

Tipo de artículo: Artículo original
Temática: Tecnologías de bases de datos
Recibido: 7/01/2013 | Aceptado: 1/03/2013

Mercado de datos para la dirección de cuadros de la Administración Provincial de Artemisa

Data mart for cuadros direction of the Provincial Administration of Artemisa

Yisel Valdés Rodríguez ^{1*}, Mayrin Ramos Maestre ², Jandy Miguel Gómez Rodríguez ³

¹ CIDI. Centro de Ideoinformática. Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Torrens, Boyeros, La Habana, Cuba. CP.: 19370

² TLM. Centro de Telemática. Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Torrens, Boyeros, La Habana, Cuba. CP.: 19370

³ CEDIN. Centro de Informática Industrial. Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Torrens, Boyeros, La Habana, Cuba. CP.: 19370

{[yvalrod](mailto:yvalrod@uci.cu), [mmaestre](mailto:mmaestre@uci.cu), [jgomez](mailto:jgomez@uci.cu)}@uci.cu

Resumen

Los almacenes de datos han progresado paulatinamente y son repositorios diseñados para facilitar la confección de informes y la realización de análisis para la toma de decisiones, los cuales se subdividen en unidades lógicas más pequeñas llamadas mercado de datos. Los mercados de datos son una versión del almacén que resuelve estudios a nivel de departamento, en específico para una necesidad de datos seleccionados. Durante el estudio preliminar de la situación existente en la dirección de cuadros, de la Administración Provincial de Artemisa, se identificaron algunos elementos tales como: el análisis de la información recibida se realizaba de forma manual, poca accesibilidad de la información y la elaboración de informes costosos en esfuerzo y tiempo, esto dificulta la disponibilidad de información estadística y el análisis de los datos por el personal administrativo. El desarrollo del mercado de datos fue guiado por la Metodología de Proceso de Desarrollo en la Línea Soluciones de Almacenes de Datos e Inteligencia de Negocio, la cual es mixta pues reúne elementos de varias metodologías de desarrollo de proyectos de integración de datos. Se utilizaron herramientas de modelado e inteligencia de negocio así como un sistema gestor de base de datos, dando como resultado la implementación del mercado de datos que será capaz de centralizar la información referente a las especialidades, así como realizar vistas de análisis, reportes del negocio y reportes ad-hoc, lo que contribuye a la disponibilidad, accesibilidad de la información y apoyo a la toma de decisiones.

Palabras clave: Accesibilidad, almacén de datos, disponibilidad, mercado de datos, toma de decisiones.

Abstract

Data warehouses have progressed gradually and are repositories designed to facilitate the preparation of reports and analysis for decision making, which are subdivided into smaller logical units called data marts. Data Marts are a warehouse version that solves department level studies. It is specific for a selected data need. During the preliminary study of the situation in the cuadros direction, of the Provincial Administration of Artemisa, some elements such as were identified: analyzing the received information is performed manually, inaccessibility of information and the costly reporting effort and time. This makes difficult the availability of statistical information and data analysis by the

administrative staff. The development of Data Mart was guided by Development Process Methodology in Data Warehousing Solutions from Business Intelligence, which is mixed because it gathers it brings together elements several development methodologies of data integration projects. Modeling and business intelligence tools were used and also a database manager given as resulting data mart implementation, that will be able to centralize information concerning specialties, also perform analysis views, business reports and ad-hoc reports, contributing to the availability, accessibility of information and support for decision making.

Keywords: *Accessibility, availability, data warehouse, data mart, decision making.*

Introducción

El nuevo marco globalizado del mundo actual y sus omnipresentes, imprescindibles y poderosas herramientas de las Tecnologías de la Información y las Comunicaciones (TIC), está induciendo una profunda revolución en todos los ámbitos sociales. Esto provoca el constante perfeccionamiento de las empresas en cuanto a tecnologías para su uso y comercialización. La principal arma que se ha desarrollado para contrarrestar esta competencia son los sistemas de análisis de información que permiten estudiar los datos históricos y actuales.

