

Tipo de artículo: Artículo original
Temática: Informática Jurídica
Recibido: 30/10/2015 | Aceptado: 13/01/2016

Mercado de Datos para los departamentos de Procesos Penales y Gestión de Cuadros y Personal de Apoyo de la Fiscalía General de la República

Data Marts for departments Criminal Proceedings and Cadre Management and Support Personnel of the Prosecutor General's

Abel Andres Irsula Tumbarell ^{1*}, Manuel Alvarez Alonso ², Jose Carlos Pupo Acosta ³, Beatriz Danmara Hidalgo García ⁴

¹ Centro de Gobierno Electrónico. Facultad 3. Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, La Lisa, Torrens, La Habana. Cuba. {aairsula, malvareza, jcpupo, bdhidalgo}@uci.cu

* Autor para correspondencia: aairsula@uci.cu

Resumen

El presente trabajo permitió el desarrollo de diferentes mercados de datos para proporcionar el acceso a datos históricos de los departamentos Procesos Penales y Gestión de Cuadros y Personal de Apoyo de la Fiscalía General de la República. Estos mercados de datos surgen por la necesidad que presentan las fiscalías de poder almacenar datos de una manera más eficiente para su explotación y análisis. Se utilizaron en la construcción de la Base de Datos de los Mercados de Datos, los procesos Extracción, Transformación y Carga y en la carga de los cubos, herramientas como Mondrian Schema Workbench y Biserver-ce. Con esto se mejora y acelera el proceso de toma de decisiones en la Fiscalía General de la República, permitiendo elevar la efectividad de la organización. Se realizan pruebas de volumen, carga y estrés para validar la capacidad y rendimiento de los mercados de datos en consonancia con los requerimientos definidos por el cliente.

Palabras claves: mercado de datos, toma de decisiones, base de datos, almacén, Fiscalía

Abstract

This work allowed the development of different data marts to provide access to historical data for the departments Criminal Proceedings and Cadre Management and Support Personnel of the Prosecutor General's. These data marts

arise from the need to present the prosecution of storing data in a more efficient manner to exploitation and analysis. They were used in the construction of the Database data marts, processes of Extraction, Transformation and Load, and loading cubes tools like Mondrian Schema Workbench and biserver-ce. With this being improves and accelerates the process of decision making in the Prosecutor General's, allowing increase the effectiveness of the organization. Evidence of volume, load and stress are performed to validate the capacity and performance of market data in line with the requirements defined by the customer.

Keywords: *data marts, decision making, database, storage, prosecutor*

Introducción

El desarrollo del Sistema de Informatización de la Gestión de la Fiscalía (SIGEF), se encarga de informatizar los procesos de la Fiscalía General de la República (FGR). Entre los módulos se encuentran, Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva y Gestión de Cuadros y Personal de Apoyo (Fernández Pupo, y otros, 2011). El sistema transaccional existente no facilita la búsqueda de información histórica con agilidad, ni la realización de comparaciones con la totalidad de los datos significativos en un determinado intervalo en el tiempo, debido a la forma en que está organizada dicha información; además provoca que los reportes estadísticos que hoy se obtiene no cumplan con los tiempos de respuestas esperados por los directivos de la FGR. La dirección de la entidad y la de estos departamentos necesitan tener y acceder a un repositorio de datos histórico que sea capaz de apoyar al proceso de toma de decisiones.

Este proceso de toma de decisiones se genera de los datos como producto secundario, que son resultado de todas las transacciones que se realizan (Keller, y otros, 2011). Es muy común, que los mismos se almacenen y administren a través de sistemas transaccionales en bases de datos relacionales, como el utilizado en la aplicación SIGEF. La idea central que se persigue es que estos datos dejen de ser solo simples datos para convertirse en información que enriquezca las decisiones. La inteligencia de negocios (*Business Intelligence - BI*) resuelve este problema con la utilización de un almacén de datos o mercado de datos, siendo el segundo un subconjunto del primero (Rodríguez Cabanillas, y otros, 2011).

