

Evolución de la gestión del conocimiento en el Grupo Empresarial CIMEX

Alina Simón Cuevas

asimonc@cimex.com.cu

Datacimex, Grupo Empresarial CIMEX

Maité Torres Sánchez

maitets@matcom.uh.cu

Universidad de La Habana

Lucina García Hernández

lucina@matcom.uh.cu

Universidad de La Habana

Raudel Ravelo Suárez

ravelo@matcom.uh.cu

Universidad de La Habana

RESUMEN

En la actualidad los sistemas para la toma de decisiones son necesarios en la conducción de cualquier organización. El Grupo Empresarial CIMEX, una de las empresas comercializadoras cubanas, tiene una amplia experiencia en el desarrollo de soluciones computacionales que en el transcurso de los años han evolucionado dando respuesta a los requerimientos informacionales de los analistas y directivos. Diferentes áreas de negocio –comercial, económica y recursos humanos– se han manejado históricamente mediante aplicaciones independientes, dificultándose la conciliación entre los resultados particulares y el análisis integral de la información. En el presente trabajo se reseña los hitos principales de la progresión de la gestión del conocimiento en la Corporación desde una perspectiva computacional y se expone la solución actual basada en la extensión y el enriquecimiento del almacén de datos existente. Las funcionalidades analíticas incorporadas se han implementado sobre el modelo tabular, ofrecido por Microsoft SQ Server, que se sustenta en las tecnologías de almacenamiento columnar y las bases de datos en memoria. Además, se brinda la versión propia para controlar el desempeño empresarial sustentado en el análisis de los principales indicadores claves de éxito, lo que posibilita la creación de cuadros de mando ajustados a cada una de las empresas con vista a la proyección, ejecución y control de sus planes tácticos y estratégicos. Finalmente se presenta una primera aproximación del proceso de minería de datos para la predicción de indicadores gerenciales.

PALABRAS CLAVES: Inteligencia de Negocios, Modelo Tabular, Bases de Datos Orientadas a Columnas, Bases de Datos en Memoria, Indicadores Claves de Éxito, Minería de Datos

ABSTRACT

Nowadays decision support systems are essential in any organization management. CIMEX Corporation, a major Cuban trading enterprise, has extensive experience in developing computer solutions over the years which progressively evolved in response to experts and managers

informational requirements. Different business areas –such as trade, economic and human resources–have historically handled through independent solutions, making difficult particular results integration and widespread information analysis. This paper outlines main milestones of knowledge management progression in the company from computational point of view. Projected solution is based on an extension and enrichment of current data warehouse. Built-in analytical capabilities have been implemented over tabular model, supported on columnar storage and in-memory database technologies and provided by Microsoft SQL Server. In addition, the own version allows driving business success based on main key performance indicators. Companies' dashboards enable design and control of tactical and strategic plans. Finally, a first approach of data mining process is applied to predict critical behavior measurements.

KEYWORDS: Business Intelligence, Tabular Model, Columnar Databases, In-Memory Databases, Key Performance Indicators, Data Mining

INTRODUCCIÓN

Desde el surgimiento del comercio, la contabilidad, las finanzas y las diversas ramas de negocios, el hombre ha estado inmerso en una continua hazaña por obtener mejores resultados en cuanto a cómo controlar y organizar cualquier institución o entidad con el fin de garantizar su existencia y evolución a largo plazo, en cuyo marco desempeña un papel esencial la calidad del proceso de toma de decisiones. Con el desarrollo de las Tecnologías de la Información y las Comunicaciones (TIC) la forma de hacer negocios y de competir en los nuevos mercados está promoviendo cambios en el modo de administrar las empresas, basado en el conocimiento. La Gestión del Conocimiento (*Knowledge Management*) es la disciplina que contribuye a transferir el conocimiento desde el origen hasta el destino donde será asimilado, empleado, valorado y/o compartido convenientemente entre las personas dentro y fuera de una organización, lo que propicia el desarrollo de la capacidad de resolución de problemas y contribuye a la sostenibilidad de las ventajas competitivas (Andrew & Sieber, 1999).

Ante la tensión de las organizaciones competidoras, los dinámicos escenarios de mercado, la necesidad de innovación permanente, el incremento del protagonismo del capital humano e intelectual y la urgencia de sostenerse y trascender en el tiempo, resulta imperioso que las empresas opten por concepciones y mecanismos cada vez más avanzados y sistemáticos para evaluar su eficacia y utilizar de manera favorable los recursos y fuentes de información para el logro de objetivos tácticos y estratégicos. De ahí que, además de concentrarse en lo inevitablemente apremiante y/o rutinario, se debe estudiar de alguna forma lo importante a corto, mediano y largo plazo. El análisis del desempeño está íntimamente asociado a la gestión del conocimiento, sin la cual no es posible determinar grados de cumplimiento o desvíos respecto a resultados esperados (Aranibar, 2013).

Por otra parte, cada día nos enfrentamos más a un fenómeno de inundación de datos y el tiempo no alcanza para explorar todo lo que se requiere. En aras de encontrar maneras para analizar automáticamente los datos, clasificarlos, resumirlos, caracterizarlos, así como descubrir información y conocimiento, estimar tendencias o detectar anomalías en ellos, se ha desarrollado

la minería de datos que constituye una de las áreas más activas de la comunidad de investigación de base de datos (Dean, 2014). Lo que diferencia a la época actual, aparte de ser considerada la era de los grandes datos “*Big Data*”, es que las empresas, los gobiernos y las organizaciones han experimentado un cambio en el pensamiento en términos de querer aprovechar todos los datos que son capaces de coleccionar para mejorar su negocio (Dean, 2014).

Una de las empresas cubanas que se encarga de la importación y comercialización de productos y servicios es el Grupo Empresarial CIMEX el cual, además de ser líder nacional en este mercado, posee un reconocido desarrollo informático. Cuenta con más de diez años de experiencia en la creación de soluciones de inteligencia de negocios que gradualmente han satisfecho las crecientes necesidades analíticas de los decisores. Estas aproximaciones comenzaron con informes estáticos sobre las bases de datos transaccionales y fueron evolucionando a través del tiempo. La solución más reciente del ciclo evolutivo de los sistemas para la toma de decisiones en CIMEX fue la concepción de un almacén de datos que integró los escenarios comercial y contable y se implementó sobre la base de nuevas tecnologías de procesamiento analítico con vistas a propiciar la experiencia interactiva de los especialistas con los datos.