Los almacenes de datos surgen por la necesidad de resolver problemas de análisis de grandes masas de información, estos se subdividen en unidades lógicas más pequeñas dependiendo del subsistema de la entidad del que procedan o para el que sea necesario; dichas unidades lógicas se denominan Mercado de Datos (MD), los cuales resuelven estudios a nivel de departamento. Un MD es una versión del almacén de datos, la diferencia principal es que la creación de un MD es específica para una necesidad de datos seleccionados, enfatizando el fácil acceso a una información relevante y el análisis de datos estadísticos que apoyan la toma de decisiones.

El análisis de la estadística es fundamental en el control de los datos para la toma de decisiones, especialmente las relacionadas con el gobierno y los órganos administrativos, en el caso del control de los datos relacionados con los cuadros, constituye una tarea estratégica debido a que hay que garantizar el completamiento de los cuadros y la política de cuadros para un funcionamiento óptimo del sistema.

La dirección de cuadros tiene como misión: auxiliar al presidente del consejo en la aplicación de la política de cuadros aprobada por el Partido Comunista de Cuba y el Gobierno.

La información referente a esta dirección, es recogida a través de diferentes modelos, finalmente se almacena la información en aplicaciones informáticas de oficina (Excel, Word) y documentos impresos. Para apoyar la toma de decisiones, es necesario llevar un control de los datos que llegan a la dirección de cuadros de los diferentes municipios de la provincia en un mismo fichero y de fácil acceso al personal autorizado, de modo que en el proceso de elaboración de informes, el costo de tiempo y esfuerzo sea el menos posible al realizar las consultas a un gran cúmulo de información de las diferentes especialidades; por lo que se hace necesario que la información se encuentre en un sistema central, donde se evite la pérdida y duplicado de la misma, y el acceso a ella sea según los permisos definidos por la administración.

El objetivo de la presente investigación consiste en desarrollar un MD que contribuya a la disponibilidad de la información, así como facilitar el proceso de análisis de los datos para mejorar la toma de decisiones en la dirección de cuadros de la Administración Provincial de Artemisa.

Materiales y métodos

La investigación fue guiada por los métodos teóricos: Histórico – lógico para analizar la secuencia lógica desde el surgimiento de los almacenes de datos hasta la actualidad. La trascendencia mundial, en Cuba y más específicamente

en la Universidad de las Ciencias Informáticas (UCI); Analítico – sintético se utilizó para el análisis de documentos, materiales, y temas relacionados con las mejores prácticas en el desarrollo de MD. Los métodos empíricos ajustados a la investigación fueron: Entrevista, este método fue utilizado mediante una conversación previamente definida por el investigador y acordada con el entrevistado, con el objetivo de obtener la mayor cantidad de información posible para diagnosticar y detectar los problemas actuales de la dirección y la encuesta se realizó mediante un cuestionario que enmarca una serie de interrogantes específicas, sin la intervención directa del investigador, donde las respuestas son seleccionadas de acuerdo con su criterio y se determinaron sus posibles variantes, lo que facilitó su procesamiento estadístico (Hernández, 2002).

En este epígrafe se expresa brevemente la información recolectada y analizada de los documentos estudiados para realizar el MD.

El mundo actual evoluciona constantemente y se ha creado un desarrollo en vista a los cambios en el ámbito informático, actualmente un almacén de datos (o *Data Warehouse*): "Es una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis. Es la unión de todos los mercados de datos de una entidad" (Kimball, 1998).

Una de las definiciones más aceptadas es la de Bill Inmon que fue uno de los primeros autores en escribir sobre el tema de los almacenes de datos, define un almacén de datos en términos de las características del repositorio de datos:

- ⌘ Orientado a temas: Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.
- ⌘ Variante en el tiempo: Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.
- ⌘ No volátil: La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas.
- ⌘ Integrado: La base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes (Inmon, 2002).

Un MD: "Es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica" (Kimball, 2002).