Se persigue como objetivo desarrollar mercados de datos que organicen la información existente en el sistema transaccional de los departamentos Procesos Penales y Gestión de Cuadros y Personal de Apoyo de la Fiscalía General de la República para disminuir los tiempos de acceso de los directivos de estas áreas a los datos contenidos en los reportes estadísticos

Materiales y métodos

1. Desarrollo teórico de los mercados de datos

Los almacenes de datos se han diseñado para superar algunos de los problemas que una organización encuentra cuando intenta realizar un análisis estratégico mediante la misma base de datos que utiliza para realizar el proceso de transacciones en línea (Nima, 2015). En muchas ocasiones los términos mercados de datos y almacén de datos son utilizados indistintamente, entre estos la diferencia es solamente en cuanto al alcance. Mientras que un almacén de datos es un sistema centralizado con datos globales de la empresa y de todos sus procesos operacionales, un mercado de datos es un subconjunto temático de datos, orientado a un proceso o un área de negocio específica propia de la organización (Ramos, 2011). El conjunto de módulos investigados apoya la creación de un mercado de datos para cada uno de estos.

De los diferentes modelos de almacenamiento de OLAP, acrónimo en inglés de procesamiento analítico en línea, y atendiendo a sus características se decide utilizar MOLAP, acrónimo en inglés de procesamiento analítico multidimensional en línea (Avila, 2015), para la realización de los mercados de datos. Este modelo permitirá tener los datos precalculados y luego podrán ser almacenados en cubos multidimensionales. El resultado se traducirá en un mejor tiempo de respuesta debido a que los datos se encuentran disponibles sin necesidad de calcularlos en cada nueva consulta (Ibarra, 2006).

2. Metodología y herramientas

Entre las metodologías conocidas para el desarrollo de los almacenes de datos se encuentran: la de Bill Inmon y la de Ralph Kimball. La descrita por Inmon propone construir primero el almacén de datos y luego los mercados de datos, por lo cual se debe conocer toda la información que posee el sistema (Inmon, 2014); mientras la propuesta por Kimball se destina a construir primeramente los mercados de datos que conformarán el almacén de datos final (Kimball, 2004), esta sería la elegida para cubrir la construcción de los mercados de datos.

Las herramientas utilizadas complementan la política de desarrollo de software con tecnologías libres. Visual Paradigm en su versión 8.0 herramienta para el modelado (Sierra, 2007); PostgreSQL 9.1 como sistema gestor de bases de datos; PgAdmin III 1.14.1 como aplicación gráfica para administrar el gestor de bases de datos (Ventura, 2014). De la plataforma Pentaho Open Source Business Intelligence, se utilizaron diferentes programas que satisfacen los requisitos de Inteligencia de negocio, estos son: Pentaho Bi Server-ce 4.8.0, aplicación 100% Java que permite gestionar todos los recursos de inteligencia de negocio; Pentaho Data Integration 4.2.1 encargada de limpiar e integrar la información, realizando el proceso de ETL (Extracción, Transformación y Carga); Mondrian Schema Workbench

3.2.0 sería el diseñador de interfaz gráfica que permite crear y probar los esquemas de cubos visualmente (Méndez, 2013).

Otras herramientas que complementan el resultado es el servidor web Apache Tomcat 5.5.12 y el Apache JMeter 2.9 (Baedke, 2015) como herramienta para medir el rendimiento del servidor de aplicaciones y así dimensionar adecuadamente la memoria del servidor y optimizar los recursos de la máquina (física o virtual) (Sanchez, 2013).

Resultados y discusión

1. Descripción de la solución

Para gestionar los datos correspondientes al sistema SIGEF se cuenta con una base de datos relacional. El departamento Índice de Peligrosidad Predelictiva tiene un total de 11 tablas de las cuales 3 son tablas nomencladoras y 8 son tablas de datos, el departamento Procesos Penales Ordinarios tiene un total de 49 tablas de las cuales 9 son tablas nomencladoras y 40 son tablas de datos, el departamento Procesos Penales Sumarios tiene un total de 26 tablas de las cuales 3 son tablas nomencladoras y 23 son tablas de datos y el departamento Gestión de Cuadros y Personal de Apoyo tiene un total de 29 tablas de las cuales 13 son tablas nomencladoras y 16 son tablas de datos.

Teniendo en cuenta las características generales de un mercado de datos se definen los componentes que intervienen en la arquitectura o ambiente. En la siguiente figura se muestra la estructura que tendrá cada uno de los mercados de datos.



Figura 1. Arquitectura de los mercados de datos

La arquitectura mostrada en la figura anterior está formada por varios elementos que interactúan entre sí y que cumplen una función específica que se resumen de la siguiente manera:

- Los datos son extraídos de la base datos SIGEF.
- Los datos son transformados, limpiados e integrados para luego ser cargados en el mercado de datos. Este proceso conocido como ETL es realizado con la herramienta *Pentaho Data Integration* (PDI).