Desde el punto de vista informacional, la presente propuesta constituye una extensión de la solución existente en CIMEX al integrar nuevos escenarios de análisis, como las finanzas y los recursos humanos. Además, se expone un nuevo proceder para controlar el desempeño de las entidades basado en el análisis de los principales indicadores de éxito con vistas particulares, jerárquicas y globales de la información. Con ello se posibilita la implementación de cuadros de mando integrales, metodología de dirección que en estos momentos está en proceso de gestación en la Corporación. Por último, se presenta una primera aproximación de un modelo de minería de datos para la predicción de indicadores, como punto de partida de futuros desarrollos.

El problema científico que motiva la realización de esta investigación es la imposibilidad de controlar de manera automatizada la eficacia y la salud organizacional de CIMEX basados en indicadores de desempeño.

El objetivo general radica en concebir una solución de inteligencia de negocios que contribuya a enriquecer la evaluación del desempeño empresarial, extendiendo el almacén de datos existente y aprovechando las fortalezas del modelo de datos tabular. Para lograrlo, se trazaron los siguientes objetivos específicos:

- Sintetizar la evolución de las soluciones computacionales en CIMEX desde la perspectiva de la gestión del conocimiento.
- Reformular el diseño conceptual de la solución de inteligencia de negocios precedente.
- Complementar la solución computacional con la incorporación de herramientas que favorezcan el análisis y la exploración de los principales indicadores de desempeño de CIMEX, sobre la base de las contribuciones del modelo tabular.
- Incursionar en las técnicas de minería de datos para la predicción de indicadores financieros e instrumentar una primera variante.
- Realizar experimentos que validen la concepción de la solución y el cumplimiento de los requerimientos informacionales.

La implementación de la solución informacional se realizó sobre el modelo tabular de *Microsoft SQL Server 2012 Analysis Services* (SSAS). Al ser un modelo analítico teóricamente más cercano al clásico modelo relacional, su utilización resulta más natural para los desarrolladores y los analistas y es posible lograr una mayor independencia entre ellos. Se eligió el algoritmo de Series de Tiempo ofrecido por SSAS para el diseño y la experimentación del proceso de minería propuesto con el objetivo de predecir indicadores de interés primordial para los especialistas y los directivos.

DESARROLLO

El volumen de datos adquiridos en las organizaciones está creciendo rápidamente, y con ello la cantidad de usuarios que necesitan acceder y analizar dichos datos. El procesamiento de los datos y la presentación de informes de negocios requieren gradualmente más recursos en aras de potenciar la gestión del conocimiento en la organización. En este contexto la Inteligencia de Negocios (BI, *Business Intelligence*) ha estado dando respuesta al ofrecer alternativas cada vez más efectivas para facilitar el manejo de la información a los analistas y los directivos, con el fin de propiciar decisiones más precisas y expeditas en entornos de cambio permanente (Matei, 2010).

Evolución de las soluciones de inteligencia de negocios en CIMEX

En CIMEX se ha priorizado tanto la automatización de las operaciones empresariales principales como la evolución necesaria de acuerdo con el desarrollo científico, tecnológico y profesional en sus entidades. No solo se han dado respuestas oportunas y pertinentes al procesamiento de los datos operacionales sino también se ha trabajado con el propósito de facilitar los procesos de dirección con herramientas y ambientes computacionales propios, que han sentado las bases para el desarrollo de la presente investigación.

Los datos primarios que han de recuperarse para el análisis de los requerimientos informacionales empresariales están localizados en diferentes sistemas computacionales de gestión empresarial (ERP, *Enterprise Resource Planning*). En CIMEX se utilizan en la actualidad tres sistemas ERP que permiten registrar esencialmente los datos de las operaciones básicas en los puntos de venta: Silver, Sentai-Trax y Sentai-Viper.

En el año 2000 se desarrolló una solución, implementada en el lenguaje de programación *Visual Foxpro* 5.0, que brindaba funcionalidades para el análisis de las estadísticas comerciales minoristas. Consistía en una aplicación de escritorio instalada en cada una de las sucursales y la versión central en la Casa Matriz. Este sistema ubicó a la Corporación entre las primeras organizaciones cubanas en diseñar una solución computacional en términos de la toma de decisiones. Constituyó el primer resultado que unificaba los datos disgregados en los sistemas operacionales, permitiendo el seguimiento de los indicadores comerciales. Logró brindar un conjunto de funcionalidades básicas para el análisis, tales como visualizar reportes parametrizados. Sin embargo, el transcurso del tiempo y el desarrollo de las nuevas concepciones y tecnologías pusieron de manifiesto los inconvenientes siguientes: al realizar consultas no era posible mostrar datos del mes en curso hasta tanto finalizara, solo se mostraba información de los meses vencidos; en una misma consulta no era posible realizar comparaciones entre diferentes años y los mecanismos de sincronización

entre las diferentes instancias de la aplicación no eran eficientes; la interfaz visual no era lo suficientemente amigable; el lenguaje de programación *Visual Foxpro* 5.0 no facilitaba la realización de actualizaciones ni la evolución de la solución en entornos contemporáneos.

En el período comprendido entre los años 2003 y 2006 se desarrolló un *data warehouse* (DW) centrado en la actividad comercial. Este sistema marcó un hito en la concepción de las soluciones de inteligencia de negocios en CIMEX, ya que por primera vez se utilizó un repositorio de datos basado en el modelo dimensional. Se brindaba al usuario la posibilidad de definir con cierto dinamismo cómo deseaba visualizar la información, así como las condiciones de búsqueda. Estas posibilidades permiten afirmar que se logró un paso de avance con respecto a los resultados precedentes. A pesar de las ventajas de esta aproximación, su puesta en práctica demostró que la organización no estaba lo suficientemente madura desde el punto de vista tecnológico ni cultural, debido a que el proceso de integración de los datos no garantizaba la actualización eficiente del DW y los usuarios demandaban reportes específicos para responder necesidades operativas, tales como alertas de reabastecimiento.