Metodología a utilizar

Para la creación de un MD es imprescindible la etapa de análisis y diseño, en esta fase se conocen las necesidades de los clientes para así obtener un sistema que responda a los intereses del negocio. Durante el análisis se sientan las bases para los posteriores procesos de diseño e implementación. Realizar este proceso necesita un estudio del negocio para entender de manera clara y transparente lo que el usuario necesita. En el transcurso de esta se generan un conjunto de artefactos que facilitan el desarrollo del sistema. Se tiene en cuenta el levantamiento de los requerimientos, creando una guía para los desarrolladores en la fase de implementación. Se elabora además el diagrama de diseño de la base de datos, donde se definen las relaciones entre los hechos y dimensiones, así como los diagramas de Casos de Uso (CU). Se especifican los actores del negocio y del sistema, y su relación con los diferentes CU. En el diseño es donde se transforman los modelos lógicos conseguidos en la fase de análisis a modelos físicos, se realiza el modelo de datos, se construye la matriz bus y se muestra el modelo de diseño realizado.

La integración de datos es el proceso que organiza el flujo de la información entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para cargar datos desde múltiples fuentes a un MD (Rivadera, 2010).

La Metodología propuesta para desarrollar el MD es la “Metodología de Proceso de Desarrollo en la Línea Soluciones de Almacenes de Datos e Inteligencia de Negocio”. Es una metodología mixta que reúne elementos de varias metodologías de desarrollo de proyectos de integración de datos, ya que cubre todas las fases por las que pasa su construcción, desde el levantamiento de información inicial, la arquitectura, el diseño, la implementación de la herramienta de inteligencia de negocio, prueba, despliegue, soporte, hasta la gestión del proyecto.

Técnica de almacenamiento

Dentro de los sistemas gestores de base de datos podemos mencionar las bases de datos OLTP (Procesamiento de transacciones en línea) y las bases de datos OLAP (Procesamiento analítico en línea). Se utilizó el sistema OLAP en la categoría de ROLAP (Relacional OLAP): Almacenamiento en un esquema en estrella (no normalizado) o copo de nieve (normalizado) ya que maneja una arquitectura de tres niveles (Galemmo, 2003).

La base de datos relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica. El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y obtención del dato. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.

Inteligencia del negocio

Sistemas de inteligencia de negocios: La Inteligencia de Negocios, (del inglés, *Business Intelligence*, BI, por sus siglas en inglés), tiene como objetivos la generación de información estratégica-gerencial e histórica, así como su despliegue y difusión entre los usuarios. Las aplicaciones de BI son herramientas de soporte de decisiones que permiten en tiempo real, acceso interactivo, análisis y manipulación de información crítica para la empresa (Thornthwaite, 2006).

Las aplicaciones de BI son la cara visible de la inteligencia de negocios, los informes y aplicaciones de análisis proporcionan información útil a los usuarios. Las aplicaciones de BI incluyen un amplio espectro de tipos de informes y herramientas de análisis, que van desde informes simples de formato fijo a sofisticadas aplicaciones analíticas que usan complejos algoritmos e información del dominio.

Modelo de almacenamiento de datos

El modelo de almacenamiento utilizado para este tipo de solución es el modelado dimensional, compuesto por dimensiones, las cuales representan categorías de información, atributos que representan un único nivel dentro de una dimensión y por último tablas de hechos, las cuales contienen datos de interés que presentan un nivel de granularidad (Kimball, *et al.*, 2002).

Para la realización del MD de la dirección de cuadros se utilizó el esquema en estrella pues posee mejor rendimiento y velocidad. La tabla de hechos representa un evento del negocio y en ella se almacenan un conjunto de medidas o atributos, que permiten medir el rendimiento en los diferentes procesos del mismo. Mientras, las de dimensión, no son más que los filtros que podemos aplicar a nuestros datos, tanto filas como columnas.

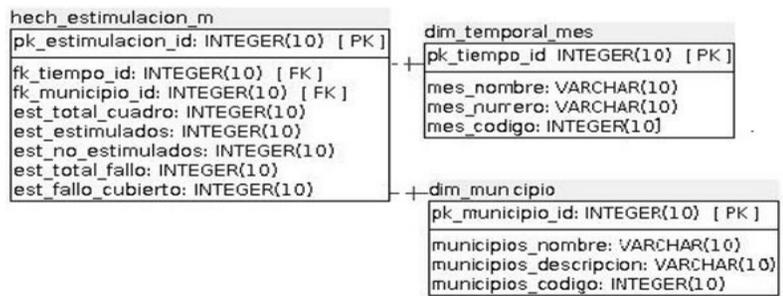


Figura. 1. Esquema en estrella. Hecho estimulación municipios.

