

Soluciones de inteligencia de negocios en la práctica: apoyo a la toma de decisiones en proyectos educativos para población infantil vulnerable en el caribe Colombiano

ALEXANDER ARMANDO BUSTAMANTE MARTÍNEZ

*Ingeniero de Sistemas, Universidad del Magdalena
Estudiante de Maestría en Ingeniería de Sistemas e Informática – Universidad Industrial de Santander
abustamante@unimagdalena.edu.co*

ERNESTO AMARU GALVIS LISTA

*Ingeniero de Sistemas, Magister en Informática
Profesor Asistente, Universidad del Magdalena
egalvis@unimagdalena.edu.co*

MAYDA PATRICIA GONZÁLEZ - ZABALA

*Ingeniera de Sistemas, Magister en Informática, PhD (C) en Ingeniería de Sistemas y Computación
Becaria del Departamento Administrativo de Ciencia y Tecnología de Colombia
Profesora Asistente, Universidad del Magdalena
mpgonzalez@unimagdalena.edu.co*

ARYS ALBERTO GARCÍA AVENDAÑO

*Ingeniero de Sistemas, Universidad del Magdalena
Estudiante de Maestría en Telecomunicaciones – Universidad Nacional de Colombia
aagarciaa@unal.edu.co*

LUIS FERNANDO BENAVIDES FAWCETT

*Ingeniero de Sistemas, Universidad del Magdalena
Estudiante de Especialización en Desarrollo de Software – Universidad del Magdalena
ibenavides@unmagdalena.edu.co*

*Fecha de recibido: 23/08/2011
Fecha de aprobación: 15/12/2011*

RESUMEN

La inteligencia de negocios se utiliza para dar respuesta a problemas relativos al acceso, la disponibilidad y la utilidad de la información. Este artículo presenta el proceso y los resultados obtenidos en la construcción de una solución de inteligencia de negocios de apoyo a la toma de decisiones en proyectos educativos para población infantil vulnerable en el Caribe colombiano. El contexto de la aplicación es la ejecución del proyecto Círculos de Aprendizaje, modelo educativo operado por la Universidad del Magdalena desde el año 2006, cuyo objetivo fundamental es vincular al sistema escolar a niños desplazados por la violencia y/o en condiciones de vulnerabilidad. Para construir la solución se utilizaron como guías metodológicas el modelo de Moss y Atre y el método ágil Programación Extrema. La principal implicación práctica de este trabajo es la de mostrar la utilidad de estas tecnologías en contexto diferentes al de los negocios.

PALABRAS CLAVE: Inteligencia de Negocios, Bodegas de Datos, OLAP.

ABSTRACT

Business intelligence is used to bring answers to problems related to access, availability and utility of information. This paper presents the process and results obtained in the construction of a business intelligence solution to support decision making in educational projects to vulnerable childhood population in the Colombian Caribbean region. The application context is the execution of a project named "Learning Circles", educational model operated by University of Magdalena since 2006, whose fundamental objective is to bring back to school to refugee and vulnerable children. To develop the solution we used as methodological guidelines Moss and Atrre's model and the agile method named Extreme Programming. The main practical implication of this work is to show the usefulness of these technologies in different context to business.

KEYWORDS: : Business Intelligence, Data Warehouse, OLAP.

1. INTRODUCCIÓN

Con el fin de garantizar que la población infantil en condición de vulnerabilidad y desplazamiento tenga acceso a la educación básica, el Ministerio de Educación Nacional de Colombia – MEN implementa modelos educativos flexibles como el denominado “Círculos de Aprendizaje” – CA [1][2][3]. El objetivo principal del modelo CA es lograr la vinculación al sistema de educación formal de niños, niñas y adolescentes - NNA con edades entre 5 y 16 años, que han sido afectados por problemas de desplazamiento forzado y vulnerabilidad. Esto se da a través de un proceso de adaptación, transición y acompañamiento. Los CA son ambientes de aprendizaje que se integran al contexto de cada comunidad y cuentan con el acompañamiento de profesionales en educación, psicología y trabajo social. Además de tener el acompañamiento y la vinculación a una Institución Educativa Madre – IEM, lo cual garantiza que se haga la transición de los NNA a la educación formal [4].