Entre los años 2007 y 2011 en el grupo de desarrollo de Inteligencia de Negocios se concibió una solución que contaba con varios almacenes de datos operacionales (ODS, *Operational Data Store*) correspondientes a las distintas áreas de gestión. En esta aproximación los datos se encuentran centralizados, consolidados y resumidos convenientemente para satisfacer los requerimientos informacionales para la toma de decisiones a corto plazo. Se brinda a los usuarios la posibilidad de generar dinámicamente informes dentro de plantillas predefinidas. Sin embargo, las respuestas que sustentan el análisis son matriciales, con posibilidades limitadas de navegación. A partir de los ODS durante varios años han sido concebidas diversas soluciones analíticas particulares como la consolidación contable (Conso), la gestión de estadísticas financieras (Sisges), el control del reabastecimiento minorista, el perfeccionamiento del sistema de estadísticas minoristas–, lo que contribuyó a fortalecer el conocimiento de las tecnologías de procesamiento analítico en línea (OLAP, *Online Analytical Processing*) en la organización. Como parte de este proceso evolutivo se desarrolló el portal web de inteligencia de negocios SAN (Sistema de Administración de Negocios) que permite el acceso a la información desde cualquier lugar de la red corporativa.

En el año 2012 los directivos de CIMEX insistieron en el interés de disponer de la información centralizada de los principales sujetos del negocio combinando los datos de varios escenarios que hasta ese momento se habían presentado en soluciones independientes. Surge así el proyecto de la construcción de un nuevo almacén de datos sobre una arquitectura de tres capas. Al mismo tiempo, *SQL Server 2012* propuso un nuevo modelo de datos para representar la lógica del negocio de manera tabular. Ello motivó que la solución fuera implementada sobre los modelos multidimensional y tabular con el propósito de aprovechar las bondades y valorar las insuficiencias de cada uno desde el punto de vista analítico. Este resultado constituyó un logro en la organización, ya que se agrupó y se concilió los datos de dos ramas principales: la contable y la comercial. Es de resaltar el ambiente analítico, completamente novedoso y atractivo para los usuarios, con funcionalidades de autoservicio para la visualización que enriquecen las alternativas

para la toma de decisiones (García et al., 2013; Maité Torres et al., 2014; Maite Torres et al., 2015). Asimismo, contribuyó al avance de la cultura empresarial en cuanto a la gestión del conocimiento.

En la figura 1 se puede observar un resumen de la evolución de las soluciones de inteligencia de negocios en la organización. Hasta el momento se puede concluir que las soluciones empleadas para el análisis de las estadísticas comerciales y contables en el Grupo Empresarial CIMEX han evolucionado respetando la llamada Pirámide del Conocimiento que expresa la relación existente entre datos, información y conocimiento.

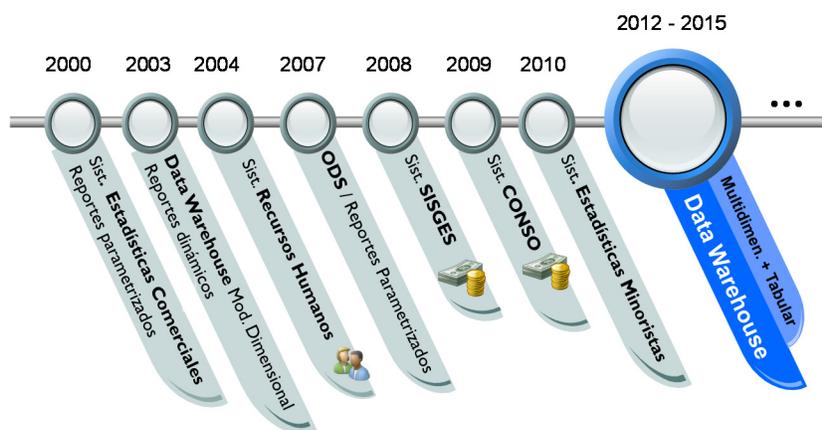


Fig. 1 Evolución de las soluciones de inteligencia de negocios en CIMEX.

Actualmente el control del desempeño empresarial basado en indicadores claves de éxito se ha convertido en un propósito para la alta dirección de la Corporación. La satisfacción de este empeño requiere no solo de la incorporación de nuevos escenarios de análisis sino también la profundización en las tecnologías para la gestión del conocimiento.

El modelo tabular para el análisis del desempeño corporativo

El rendimiento de un sistema de gestión de bases de datos (DBMS, *Database Management System*) está directamente relacionado con la eficiencia del almacenamiento de datos en disco y su movimiento hacia los registros del CPU para el procesamiento. El almacenamiento por filas/registros (*row-oriented storage*) utilizado históricamente por la mayoría de los DBMS, está optimizado para el procesamiento transaccional de los datos (OLTP, *Online Transactional Processing*) (Albano et al., 2006; Matei, 2010). En este tipo de almacenamiento cada fila de una tabla se almacena de forma consecutiva y generalmente la tabla se organiza en páginas de datos, de modo que cada página contiene una cantidad de filas en dependencia del tamaño de los registros. Una de sus desventajas radica en que, si se desea obtener todos los valores de una misma columna de una tabla, es preciso leer todas sus filas (Russo et al., 2012).

En búsqueda de la eficiencia en el almacenamiento y el procesamiento de los datos, en los últimos años se ha renovado el interés en el almacenamiento por columnas (*column-oriented storage*) y diversos autores han defendido sus beneficios (D. J. Abadi et al., 2008; Albano et al., 2006; Cattell, 2010a, 2010b). En este tipo de distribución cada página de datos contiene los valores correspondientes a una sola columna y, al tratarse de datos estructurados con valores

generalmente repetidos, se favorece la compresión y rápida recuperación. En la figura 2 se presenta un diagrama por cada tipo de almacenamiento.

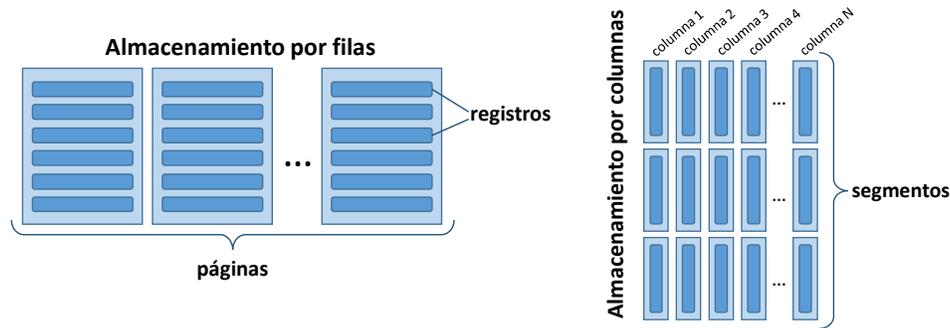


Fig. 2 Almacenamiento por filas y almacenamiento por columnas.