Estudio preliminar del negocio

En la presente investigación se propone un producto que apoya a los procesos que se realizan en la dirección de cuadros de la Provincia Artemisa, relacionados con información referente a las cinco especialidades agrupadas en 8 temas de análisis. Por tal motivo, fue de vital importancia identificar las necesidades de información de los especialistas de dichas áreas, pues constituyen la base para un correcto diseño del MD. En la dirección de cuadro se atienden los procesos relacionados con la estadística, estimulación, evaluación, chequeo médico, superación, reserva, los movimientos de los cuadros y los balances generales de los cuadros ya sea los que pertenecen a las 32 direcciones de la administración provincial o a los municipios.

Para lograr satisfacer las necesidades del cliente precisadas con anterioridad se identificaron 6 requisitos funcionales, 22 requisitos informativos y 32 requisitos no funcionales simbolizando aquellas características del sistema que le reportan al cliente ventajas, como el rendimiento y la fiabilidad, dándole más confianza y seguridad en la aplicación.

Resultados y discusión

Durante la etapa de análisis y diseño del MD, fueron definidos los hechos y las dimensiones de los diferentes procesos que se convertirían en las tablas de la base de datos. Para el MD de la dirección de cuadros se definieron dos esquemas principales, el primero para almacenar las dimensiones comunes de todas las direcciones de la administración y un segundo esquema para las dimensiones y hechos específicos de la dirección.

Tabla. Esquemas y tablas de la base de datos.

Esquemas	Tablas que lo componen
dimensiones	dim_temporal_mes, dim_municipios
mart_cuadro	dim_direcciones, chequeo_medico, estadistica_direcciones, estadistica_municipios, estimulacion_direcciones, estimulacion_municipios, evaluacion_direcciones, evaluacion_municipios, medidas_evaluacion_direcciones, medidas_evaluacion_municipios, movimientos_cuadros, reserva_direcciones, reserva_municipios, superación_direcciones, superación_municipios

El total de hechos definidos para un MD, indica la cantidad de cubos OLAP que deben ser creados para la visualización de la información. En correspondencia con los hechos definidos para la dirección de cuadros, se implementaron 32 cubos OLAP, 16 de las tablas de hecho y 16 de las vistas materializadas.

Implementación de los subsistemas de integración

El proceso ETL (extracción, transformación y carga) se hace necesario para eliminar errores que pueden dañar los reportes generados, es la base sobre la cual se alimenta el almacén de datos.

Extraer: La primera parte del proceso ETL consiste en extraer los datos desde los sistemas de origen.

Transformar: La fase de transformación aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados.

Carga: La fase de carga es el momento en el cual los datos de la fase anterior son cargados en el sistema de destino del MD para la dirección de cuadros dependiendo de los requerimientos de la organización. (Kimball, *et al.*, 2004).

En la figura 2 se describe de forma general la estrategia de integración definida para realizar los procesos de ETL, correspondientes a los hechos y dimensiones del MD. Se extraen los datos desde el sistema de gestión de información.

Una vez efectuada la extracción de los datos, se realizan las validaciones necesarias teniendo en cuenta las reglas del negocio identificadas, mediante un código Java se calculan los datos y se filtran las filas. Si se comprueba que los datos poseen la calidad requerida, se elimina por la fecha para que no existan valores duplicados, se ordenan las filas,

se agrupa mediante el tipo de consulta y se procede a su inserción en la tabla de la base de datos señalada, en caso contrario de existir alguna falta los datos se guardan en un tabla error.