Para la implementación, expansión y fortalecimiento de los Círculos de Aprendizaje en la región Caribe colombiana, el MEN ha trabajado en alianza con la Universidad del Magdalena los últimos cinco años. La ejecución de este proyecto ha sido posible gracias a una inversión que supera los cinco millones de dólares, lo cual ha permitido atender a más de 6.000 NNA. Así mismo, se ha contado con un equipo, de más de 300 personas, conformado por un coordinador general, uno pedagógico, uno psicosocial, seis coordinadores departamentales, asesores pedagógicos, asesores psicosociales, tutores, auxiliares e ingenieros, los cuales se distribuyen en las localidades donde se lleva a cabo el Proyecto Círculos de Aprendizaje - PCA.

En el caso específico de la Universidad del Magdalena el equipo se encuentra conformado por una coordinación general que integra las coordinaciones departamentales,

en las cuales se distribuyen los CA con sus respectivos asesores y tutores pedagógicos y psicosociales. La dinámica del PCA requiere que para cada uno de los NNA pertenecientes a los CA se registre información de tipo personal, familiar, académica y psicosocial. Dicha actividad es realizada por los tutores pedagógicos y psicosociales que están a cargo de los NNA. El registro de información se realiza en la Plataforma de Gestión y Seguimiento de los componentes pedagógico y comunitario del Modelo Círculos de Aprendizaje - GYS-MCA desarrollada por la Coordinación de Tecnología de Información vinculada al proyecto. Por su parte, los asesores supervisan, complementan y corrigen la información registrada por los tutores a su cargo.

De otro lado, los coordinadores son los encargados de procesar la información con el fin de generar reportes orientados al seguimiento y desempeño de los NNA, así como el cumplimiento de los objetivos del proyecto [5]. En este sentido, la ejecución del PCA presenta dos retos concernientes a la administración de la información: i) la gestión del proyecto y el procesamiento de las transacciones en línea, considerando que el proyecto se opera en siete departamentos de la costa Caribe; y ii) la generación de consultas y reportes, a partir de la información registrada, que permitan apoyar la toma de decisiones. Tales retos motivaron el desarrollo de un proyecto orientado a proponer una solución fundamentada en Inteligencia de Negocios (*Business Intelligence* – BI) que permitiera apoyar el proceso de toma de decisiones y de gestión del PCA operado por la Universidad del Magdalena.

A continuación se presenta la solución propuesta, para lo cual se inicia con una síntesis de los conceptos teóricos que fundamentan el proyecto, seguidamente se describe el proceso metodológico que se siguió en el proceso investigativo y los resultados obtenidos; y finalmente se presentan las conclusiones obtenidas y los referentes bibliográficos que sustentan el presente documento.

2. FUNDAMENTACIÓN TEÓRICA Y METODOLÓGICA

2.1 Inteligencia de Negocios

El término inteligencia de negocios ha tenido diferentes acepciones durante los últimos años, un pionero en la creación del término, fue Richard Green, quien en 1966 lo definió como “la información procesada, de interés para la administración, acerca del presente y futuro del entorno en el cual el negocio debe operar” [6]. Para Howard Dresner [7], la inteligencia de negocios es “... un conjunto de conceptos y métodos para mejorar el proceso de decisión utilizando un sistema de soporte basado en hechos...”. Por otra parte, para el Instituto de Bodegas de Datos (The Data Warehouse Institute - TDWI), la inteligencia de negocios es un término sombrilla, es decir, la combinación de tecnología, herramientas y procesos que permiten transformar los datos almacenados en información, esta información en conocimiento y este conocimiento aplicarlos a un plan o una estrategia comercial. En este sentido, la inteligencia de negocios debe ser parte de la estrategia empresarial pues permite optimizar la utilización de recursos, monitorear el cumplimiento de los objetivos de la empresa y la capacidad de tomar buenas decisiones para así obtener mejores resultados [8].