Si bien las técnicas de almacenamiento orientadas a las columnas aparecieron en los años 70s, no fue hasta la primera década del siglo XXI cuando las investigaciones al respecto ganaron auge, potenciadas fundamentalmente por el desarrollo del *hardware* –CPUs más rápidos, memorias más capaces, procesamiento en disco más eficiente– lo que provocó su implementación en el marco de prestigiosos DBMS (D. Abadi et al., 2012; Matei, 2010). El acceso a estas técnicas ha tenido un gran impacto en el diseño y la instrumentación de los sistemas de recuperación de información, *data warehousing* y soluciones OLAP, en los cuales la información se obtiene resumiendo grandes volúmenes de datos.

Varios acontecimientos de los últimos años han despertado el interés en el procesamiento analítico en memoria y lo han hecho más asequible las plataformas de 64 bits, los servidores multi-procesadores y la disminución constante de los precios de las memorias RAM (D. Abadi et al., 2012; Matei, 2010). Las bases de datos en memoria (IMDB, *in-memory databases or main memory database system*) son bases de datos cuya información se almacena en la memoria principal de la máquina en un formato comprimido. Resulta obvio que estas bases de datos son mucho más eficientes que las tradicionales almacenadas en disco y se recomiendan en ambientes donde el tiempo de respuesta a las consultas es crítico. Aun cuando no se ha logrado producir memorias principales que respondan a los crecientes volúmenes de datos que se manejan en la actualidad, el uso de esta tecnología se fomenta cada vez con mayor ímpetu en las empresas que lidian con las transacciones en línea de gran escala o requieren la predicción en tiempo real (Plattner, 2014).

El Grupo Empresarial CIMEX ha dedicado esfuerzos y recursos durante años en función de la evolución de las concepciones y las técnicas de dirección con vistas a enriquecer la toma de decisiones. Desde la perspectiva tecnológica, existe un interés en profundizar y aplicar los últimos avances en la infraestructura actual en aras de continuar sustentando los procesos para gestionar el conocimiento.

Microsoft ofrece las técnicas previamente reseñadas –almacenamiento por columnas y bases de datos en memoria– en su gestor de bases de datos SQL Server a partir de su versión 2012. Sumándole a estos recursos la experiencia alcanzada por los especialistas de CIMEX en el empleo

de este sistema desde el siglo pasado, la plataforma de inteligencia de negocios de Microsoft se reafirma como soporte para el despliegue de las soluciones computacionales.

En el contexto actual de la Corporación se percibe un cambio en la forma de pensar de analistas y directivos en términos de potenciar la utilización de los datos tan rápido y ampliamente como sea posible. A ello se intenta responder en esta aproximación incorporando nuevos escenarios de análisis como las finanzas y los recursos humanos –sobre la base del modelo tabular– y potenciando las alternativas de evaluación del funcionamiento de la organización a través de cuadros de mando integrales.

El concepto de cuadro de mando integral (CMI), conocido también como tablero de comando o *balanced scorecard* (BSC), fue propuesto en la revista *Harvard Business Review* de la Universidad de Harvard y se destaca como metodología de dirección empresarial (Kaplan & Norton, 1992). No es posible lograr un CMI efectivo si no se cuenta con datos depurados, conciliados, integrados –siempre que sea conveniente– y organizados eficientemente en función del análisis dinámico y proactivo, lo que permite conducir el desempeño actual y proyectar el futuro. Se puede afirmar que, gracias a la evolución satisfactoria de las soluciones computacionales abordadas, hoy se cuenta con condiciones suficientes para explotar el empleo de esta metodología.

Los indicadores de desempeño claves (KPI, *Key Performance Indicator*) constituyen un componente fundamental del cuadro de mando integral. Un KPI es una métrica que permite comparar el valor de una variable contra un valor deseado que está representado en otra variable o como un valor absoluto dado. Los KPI se utilizan para cuantificar y monitorear gráficamente los objetivos del plan estratégico de la organización, lo que contribuye a tener una visión concreta de su comportamiento (Aranibar, 2013). La explotación sistemática de los cuadros de mando integrales no solo contribuye a tomar decisiones certeras y pertinentes, sino que cambia la manera en que se mide y gerencia un negocio (Kaplan, 2010; Kaplan & Norton, 1992).

Durante la concepción de la presente propuesta, se identificaron los requerimientos informacionales a partir de reuniones de intercambio y se estudiaron los informes gerenciales mensuales que se elaboran por los directivos. Entre los requerimientos examinados se pueden mencionar:

- Analizar mensualmente el cumplimiento de los indicadores financieros, tales como los ingresos totales detallados por tipos de servicios y por monedas, los costos y los gastos, las utilidades detalladas por monedas, las cuentas por cobrar detalladas por tipos de clientes y por edades, las cuentas por pagar detalladas por tipo de proveedores y por edades, el estado de los inventarios por tipo y ciclos de rotación.
- Analizar mensualmente el cumplimiento de los indicadores de recursos humanos, entre los que se encuentran el gasto de salario por peso de ingreso, la correlación entre el salario medio y la productividad, el aprovechamiento de la jornada laboral por tipos de ausencias, el completamiento de la plantilla por cargos, el salario medio mensual y las incidencias laborales detalladas por tipo.

- Comparar los resultados de los indicadores en diferentes períodos (mensuales, trimestrales y anuales).
- Resumir el estado del negocio mediante cuadros de mando que reflejen el funcionamiento de las empresas y el desempeño global de la organización.

Para dar respuesta a las demandas esbozadas fue necesario realizar un análisis crítico y constructivo de la solución precedente, manteniendo lo valioso y desechando o modificando lo necesario. En la figura 3 se presenta el modelo conceptual de la solución de inteligencia de negocios extendida. El principal componente lo constituye el almacén de datos, donde no solo se integran el escenario comercial y el escenario contables, sino también los nuevos escenarios de interés, económico-financiero y recursos humanos. Para ello fue preciso desarrollar procesos de extracción, transformación, depuración y carga con el propósito de incorporar al *data warehouse* empresarial datos primarios presentes en nuevas fuentes.