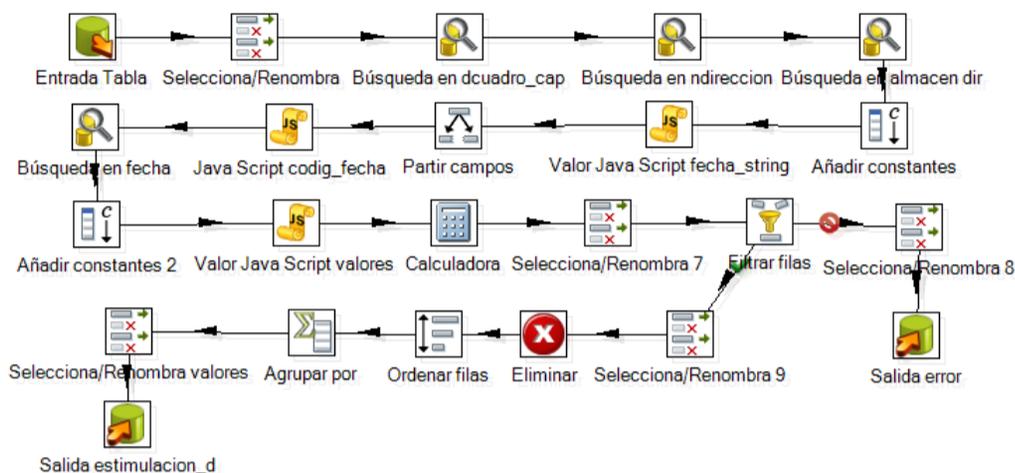


Figura 2. ETL. Hecho estimulación de las direcciones.

Implementación de los trabajos (Jobs)

Un Job o trabajo, es una tarea que es asignada para que se realice a determinada hora o fecha, es el encargado de ejecutar todas las transformaciones realizadas, hay diversas formas de hacer un Job, esto de acuerdo a las necesidades o ejecutar procedimientos almacenados. Una vez que la conexión al MD se encuentra en perfecto estado, se procede a la carga del mismo, y el Job o trabajo es la forma en que se realiza la carga de los datos hacia el mercado.

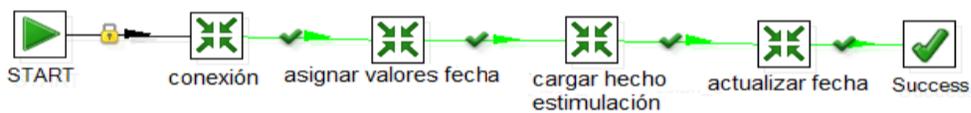


Figura 3. Job para la transformación estimulación de las direcciones.

Una vez realizado todos los trabajos necesarios se realiza una Job general para cargar los datos hacia todas las tablas de hechos correspondientes.



Figura 4. Job para las transformaciones de la base de datos.

Implementación del subsistema de visualización de datos

En correspondencia con los requisitos de información, se definieron 22 reportes agrupados en 8 libros de trabajo (LT) ubicados dentro de 8 áreas de análisis (AA). Dichas áreas se corresponde con las áreas de las cuales se van a analizar los reportes, mientras que los LT representan las diferentes categorías a las que pueden pertenecer los reportes. Cuando se accede al reporte deseado se pueden observar las vistas de análisis creadas.

Con el fin de generar reportes que precisan ser entregados para su análisis, fue necesario realizar reportes basados en distintas tablas construidos con la herramienta Pentaho Report Designer la cual pertenece a la suite de Pentaho.

Es posible que el usuario construya sus propios reportes a partir de la información de los metadatos definidos en el sistema con el uso de la herramienta Pentaho Metadata Editor. Los metadatos son los nombrados reportes ad-hoc, y se clasifican como reportes inmediatos que el usuario puede crear al instante (Pentaho, 2012).

Para ello se construyeron 16 metadatos.



Figura 5. Fragmento del mapa de navegación.

Pruebas

En el desarrollo de cualquier producto de software se realizan diferentes actividades desde que surge la idea inicial hasta la obtención del producto final. En la etapa de prueba del MD se desarrollan una serie de comprobaciones la cual son un conjunto de acciones en las que un sistema o componente es ejecutado bajo unas condiciones específicas, donde los resultados son observados y registrados para dar una evaluación de algún aspecto del sistema.

Se realizaron las pruebas a nivel de desarrollador: Donde se probó que se aplicaran apropiadamente cada regla del negocio y que los resultados esperados ocurriesen cuando se usaran datos válidos. Pruebas de acceso a la información por roles o niveles de usuario: Cumpliendo con los privilegios de acceso establecidos por la dirección.

