y en la última fase se ejecutaron todos los procesos ETCL para un mes, que es el período de

actualización del DWE. Como resultado, se comprobó que los procesos ETCL funcionan correctamente. Se cargaron 1.2 millones de registros de dimensiones en 40 min. y 830 mil registros de hechos de un mes (marzo/2013) en 42 min., ocupándose finalmente un espacio de 2.6 Gb. El tiempo de ejecución en este experimento se considera aceptable, dado el volumen de datos manejados y el procesamiento necesario para garantizar tanto la integridad referencial como la integración real de ambos escenarios.

El segundo experimento corresponde a la población de la capa de datos derivados y se concibió con el objetivo de analizar el comportamiento del proceso de población del WI para los dos modelos implementados, así como comparar los resultados obtenidos en cada caso. Para ello, se consideraron dos fases, correspondientes a la población de cada una de las bases de datos analíticas de la solución implementada para un mes (marzo/2013). Pudo concluirse que el modelo tabular es más rápido que el modelo multidimensional en cuanto a tiempo de procesamiento, obteniéndose resultados totales de 36 y 51 minutos respectivamente para el mismo volumen de datos. Los resultados de este experimento evidencian la capacidad del motor analítico *xVelocity* para el procesamiento y la compresión de los datos, pues la base de datos tabular ocupa alrededor del 37% de lo requerido para el almacenamiento de la base de datos multidimensional. Es preciso acotar que en el modelo tabular el volumen de datos se maneja completamente en la RAM, a diferencia del multidimensional, que almacena los datos en disco, lo que por el momento resulta más apropiado para enormes volúmenes de datos.

El tercer experimento se diseñó con el objetivo de comparar ambos modelos en cuanto a eficiencia, así como comprobar la

validez de la solución que para satisfacer los requerimientos informacionales. Fueron concibidas seis consultas con diferentes niveles de complejidad, ejecutadas cinco veces cada una buscando incrementar la precisión en las mediciones. Cabe señalar que todas las consultas se ejecutaron de forma satisfactoria devolviendo iguales resultados sobre ambos modelos, a excepción de una de ellas que devolvió *time-out* para el WI tabular. Con este comportamiento se ratifica la necesidad de poseer elevados recursos de *hardware* para utilizar el modo tabular con grandes volúmenes de datos. Resulta interesante destacar el comportamiento desigual en cuanto al consumo de memoria RAM, evidenciando el elevado consumo que genera el modo tabular en cuanto a este recurso en todos los casos. Vale señalar que entre una consulta y otra varía el comportamiento del tiempo de ejecución, atendiendo al nivel de complejidad que se demanda.

El último experimento este relacionado con el análisis informacional a partir de la solución implementada, y se concibió con el objetivo de explorar la capacidad de la solución BI para responder a los intereses organizacionales, ofreciendo mas facilidades para la navegación, para la realización de consultas dinámicas y para el enriquecimiento visual. Para este fin, se elaboraron ocho consultas diferentes, haciendo énfasis en la presentación de los resultados, en función de los requerimientos informacionales. Además, se indagó en las facilidades proporcionadas por las herramientas *Excel* y *Power View* para satisfacer las expectativas de los analistas y los ejecutivos. Todos los casos se da respuesta a requerimientos informacionales, así como a las solicitudes identificadas en relación con la presentación de los resultados. Por último, la solución de BI concebida e implementada ofrece un conjunto de funcionalidades que no se brindaban en las soluciones precedentes que favorecen la observación, la reflexión y el análisis de los datos para la toma de decisiones en CIMEX.

De lo experimentos que se realizaron, los resultados muestran que el uso del modelo tabular con grandes volúmenes de datos exige elevados recursos de hardware. Teniendo en cuenta las condiciones reales de CIMEX, en cuanto a soporte tecnológico y volumen de datos, no es conveniente

seleccionar el warehouse informacional tabular como único componente en la capa de datos derivados. Se recomienda emplear ambos tipos de modelos a la vez en dependencia de los recursos disponibles y aplicarlos convenientemente según los requerimientos presentes.

El emplazamiento del modelo tabular resulta apropiado en el contexto de las empresas cubanas que requieran la creación de un *data warehouse* con pequeño volumen de datos, aunque posean capacidades limitadas de *hardware*. Además, en este tipo de empresas los desarrolladores de bases de datos, suelen estar familiarizados con el modelo relacional, por lo que pueden realizar, con relativa rapidez, la instrumentación de soluciones sobre otro modelo basado en tablas e interrelaciones. El modelo multidimensional se adecua mejor para la creación de soluciones de BI que requieren la modelación de una lógica de negocio compleja (Vitt & Cameron, 2012). Es por esta razón se considera que este modelo sigue siendo la opción más completa para el desarrollo de bases de datos analíticas.