Tomando como base la definición del TDWI, se concibe a la inteligencia de negocios para el presente proyecto como el grupo de tecnologías, técnicas, conceptos, herramientas y métodos que permiten aprovechar los datos históricos y actuales, del PCA y su entorno, para soportar la toma de decisiones mediante reportes, proyecciones, y predicciones. Propiciando así, que los datos manejados contribuyan al logro de los objetivos del proyecto.

En este sentido, es importante especificar que la solución desarrollada en este proyecto involucró tecnologías, técnicas, conceptos y herramientas como sistemas OLTP o bases de datos operacionales, bodegas de datos, procesos de ETL, aplicaciones OLAP y herramientas de reportes. La relación entre estos componentes se puede apreciar en la Figura 1. Además, en el siguiente numeral se describe de forma sintética cada uno de estos elementos.

2.2 Tecnologías para el desarrollo de soluciones de Inteligencia de Negocios

Como se ha mencionado hasta este punto, el concepto inteligencia de negocios constituye un término sombrilla, que agrupa un conjunto de tecnologías, técnicas, conceptos y herramientas en dos componentes: (i) integración de datos y (ii) análisis de datos [9] [10] [11].

Bill Inmon en [12] define bodega de datos como: “...una colección de datos integrados orientados a temas, integrados, no-volátiles y variables en el tiempo, organizados para soportar necesidades empresariales...” y Ralph Kimball en [10] la define como: “...una colección de datos en forma de una base de datos que guarda y ordena información que se extrae directamente de los sistemas operacionales (ventas, producción, finanzas, marketing, etc.) y de datos externos...”.

Por otra parte, el proceso de extracción transformación y carga (*Extraction, Transformation and Load - ETL*) es el encargado de seleccionar los datos, formatearlos y depositarlos en la bodega de datos. Para cumplir con esta función, se subdivide en tres subprocesos. En la Tabla 1, se describen las operaciones y elementos involucrados en cada subproceso [9] [12] [13].

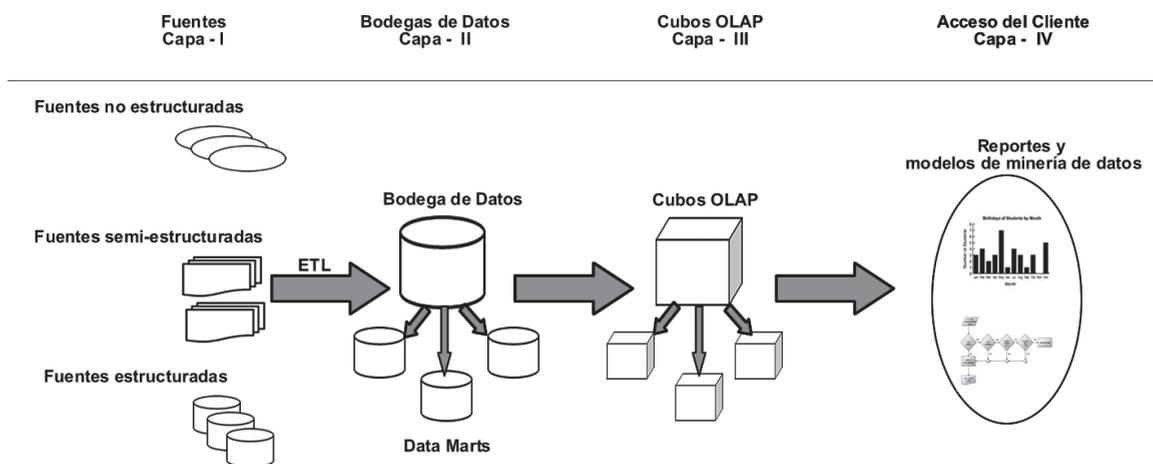


Figura 1. Componentes de la solución de BI desarrollada

Tabla 1. Componentes del Proceso de ETL

Componente	Elementos Objetivos (entrada)	Operaciones realizadas (proceso)	Resultado de la tarea (salida)
Extracción	Fuentes de datos, sistemas transaccionales, hojas de cálculo, archivos de texto.	Selección	Datos crudos(cargados en memoria)
Transformación	Datos crudos(cargados en memoria)	Limpieza, transformación, personalización, realización de cálculos y aplicación de funciones de agregación.	Datos formateados, estructurados y resumidos de acuerdo a las necesidades(aún en memoria)
Carga	Datos formateados, estructurados y resumidos de acuerdo a las necesidades(aún en memoria)	Inserción	Datos formateados, estructurados y resumidos con persistencia en el DW