Diseño de los casos de pruebas: Para la conformidad del MD de la dirección de cuadros, se diseñaron 8 casos de prueba por casos de uso y 22 casos de prueba por requisitos, con el propósito de verificar los requisitos, agrupados en ocho casos de uso del sistema que fueron definidos previamente durante la etapa de análisis.

Resultados y funcionalidades obtenidas

El MD para la dirección de cuadros brinda una vía eficiente y rápida de realizar el trabajo de los especialistas consultando gran cúmulo de información de las diferentes áreas, favorece el desempeño profesional, lo que permite facilitar la toma de decisiones de la dirección de cuadros, esta propuesta ahorra recursos dando la posibilidad de que la información se administre utilizando las nuevas tecnologías y permitiendo el acceso a esta en el momento preciso, evita la pérdida de tiempo que ocasionaba el análisis que hasta este momento realizaban los especialistas manualmente.

Como resultado de la propuesta se tiene disponible en su versión 1.0 el MD para la dirección de cuadros de la Administración Provincial de Artemisa. Se obtuvo un sistema que cumple con todas las especificaciones, para garantizar la accesibilidad, calidad, disponibilidad y el análisis de la información en apoyo a la toma de decisiones en la dirección. Entre las funcionalidades más significativas del MD se pueden señalar:

- ⌘ Permite crear vistas de análisis.
- ⌘ Permite realizar los reportes del negocio y crear reportes ad-hoc.

Proporciona al usuario una interfaz consolidada, única para los datos, que hace más fácil el trabajo con las consultas para la toma de decisiones

Conclusiones

Con esta investigación se abordaron los aspectos teóricos relacionados con almacenes y MD. Se definió seguir la Metodología de Proceso de Desarrollo en la Línea Soluciones de Almacenes de Datos e Inteligencia de Negocio para especificar las fases de construcción del MD. Se realizó el análisis y diseño correspondiente al MD, del cual se obtuvo las tablas de hechos que describen los procesos de los cuales se analizará la información, así como las vistas materializadas que permiten mostrar la información de manera rápida y concreta permitiendo la implementación de

todas las transformaciones y trabajos necesarios se pobló el MD. Las vistas de análisis y los reportes obtenidos a partir del uso de buenas prácticas, permiten la disponibilidad de los datos para su posterior análisis en apoyo a la toma de decisiones. Con las pruebas realizadas se pudo evaluar el cumplimiento de las funcionalidades del MD a partir de los requisitos adquiridos por el cliente. Al quedar validado el mercado se puede proceder su despliegue en la Administración Provincial de Artemisa para ser utilizado por los especialistas de la dirección de cuadros.

Referencias

- GALEMMO, I. &. *Mastering Data Warehouse Design: Relational and Dimensional Techniques*. s.l.: Wiley Publishing, 2003.
- HERNÁNDEZ, A. COELLO. *El paradigma cuantitativo de la investigación científica*. La Habana: Editorial universitaria, 2002.
- INMON, B. *Building the Data Warehouse (Third Edition)*. s.l.: John Wiley & Sons, 2002.
- KIMBALL, R. *The Data Warehouse Lifecycle Toolkit*. s.l.: Ed. John Wiley, 1998.
- KIMBALL, R. y CASERTA, J. *The Data Warehouse ETL Toolkit Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. s.l.: Wiley Publishing. ISBN: 0-764-57923-1.
- KIMBALL, R. y ROSS, M. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition)*. New York: Wiley, 2002.
- KIMBALL, R. y ROSS, M. New York: Wiley Computer Publishing. 0-471-20024-7.
- PENTAHO. [en línea] [Consultado el: 12 de octubre de 2012]. Disponible en: [www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/].
- RIVADERA, G. R. *Integración de datos*. [en línea] [Consultado el: 21 de mayo de 2012]. Disponible en: [<http://www.ucasal.net/templates/unid-academicas/ingenieria/apps/5-p56-rivadera-formateado.pdf>].
- THORNTHWAITE, M. &. *The Microsoft Data Warehouse Toolkit—With SQL Server 2005 and the Microsoft Business Intelligence Toolset*. Indianapolis: Wiley, 2006.