Los resultados alcanzados respaldan la concepción realizada en los términos informacionales y el modelo de solución propuesto, con lo que se comprueba la importancia de la etapa de diseño en el desarrollo de una solución de BI. Además, se constató la eficacia de los procesos de integración para la población de las capas de datos conciliados y derivados. Por último, a partir de los experimentos realizados, puede asumirse la validez de la nueva solución para la Inteligencia de Negocios que actúe en correspondencia a los requerimientos informacionales prioritarios de CIMEX.

## Conclusiones

Como resultado del presente trabajo se concibió, diseñó e implementó una nueva aproximación de solución de Inteligencia de Negocios para el Grupo Empresarial CIMEX, en la que se integra la información comercial y contable en pos de satisfacer requerimientos informacionales particulares y concertados. La solución funcionó como entorno práctico para el estudio de las características del clásico modelo dimensional y el novedoso modelo tabular, así como en las funcionalidades de

herramientas de visualización como *Microsoft Excel y Power View*, en función de agilizar y enriquecer el ambiente analítico puesto a disposición de los especialistas y ejecutivos. Se constató que el modelo tabular no constituye un aporte conceptualmente diferente, sino una implementación alternativa del modelo dimensional de datos para la herramienta de procesamiento analítico *SQL Server 2012 Analysis Services*, aun cuando el soporte sobre el almacenamiento columnar y las bases de datos *in-memory* le aportan un interés actual y una perspectiva fructífera al modelo tabular, más allá de lo que puede ofrecer Microsoft.

A partir del estudio de ambos modelos, la experiencia adquirida en el desarrollo del sistema y el comportamiento de los resultados obtenidos en los experimentos, se pudo realizar una comparación práctica y cualitativa de ambas opciones. Como resultado, se esbozaron consideraciones generales respecto a fortalezas y debilidades contribuyendo al trabajo ulterior en esta área del conocimiento. Se concluye que el modelo tabular ofrece una alternativa de modelación con la que los usuarios se identifican mejor a partir de necesidades de análisis y consultas, pero se basa en el uso y el aprovechamiento de elevados recursos de *hardware*, principalmente memoria *RAM*. Por otro lado, se corroboró que el modelo multidimensional puede constituir la mejor opción si se requiere una compleja modelación de la lógica del negocio o se necesita una solución escalable y que maneje grandes volúmenes de datos.

La solución de BI propuesta sirve como punto de partida para el desarrollo de soluciones computacionales que integren múltiples escenarios y brinden una experiencia de autoservicio a los ejecutivos y analistas en relación con la generación de consultas «al vuelo» y la creación de informes respaldados visualmente. También sienta las bases para futuras aplicaciones de descubrimiento de conocimiento que enriquezcan la toma de decisiones en las organizaciones.

## Bibliografía

A., Karagiannis, P., Vassiliadis, & A., Simitsis. (2013). Scheduling strategies for efficient ETL

execution. *Journal Information Systems*, vol. 38(no. 6), p. 927-945

Albano A., De Rosa L., Dumitrescu C., Goglia L., Goglia R., & V, Minei. (2006). Another Example of a Data Warehouse System Based on Transposed Files. Paper presented at the 10th International Conference on Advances in Database Technology.

B., Evelson, & N., Nicolson. (2008). Topic Overview: Business Intelligence. Retrieved from Forester Research website: <http://bit.ly/foresterbi>

Ballard, C., Farrell, D., Gupta, A., Mazuela, C., & Vohnik, S. (2006). Dimensional Modeling: In a Business Intelligence Environment. United States of American: International Business Machines Corporation.

Bernabeu, R. (2010). HEFESTO. Data Warehousing: Investigación y Sistematización de Conceptos. Hefesto: Metodología para la Construcción de un Data Warehouse. Córdoba, Argentina.

Boateng, Ofori, Singh, Jagir, Greeshma, & Singh, P. (2012). DATA WAREHOUSING. *Business Intelligence Journal*, vol. 5(no. 2), pp. 224-234.

Cano, J. (2007). Business Intelligence: Competir con Información. Barcelona, España: Universidad Ramon Llull. Escuela Superior de Administración y Dirección de Empresas (ESADE).