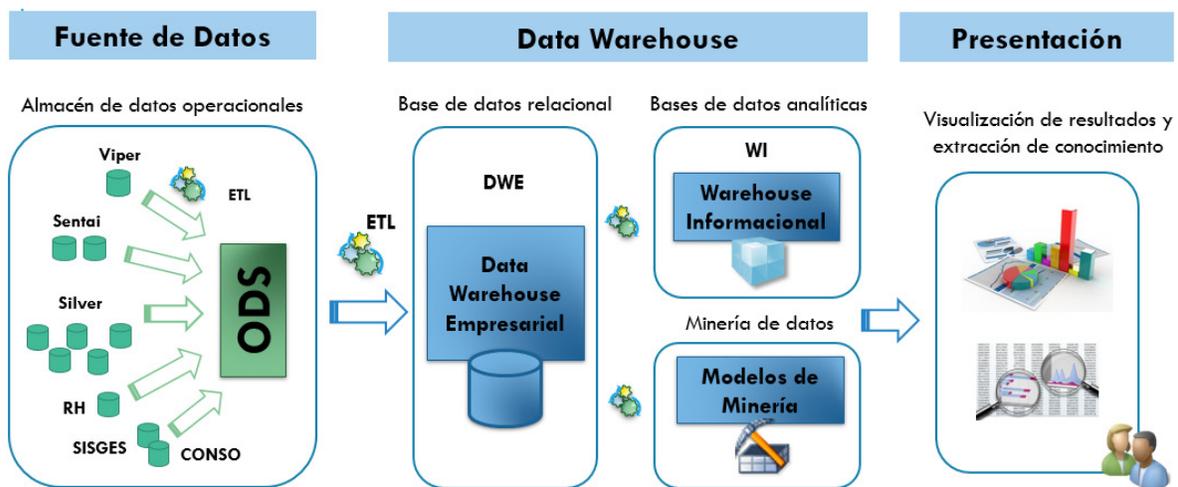


Fig. 3 Modelo general de la solución.

El diseño de la base de datos analítica consistió en la reformulación del *warehouse* informacional, adicionando conceptualmente los modelos multidimensionales que respondieran a los nuevos requerimientos según las áreas de negocio incorporadas. En el contexto económico-financiero se modelaron Indicador Económico e Indicador Financiero y, en el escenario de recursos humanos, los sujetos modelados fueron Pago Salario, Incidencia Laboral e Indicador General. Se modelaron diversas jerarquías entre los atributos presentes en las dimensiones para enriquecer el análisis. La jerarquía irregular Tipo de Incidencia contribuye a examinar el comportamiento de las incidencias según el grado de afectación al tiempo laboral y al ausentismo. En las figuras 4a y 4b se muestran los esquemas dimensionales más significativos. Cabe observar que varios criterios de análisis se comportan como dimensiones compartidas entre los múltiples esquemas existentes en el almacén de datos.



Fig. 4a Esquema dimensional para los Indicadores Financieros.

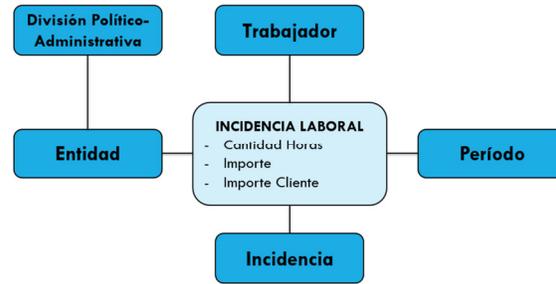


Fig. 4b Esquema dimensional para las Incidencias Laborales de los Trabajadores.

La capa de presentación de la información se modificó para proveer las nuevas funcionalidades analíticas, conservando la concepción de favorecer el autoservicio. Se diseñaron, asimismo, indicadores claves de desempeño para reflejar el funcionamiento de la organización en el devenir del tiempo.

Si bien los datos pudieran tener cierto valor en sí mismos, más aún cuando son preparados convenientemente para el análisis y convertidos en información, el conocimiento que de ellos se puede derivar resulta ser una motivación novedosa en esta investigación. Descubrir nuevo conocimiento es generalmente difícil de lograr con el procesamiento analítico en línea y la presentación de informes, de ahí que se incorporó al diseño conceptual del sistema un primer acercamiento a un proceso de minería de datos.

En una primera intención abordamos la predicción de los indicadores económicos y financieros por ser una de las necesidades más acuciantes en CIMEX y contar con un volumen significativo de datos históricos. La Minería de Datos (*Data Mining*) es un proceso de exploración y análisis, por medios automáticos o semiautomáticos, de los datos históricos para descubrir patrones y reglas, que se pueden aplicar sobre nuevos datos para las predicciones y previsiones (Aggarwal, 2015; Dean, 2014).

Entre los tipos de problemas de minería de datos más conocidos se encuentran la clasificación, el agrupamiento, la similitud y la regresión, constituyendo este último el más apropiado para responder a los requerimientos identificados. Las series de tiempo constituyen métodos de regresión para la previsión de valores continuos en el tiempo, por lo que en el presente trabajo se han utilizado para la predicción de los indicadores económico-financieros, tales como los ingresos y los gastos (Aggarwal, 2015; Chessman, 2013; Singh Bassan & Sarkar, 2014).

Vale señalar que el proceso de minería se presenta como un componente independiente en el sistema, que se nutre directamente de la capa de datos conciliados DWE para la definición de las vistas minables. A partir de las vistas minables se seleccionan los atributos que expresan el tiempo (período) y los atributos a predecir (los indicadores), así como el algoritmo a utilizar.

Implementación de la propuesta de solución

Durante más de quince años se ha acumulado una significativa experiencia en la explotación de *Microsoft SQL Server* como sistema de gestión de bases de datos y plataforma de desarrollo de soluciones de inteligencia de negocios en CIMEX. Las facilidades que han ofrecido las diferentes herramientas para la gestión de las bases de datos, así como para la creación y la explotación de los almacenes de datos, se han aprovechado convenientemente en las soluciones computacionales predecesoras. Las variadas y originales funcionalidades provistas para el procesamiento analítico y la visualización de los resultados, según los intereses de los usuarios, fundamentan la elección de *Microsoft SQL Server 2012* para la implementación de la propuesta de solución actual en el marco de la presente investigación.