- La información del mercado de datos es almacenada en cubos multidimensionales.

Modelado Dimensional

El modelado dimensional, es la forma en que se transforman los datos y son convertidos en información útil para los usuarios del negocio. El objetivo final es que estos puedan encontrar de manera intuitiva y rápida la información que necesitan. Para la creación del modelo se definirán los procesos del negocio a los cuales se les realizarán los mercados de datos. Se identificarán las dimensiones y se le establecerá el nivel de granularidad. Posteriormente se describirán las tablas dimensiones y hechos (Kimball, 2004).

Elegir el proceso de negocio

Las áreas de proceso a modelar serán los procesos de negocio que pertenecen a los departamentos Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva y Gestión de Cuadros y Personal de Apoyo.

Identificación de las dimensiones

Las dimensiones se obtuvieron mediante el análisis de los reportes de los departamentos Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva y Gestión de Cuadros y Personal de Apoyo y de la base de datos del sistema SIGEF.

Dimensiones comunes para los mercados de datos de los departamentos Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva.

Dimensión tiempo: Representa el espacio de tiempo en el que se está analizando cada acción realizada en la fiscalía, en este caso se analizarán los documentos diarios y en rangos de fechas.

Dimensión instancia: Define en que instancia de las fiscalías existentes de nuestro país se están obteniendo los datos, las cuales serían Fiscalía General de la República, Provincial y Municipal.

Dimensión indicador: Define el tipo de documento y los indicadores por los que se desea analizar.

Dimensión provincia: Contiene todas las provincias de Cuba.

Dimensiones para el mercado de datos del departamento Gestión de Cuadros y Personal de Apoyo.

Dimensión instancia: Define en que instancia de las fiscalías existentes de nuestro país se están obteniendo los datos, las cuales serían Fiscalía General de la República, Provincial y Municipal.

Dimensión provincia: Contiene todas las provincias de Cuba.

Dimensión cargo: Define los cargos que se pueden ocupar en la fiscalía.

Establecimiento de granularidad

La granularidad representa el nivel de detalle en el que se desea almacenar la información sobre el negocio. Mientras mayor sea el nivel de detalle de los datos, se tendrán mayores posibilidades de análisis. Al realizarse este paso es importante explicar el tipo de dato y los valores que pueden tomar estos.

Descripción las tablas dimensiones y hechos

Los hechos proporcionan información sobre el historial de las operaciones del negocio. Cada tabla de hechos posee un nombre que la identifica y contiene las claves de todas las tablas de dimensiones con las que se relaciona. Estas tablas se obtuvieron mediante el análisis de los reportes de los departamentos Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva y Gestión de Cuadros y Personal de Apoyo.

Como buenas prácticas en la base de datos de los mercados de datos resultantes fueron utilizados patrones. El patrón de llaves subrogadas se utiliza para generar una llave primaria única para cada tabla de dimensión en vez de usar un atributo identificador en el contexto dado (Blaha, 2010). Para el modelado de los mercados de datos de los departamentos Gestión de Cuadros y Personal de Apoyo, Procesos Penales Ordinarios, Índice de Peligrosidad Predelictiva se utiliza el patrón constelación de hechos, pues estos contienen más de una tabla de hechos. En el modelado del mercado de datos del departamento Procesos Penales Sumarios se selecciona el esquema en estrella debido a que la información se almacena en una sola tabla de hechos (Bernabeu, 2010).

Las figuras siguientes muestran los resultados obtenidos.

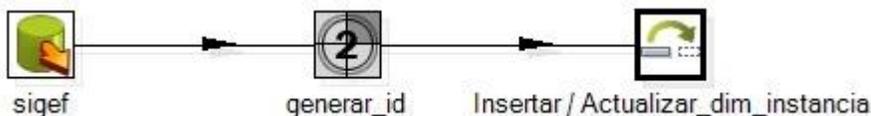


Figura. 2 Transformación para la dimensión instancia

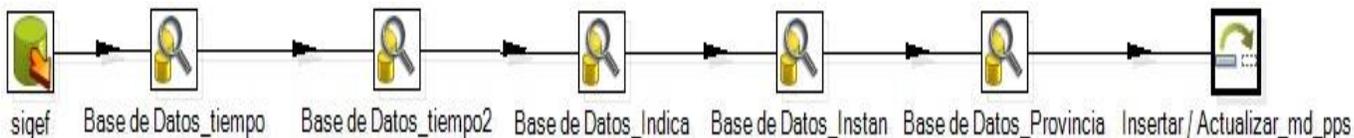


Figura. 3 Fragmento de transformación para el hecho denuncias recibidas en la fiscalía