Devlin, B. (1997). Data Warehouse from Architecture to Implementation: Addison Wesley Longman, Inc. .

García Hernández, L., Oliva-Santos, R., Prendes-Arencibia, H., Velázquez-Vidal, L., & Véliz - Monteagudo, M. (2008). La Inteligencia de Negocios desde la perspectiva de los datos. Paper presented at the 4to Evento Nacional de Informática/08, Cuba.

García, L., Simón, A., Torres, M., & Espinosa, Y. (2013). Solución

- de inteligencia de negocios para la integración de la información comercial y contable. Paper presented at the Congreso Internacional de Matemática y Computación 2013, La Habana, Cuba.
- J., Sirmon, G., Galloway, C., Gross, & K., Gulati. (2013). Performance Tuning of Tabular Models in SQL Server 2012 Analysis Services. SQL Server Technical Article.
- J., Sirmon, & H., Steen. (2013). Hardware Sizing a Tabular Solution (SQL Server Analysis Services). SQL Server Technical Article.
- K., Srinivasu, & S., Satapathy. (2012). Optimization of ETL Work Flow in Data Warehouse. International Journal on Computer Science and Engineering (IJCSE), vol. 4(no. 9), p. 1579-1586.
- Kimball, R., & Ross, M. (2002). The Data Warehouse Toolkit : The Complete Guide to Dimensional Modeling (3th ed.). United States of America: Wiley Computer Publishing.
- Lachev, T. (2012). Applied Microsoft SQL Server 2012 Analysis Services: Tabular Modeling.
- Leland, D. (2012). Introducing xVelocity in-memory technologies in SQL Server 2012 for 10-100X performance SQL Server Blog (Vol. 2014).
- M., Stonebraker, D., Abadi, A, Batkin, & X., Chen. (2005). C-Store: A Column-oriented DBMS Paper presented at the 31st Very Large Data Bases (VLDB) Conference, Trondheim, Norway.
- Myers, P. (2012). Introducing the BI Semantic Model in Microsoft® SQL Server®
- P., Savjani. (2014). BI Solutions using SSAS Tabular Model. Morrisville, USA: Synsfusion Inc.
- Power Pivot Add-in. (2010). <http://office.microsoft.com/en-001/excel-help/power-pivot-add-in-HA101811050.aspx> Power View: Explore, visualize, and present your data. (2013). <http://office.microsoft.com/en-001/excel-help/power-view-explore-visualize-and-present-your-data-HA102835634.aspx>
- Russo, M. (2012). BISM: Multidimensional vs. Tabular. Paper presented at the Microsoft TechEd 2012.
- Russo, M., Ferrari, A., & Webb, C. (2012). Microsoft SQL Server 2012 Analysis Services: The BISM Tabular Model: Microsoft Press.
- Velázquez-Vidal, L. (2009). Herramienta genérica para la población del Data Warehouse Informacional. (Tesis en opción del título de Máster en Ciencia de la Computación.), Facultad de Matemática y Computación. Universidad de La Habana, Cuba.
- Véliz Monteagudo, M. (2009). Herramienta genérica para la población del Data Warehouse Empresarial. (Tesis en opción del grado de Máster en Ciencia de la Computación.), Facultad de Matemática y Computación. Universidad de La Habana, Cuba.
- Vitt, L., & Cameron, S. (2012). Choosing a Tabular or Multidimensional Modeling Experience in SQL Server 2012 Analysis Services. Microsoft Business Intelligence.

Recibido: 4 de agosto de 2014.  
Aprobado en su forma definitiva:  
16 de enero de 2015

---

**Maité Torres Sánchez**  
Universidad de La Habana, Cuba.  
Correo-e.: maitets@matcom.uh.cu

**Yissel Espinosa Cervantes**  
Universidad de La Habana, Cuba.  
Correo-e.: yisselec@yahoo.com

**Alina Simón Cuevas**  
Datacimex, Grupo Empresarial CIMEX, La Habana, Cuba.  
Correo-e.: asimonc@cimex.com.cu

**Lucina García Hernández**  
Universidad de La Habana, Cuba.  
Correo-e.: lucina@matcom.uh.cu

**Alfredo J. Simón Cuevas**  
Instituto Superior Politécnico José Antonio Echeverría (CUJAE), La Habana, Cuba.  
Correo-e.: asimon@ceis.cujae.edu.cu

---