El proceso ETL (*Extract, Transform and Load*) se ha extendido convenientemente para la población del *data warehouse* empresarial (DWE) y se compone de cinco módulos principales: carga de las dimensiones → carga de hechos comerciales → carga de hechos contables → carga de hechos financieros → carga de hechos de recursos humanos. Con vistas a apoyar el ciclo de inteligencia de negocios en pos de mejorar la captura y el almacenamiento de los datos primarios, se concibieron mecanismos para la depuración de los datos como la detección de incongruencias o información omitida, eventos que se registran en tablas de trazas (*logs*) para lograr la retroalimentación hacia los sistemas transaccionales fuentes. Para el desarrollo de estas tareas se utilizó *SQL Server 2012 Integration Services* (SSIS).

Microsoft SQL Server 2012 Analysis Services (SSAS) brinda dos alternativas para la creación de los modelos analíticos, el clásico modelo multidimensional y el reciente modelo tabular. Al mismo tiempo, ofrece el Modelo Semántico de Inteligencia de Negocios (BISM, *Business Intelligence Semantic Model*), que permite combinar datos desde múltiples fuentes heterogéneas en un único modelo común para las aplicaciones clientes, independientemente de la forma de visualización de la información y viable para todos los tipos de usuarios (Myers, 2012; Russo et al., 2012). A partir de la arquitectura conceptual de datos en tres capas de la solución de inteligencia de negocios concebida, BISM se utiliza esencialmente como una capa de abstracción entre los usuarios finales y los modelos de datos analíticos. De manera transparente, las consultas DAX (*Data Analysis Expression*) pueden ejecutarse sobre los cubos dimensionales y las consultas MDX (*Multi-Dimensional Expression*), sobre los modelos tabulares con funcionalidades similares.

Desde la perspectiva funcional el modelo tabular de SSAS brinda prácticamente las mismas prestaciones que el modelo multidimensional. Para los usuarios finales no existe diferencia respecto a si están consultado uno u otro modo informacional, pero las tecnologías subyacentes y las implementaciones por los desarrolladores son muy diferentes. En el modelo de datos tabular los conceptos básicos son heredados del modelo relacional, aunque debe aclararse que no juegan los mismos roles. Los conjuntos de datos se organizan en tablas y se establecen relaciones entre ellos mediante los valores de las columnas, lo cual lo diferencia del diseño de cubos y dimensiones de los modelos multidimensionales. Este enfoque –similar al de *PowerPivot*– lo hace más simple, flexible y más cercano a los usuarios finales, pero sin dudas es más engorroso manejarlo ante la presencia de requerimientos complejos (Russo et al., 2012; Maite Torres et al., 2015).

Los nuevos esquemas dimensionales correspondientes a la base de datos analítica (WI) se implementaron solo sobre el modelo tabular, utilizando la herramienta *SQL Server Data Tools (SSDT)* (Harinath et al., 2012). Se aprovecharon las facilidades que brinda el motor analítico en memoria *xVelocity (in-memory analytics engine)* –almacenamiento columnar, los algoritmos avanzados de compresión, procesador de consultas multiproceso– con vistas a lograr un acceso rápido a las bases de datos tabulares desde aplicaciones clientes como *Microsoft Excel* y *Microsoft Power View* (Russo, 2014; Maite Torres et al., 2015).

Numerosos indicadores de desempeño se calcularon para satisfacer las necesidades de los especialistas; vale mencionar la utilidad, los costos, la solvencia y la liquidez en el escenario financiero, así como el acumulado de vacaciones, el índice de ausentismo y el completamiento de la plantilla en recursos humanos. La incorporación de estos indicadores ha motivado a un grupo de usuarios no solo a asimilar las facilidades de las herramientas clientes sino también a plantear nuevos requerimientos de información, lo que ha conllevado a la sistematización de un procedimiento para facilitar la extensión de la solución instrumentada.

Para la implementación de la predicción del comportamiento de los indicadores de desempeño, en el marco de la primera aproximación a un proceso de minería de datos en CIMEX, se eligió el método de Series de Tiempo proporcionado por *Microsoft*. El objetivo fundamental del método es predecir series de variables continuas basándose en datos históricos y aprovechando convenientemente las bondades de las técnicas estadística y de clasificación. Debido al funcionamiento subyacente, el método de predicción utiliza dos algoritmos internos que pueden ser combinados inteligentemente: ARTxp (optimizado para predicciones a corto plazo) y ARIMA (optimizado para predicciones a largo plazo) (Singh Bassan & Sarkar, 2014).

En esta fase del desarrollo de la solución, se implementaron las vistas minables en el DWE, donde ya se cuenta con los datos preparados. Luego, dentro de un proyecto de tipo *Analysis Services Multidimensional and Data Mining* se definió para cada indicador la estructura minable, especificando las columnas que forman parte del proceso. Finalmente se aplicó el algoritmo, ajustando en algunos casos ciertos parámetros de configuración como: *Forecast_Method*, *Periodicity_Hint*, *Missing_Value_Substitution* y *Minimum_Series_Value*.

En la capa final de presentación de la información se suministran reportes gerenciales para las nuevas áreas de negocio integradas, que se publican tanto en SAN como en los cuadros de mando integrales implementados con *Power View* sobre *Share Point*. Para impulsar el análisis proactivo se utilizó *Excel* por sus facilidades analíticas con tablas dinámicas y *Power BI* para compartir los resultados y colaborar de manera intuitiva. Ambas alternativas no solo propician el autoservicio a la hora de generar reportes e investigar comportamientos a la medida de los intereses del usuario sino también respaldan la independencia respecto del especialista informático.

EXPERIMENTOS Y RESULTADOS

En la fase de experimentación se concibió e implementó un conjunto de pruebas que permitieran validar la solución propuesta en términos de eficiencia, cumplimiento de los requerimientos y presentación de los resultados. Resulta preciso aclarar que el *hardware* empleado es considerablemente superior que el utilizado durante la aproximación anterior de la solución de inteligencia de negocios (Maite Torres et al., 2015). Se contó con una PC con sistema operativo Windows 8.1 sobre una arquitectura de 64 bit, CPU QuadCore Intel Core i7-4790 K, 4000 MHz (4400 Mhz with Turbo Boost), 8Mb Cache L3, 2 Mb Cache L2, con 24 Gb (DDR3 SDRAM) 2400Mhz, 240 Gb SSD Kingston de almacenamiento (Simón et al., 2015).