		Medidas			
Fecha inicio	Fecha fin	● Cantidad en el Período	● Diferencia en el período del año anterior	● Acumulado	● Diferencia de acumulado en el año anterior
+ Total	+ Total	1.220.218	2.410.307	127.897	3.630.283
+ Total	+ Total	1.220.218	2.410.307	127.897	3.630.283
+ Total	+ Total	619.063	1.207.997	127.897	1.826.818
- Total	+ Total	68.781	134.216	14.210	202.971
- 2014	+ Total	68.781	134.216	14.210	202.971
- enero	+ Total	12.836	22.326	14.210	35.136
2014-01-01	- Total	2.351	1.356	14.210	3.681
	- 2014	2.351	1.356	14.210	3.681
	+ enero	245	123	1.617	376
	- febrero	249	137	1.473	362
	2014-02-01	22	7	90	17
	2014-02-02	8	9	61	13
	2014-02-03	12	9	57	16
	2014-02-04	8	2	53	4
2014-02-05	14	8	27	31	

Figura: 4. Fragmento de visualización de los datos del cubo Procesos Penales Sumarios.

Pruebas

En busca de mejorar los tiempos de respuesta esperados, se definió cargar los mercados de datos con un volumen superior al esperado y sobre estos realizar consultas de acceso a los datos que devuelvan la información solicitada para el análisis en estas áreas. Para garantizar la calidad de los mercados de datos se realizaron pruebas de rendimiento tales como: pruebas de carga, volumen y estrés, las cuales validarán el correcto funcionamiento de los mercados de datos.

Para realizar la prueba de volumen se utilizó una carga de generada de 5 millones de tuplas, la cual no provocó ningún problema al gestor. En las siguientes comprobaciones se recurrió a simular un plan de pruebas con una cantidad de usuarios conectados concurrentemente haciéndole peticiones al mercado de datos y a los tiempos de respuestas definidos en desarrollo; Aceptables entre 0 y 10 segundos y como No Aceptables para tiempos superiores a 10 segundos.

Las consultas utilizadas buscan obtener cantidades de procesos para un módulo, su ejecución a través de la herramienta JMeter, permitió crear un ambiente de concurrencia para situaciones críticas de búsqueda de información. Los tiempos obtenidos inferiores a 5 segundos demuestran que los mercados de datos son capaces de mantener el máximo de conexiones estimadas de manera concurrente sin presentar problemas.