Primeramente se repitieron las pruebas diseñadas en la investigación anterior para los escenarios contable y comercial sobre el modelo tabular, pero utilizando el *hardware* actual. En la tabla 1 se muestra la comparación de los resultados en el caso de la población del WI. Es de destacar que el procesamiento de la base de datos analítica con 60 veces más información se realizó en 6 veces menos tiempo, ocupándose solamente 4 veces más espacio. Se corroboró que con el modelo tabular se obtienen mejores resultados cuando se cuenta con elevados recursos, reafirmando que es un requerimiento imprescindible para su utilización con grandes volúmenes de datos, como también plantean distintos autores (Russo et al., 2012; Sirmon et al., 2013). La reproducción de todos estos experimentos permitió comprobar que los componentes de la solución precedente funcionan perfectamente en el marco de la presente extensión.

Tabla 1. Población del WI tabular con los escenarios precedentes

Parámetro medido	Solución anterior	Solución actual
Datos procesados	1 mes: marzo 2013	60 meses: 2010-2014
Tiempo total de ejecución	36 min.	9 min.
Tamaño de la BD tabular	230 MB	877 Mb

La segunda prueba consistió en analizar los tiempos de ejecución de la población inicial del DWE, considerando 15 meses de historia (enero 2014 - marzo 2015) de los nuevos módulos económico-financiero y de recursos humanos. La tabla 2 muestra los resultados cuantitativos de la carga de las dimensiones y los hechos. La base de datos reconciliados ocupó en su totalidad un espacio de 2.3 Gb. Se comprobó que los procesos ETL funcionan correctamente con tiempos aceptables.

Tabla 2. Población del DWE con los nuevos escenarios

Parámetro medido	Dimensiones	Hechos
Total de registros cargados	61 517	3 063 408
Tiempo total de ejecución	90 seg.	17 min.
Tamaño en disco	276 Mb	

Se concibió otro experimento con el objetivo de validar si la solución implementada responde al análisis informacional y a las expectativas de los usuarios con relación a la visualización de los resultados. Para ello, se diseñaron consultas dinámicas en correspondencia con los requerimientos informacionales y se elaboraron convenientemente informes y tableros de control en *Excel*, *Power BI* y *Power View* sobre *SharePoint*. En la figura 5a se muestra un *dashboard* que favorece el estudio

de las utilidades en una entidad dada teniendo en cuenta el cumplimiento del plan mensual y acumulado hasta la fecha y su comparación con el año anterior en gráficos de barra, así como el comportamiento histórico en un gráfico de línea. En la figura 5b se presenta el estado de un conjunto de indicadores generales por trimestre.

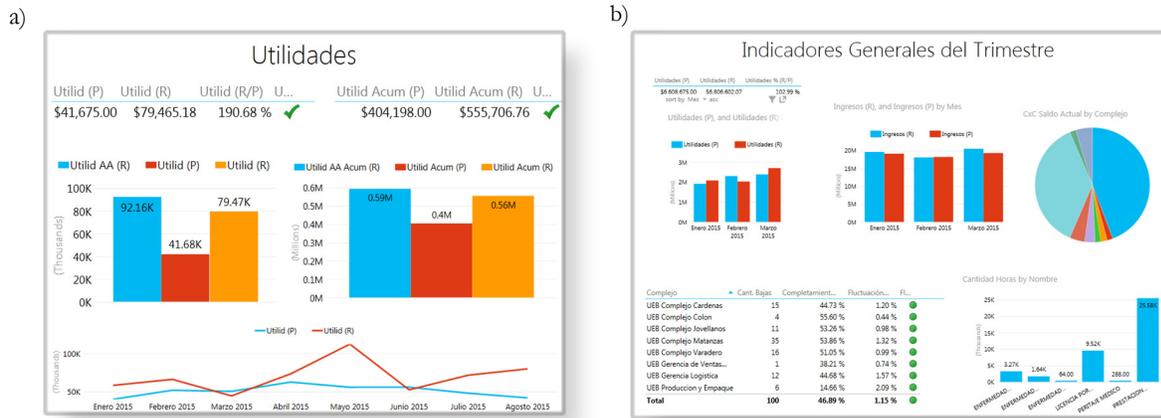


Fig. 5 Dashboards para el análisis de indicadores.

En la figura 6a es posible apreciar los KPI asociados a la fluctuación laboral y la estimulación de los trabajadores. En la figura 6b se propicia la exploración del comportamiento de los indicadores desde la perspectiva geográfica. La incorporación de riqueza visual para el análisis de los indicadores claves de desempeño en la solución actual ha constituido un incentivo para impulsar el empleo de la tecnología en la toma de decisiones.



Fig. 6 Dinamismo en la visualización de los indicadores.

Por último, se incluyó un experimento con el fin de predecir indicadores de desempeño para una entidad y verificar el proceso de minería propuesto. Fueron considerados 42 meses de historia (enero 2012 - junio 2015) y se estimaron 18 meses (julio 2015 – diciembre 2016). En la figura 7 se presenta el modelo resultante de aplicar el método de Series de Tiempo con el algoritmo ARTxp para la predicción del indicador Ingresos en una entidad, utilizando *Mining Model Viewer* de *SQL Server Data Tools*. Cabe observar que los valores reales se expresan con una línea continua y los valores estimados, mediante una línea discontinua. Las posibles desviaciones del modelo se muestran mediante barras verticales, lo cual expresa el grado de precisión de la predicción. El conocimiento acerca del modo en que los ingresos tienden a aumentar no existía con anterioridad.

De manera similar, fueron generados varios modelos y se comprobó que generalmente es necesario modificar los parámetros del algoritmo para obtener valores más certeros (Dean, 2014).



Fig. 7 Gráfico de predicción del indicador Ingresos para una entidad de CIMEX

La solución de inteligencia de negocios concebida e instrumentada en una nueva aproximación está orientada a evaluar el desempeño empresarial aprovechando las fortalezas del modelo de datos tabular. La inclusión de técnicas de minería de datos para la previsión del comportamiento futuro de los indicadores gerenciales contribuye a mejorar las capacidades y el rendimiento de la organización en el marco de la gestión del conocimiento.

CONCLUSIONES

El Grupo Empresarial CIMEX ha estado inmerso en un proceso evolutivo ascendente en el que no solo se ha dado respuestas oportunas y pertinentes al procesamiento de los datos operacionales, sino también se han desarrollado diversas soluciones de inteligencia de negocios propias que gradualmente han potenciado la gestión del conocimiento en la organización.

Se logró enriquecer el diseño conceptual del almacén de datos, incorporando nuevos sujetos de análisis y profundizando en el modelo tabular como alternativa de implementación del modelo de datos multidimensional. El aprovechamiento del modelo semántico de inteligencia de negocios de Microsoft constituye una práctica efectiva para satisfacer los requerimientos informacionales ante contextos muy variables, ya que brinda diversas formas de visualización y recreación de los resultados sobre los enfoques dimensionales.

Los usuarios se han apropiado con buena disposición de las herramientas analíticas que aportan funcionalidades dinámicas y han ganado cierta independencia de los profesionales informáticos al incorporarlas paulatinamente a la administración del negocio. La incursión en la predicción de comportamientos contribuye al avance del proceso de toma de decisiones oportunas en función de la anticipación de riesgos y el cumplimiento de los planes estratégicos.

En el futuro inmediato se proyecta continuar investigando en la aplicación consecuente del proceso de extracción de conocimiento con vistas a garantizar de manera preventiva la salud de la organización y el avance ascendente como tendencia. Como estrategia se pretende incursionar en *Big Data* como nuevo paradigma para el procesamiento de datos.

REFERENCIAS

- Abadi, D., Boncz, P., Harizopoulos, S., Idreos, S., and Madden, S. (2012). The Design and Implementation of Modern Column-Oriented Database Systems. *Foundations and Trends in Databases*, 5(3): 197–280.
- Abadi, D. J., Madden, S. R., and Hachem, N. (2008). *Column-Stores vs. Row-Stores: How Different Are They Really?* Paper presented at the 28th ACM SIGMOD/PODS International Conference on Management of Data / Principles of Database Systems, Vancouver, BC, Canada.
- Aggarwal, C. C. (2015). *Data Mining: The Text Book* (Vol. 14). Switzerland: Springer.
- Albano, A., Rosa, L. D., Dumitrescu, C., Goglia, L., Goglia, R., and Minei, V. (2006). *Another Example of a Data Warehouse System Based on Transposed Files*. Paper presented at the 10th International Conference on Advances in Database Technology.
- Andrew, R., and Sieber, S. (1999). La gestión integral del conocimiento y del aprendizaje. *Economía Industrial*, 326: 63-72.
- Aranibar, J. C. (2013). *Sistemas de Información Gerencial para la Administración del Desempeño Empresarial*. La Paz, Bolivia: Gráfica Holding s.r.l.
- Cattell, R. (2010a). High Performance Scalable Data Stores. <http://www.scribd.com/doc/27378891/High-Performance-Scalable-Data-Stores-Rick-Cattell>
- Cattell, R. (2010b). Scalable SQL and NoSQL Data Stores. www.cattell.net/datastores/Datastores.pdf
- Chessman, T. (2013). Understanding Time Series Forecasting Concepts. *SQL Server Pro*. <http://sqlmag.com/sql-server-analysis-services/understanding-time-series-forecasting-concepts>
- Dean, J. (2014). *Big Data, Data Mining, and Machine Learning* (Vol. 8). New Jersey and Canada: John Wiley & Sons, Inc.
- García, L., Simón, A., Torres, M., and Espinosa, Y. (2013). *Solución de inteligencia de negocios para la integración de la información comercial y contable*. Paper presented at the COMPUMAT 2013, Congreso Internacional de Matemática y Computación, La Habana, Cuba.
- Harinath, S., Pihlgren, R., Guang-Yeu Lee, P., Sirmon, J., and M. Bruckner, R. (2012). *Professional Microsoft SQL Server 2012 Analysis Services with MDX and DAX*. Indianapolis: John Wiley & Sons, Inc.
- Kaplan, R. S. (2010). *Conceptual Foundations of the Balanced Scorecard*. *Harvard Business School Press*, 10 no. 74.
- Kaplan, R. S., and Norton, D. P. (1992). *Balanced Scorecard - Measures that drive performance*. *Harvard Business Review*.
- Matei, G. (2010). Column-Oriented Databases, an Alternative for Analytical Environment *Database System Journal*, 1 no. 2.
- Myers, P. (2012). *Introducing the BI Semantic Model in Microsoft SQL Server*.
- Plattner, H. (2014). *The impact of columnar in-memory databases on enterprise systems: implications of eliminating transaction-maintained aggregates*. Paper presented at the VLDB Endowment.
- Russo, M. (2014). *SSAS Tabular as Analytical Engine*. *SQLBI*.
- Russo, M., Ferrari, A., and Webb, C. (2012). *Microsoft SQL Server 2012 Analysis Services: The BISM Tabular Model*. United States of America: O'Reilly.
- Simón, A., Torres, M., García, L., and Ravelo, R. (2015). *Extensión del Data Warehouse de CIMEX basado en el Modelo Tabular*. Paper presented at the COMPUMAT 2015, Congreso Internacional de Matemática y Computación, La Habana, Cuba.
- Singh Bassan, A., and Sarkar, D. (2014). *Mastering SQL Server 2014 Data Mining* (Vol. chapters 7, 8). United Kingdom: Packt Publishing.
- Sirmon, J., Galloway, G., Gross, C., and Gulati, K. (2013). *Performance Tuning of Tabular Models in SQL Server 2012 Analysis Services*. *SQL Server Technical Article*.
- Torres, M., Espinosa, Y., Simón, A., García, L., and Simón, A. (2014). *Modelación Tabular: una alternativa sugerente para el análisis de los datos*. Paper presented at the CIISI 2014, II

Congreso Internacional de Ingeniería Informática y Sistemas de Información, La Habana, Cuba.

Torres, M., Espinosa, Y., Simón, A., García, L., and Simón, A. (2015). Modelación Tabular: una alternativa sugerente para el análisis de los datos. *Revista Ciencias de la Información*, 46 no. 1(1): 3-10.