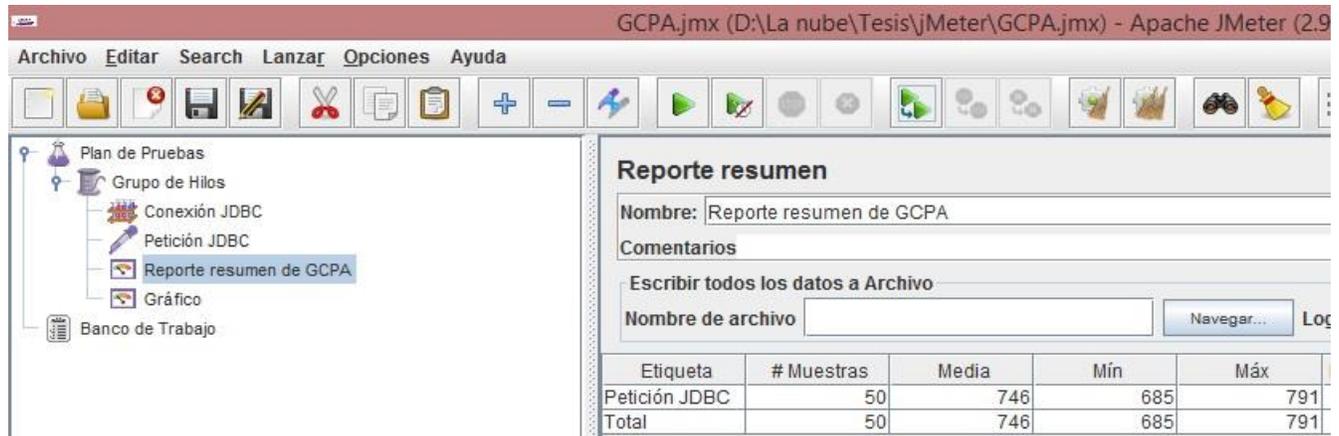


Figura. 5 Resultados del plan de prueba simulado

Conclusiones

Se conocieron los principales conceptos relacionados con los almacenes de datos; tales como metodologías y herramientas utilizadas para su desarrollo los cuales sirvieron de punto de partida para la comprensión de la investigación.

Se definió la metodología adecuada para solución, resultando ser la de Kimball la más apropiada atendiendo a las necesidades del negocio.

Se logró implementar y validar los mercados de datos, para esto último se utilizaron las pruebas de volumen, carga y estrés, arrojando resultados satisfactorios.

Referencias

- AVILA, KATTY. CAVSI. [En línea] [Citado el: 2 de 6 de 2015.] <http://www.cavsi.com/preguntasrespuestas/que-es-procesamiento-analitico-en-liea-olap/>.
- BAEDKE, MANFRED. 2015. The Apache software. *The Apache software*. [Online] 2015. <http://www.apache.org/>.
- BERNABEU, RICARDO DARIO. 2010. *Hefesto: Metodología para la construcción de un Datawarehouse*. Córdoba,Argentina : s.n., 2010.
- BLAHA, MICHEL. 2010. Patterns of Data Modeling. Emerging Directions in Database Systems and Applications. Taylor &. [book auth.] Michel laha. *Patterns of Data Modeling. Emerging Directions in Database Systems and Applications*. Taylor &. Atlanta,Giorgio,USA : Sham Navathe, 2010.
- FERNÁNDEZ PUPO, JAVIER, Y OTROS. 2011. *Implementación del Módulo Sumario del Proyecto Sistema de Informatización de la Gestión de la Fiscalía*. 2011.
- IBARRA, MARÍA DE LOS ANGELES. 2006. Procesamiento Analítico en Línea. Corrientes,Argentina : s.n., 2006.
- INMON, BILL. Dataprix. [En línea] [Citado el: 2 de 2 de 2015.] <http://www.dataprix.com/qu-es-un-data-warehouse>.
- KELLER RODRÍGUEZ CABANILLAS, ANGELA LUCIA MENDOZA PEÑA. 2011. Análisis diseño e implementación de una solución de inteligencia de negocios para el área de compras y ventas de una empresa comercializadora de electrodomésticos. Lima, Perú: s.n., 2011.
- KIMBALL, RALPH. 2004. *The Data Warehouse Lifecycle Toolkit*. New York : s.n., 2004.
- MARTINEZ, RAFAEL. 2013. PostgreSQL-es. *PostgreSQL-es*. [En línea] 2013. [Citado el: 2 de 9 de 2015.] http://www.postgresql.org.es/sobre_postgresql.
- MÉNDEZ, GEIDY ACOSTA. 2013. Informática en salud 2013. *Informática en salud 2013*. [En línea] 2013. <http://www.informatica2013.sld.cu/index.php/informaticasalud/2013/paper/view/195>.
- NIMA, JONATHAN DAVID RAMOS. 2015. EUMEDNET. *EUMEDNET*. [En línea] 2015. [Citado el: 01 de 05 de 2015.] <http://www.eumed.net/libros-gratis/2009c/574/TECNOLOGIA%20OLTP.htm>.
- RAMOS, SALVADOR. 2011. Microsoft Business Intelligence:vea el cubo medio lleno. *Microsoft Business Intelligence:vea el cubo medio lleno*. España: SolidQ Press, 2011.

SANCHEZ, ANTONIO. 2013. Servidores de Aplicaciones. *Servidores de Aplicaciones*. [En línea] 2013. [Citado el: 9 de 2 de 2015.] <http://blog.servidoresdeaplicaciones.com/2013/08/03/aprendiendo-a-usar-apache-jmeter/>.

SIERRRA, DANIEL. 2007. slideshare. *slideshare*. [En línea] 11 15, 2007. [Citado el: 2 de 8 de 2015.] <http://es.slideshare.net/vanquishdarkenigma/visual-paradigm-for-uml>.