Respecto del Procesamiento OLAP Codd en [14] lo define como "...un tipo de procesamiento de datos que se caracteriza, entre otras cosas, por permitir el análisis multidimensional de datos...". Este análisis se basa en modelar la información en forma de medidas, hechos y dimensiones. Las medidas son los valores de un dato en particular, las dimensiones son las descripciones de las características que definen dicho dato y los hechos son la definición de una o más medidas para una combinación particular de dimensiones.

El modelo derivado del procesamiento OLAP se representa vectorialmente: los hechos se ubican lógicamente en una celda que queda en la intersección de ciertas coordenadas según el modelo de coordenadas (x, y, z,...), donde cada una de las coordenadas de la celda representa una dimensión. Esto es conocido como análisis multidimensional y para materializarlo en una base de datos, se usa la correspondencia entre los elementos del modelo (hechos y coordenadas) en la bodega de datos (tabla de hechos y tablas de dimensiones).

Por último, Fayyad en [15] define la minería de datos como: "un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos".

2.3. Metodologías para el desarrollo de soluciones de Inteligencia de Negocios

En la revisión de la literatura realizada, se encontró que existe un amplio número de trabajos alrededor de las técnicas y tecnologías que conforman una solución de inteligencia de negocios, sin embargo, no existen consensos sólidos respecto de los modelos de

proceso de desarrollo que pueden utilizar los equipos de ingeniería de software para construir este tipo de soluciones. Para mostrar esta diversidad, en la Tabla 2 se presenta una síntesis de las metodologías para el desarrollo de soluciones de BI en donde se identifica el modelo de ciclo de vida subyacente, el nivel de detalle en la especificación de las actividades a realizar, y el alcance en términos de los componentes de la solución de BI que se construyen con su guía.

Como se ve en la Tabla 2, existen varios modelos que abarcan todo el proceso para el desarrollo de una solución de inteligencia de negocios como [11][23] [25][24], sin embargo, la metodología expuesta en [24] ofrece un nivel de detalle muy bajo, pues es un modelo de ciclo de vida tradicional que no especifica aspectos particulares relativos a los componentes de una solución de BI. Por otra parte, lo planteado en [23] es una metodología demasiado pesada para equipos pequeños y con bajos niveles de madurez, lo cual impide una apropiación rápida por parte del equipo de desarrollo. Por su parte, en [11] se presenta una metodología que, pese a ofrecer una descripción global de muchas de las tareas del proceso de BI, sigue un enfoque de cascada y expone un nivel de detalle muy bajo. Finalmente [25] propone una metodología ágil pero con un nivel de detalle demasiado bajo, pues únicamente define los procesos generales, no define las actividades y las tareas específicas que debe cumplir el equipo de desarrollo, además está enfocada en el uso de las herramientas de desarrollo ofrecidas por la organización Pentaho.

Con base en el análisis anterior, se decidió que la metodología empleada para el desarrollo de este proyecto es la planteada por Larissa T. Moss y Shaku Atre en su libro "Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support

Tabla 2. Metodologías para BI y sus componentes.

Metodologías y Técnicas / característica	Modelo de Ciclo de vida	Nivel de Detalle	Alcance
Ciclo de vida de Ralph Kimball [16]	Cascada	Alto	Bodega de datos
Método de desarrollo de sistemas dinámicos para bodegas de datos (DSDM DW) [17]	Iterativo	Bajo	Bodega de datos
Metodología para la construcción de un DataWarehouse(Hefesto) [18]	Cascada	Bajo	Bodega de datos
Metodología DWEF [19]	iterativo	Alto	Bodega de datos
MBD [20]	Ágil	Bajo	Bodega de datos
SEMMA [21]	Cascada	Medio	Minería de Datos
CRISP-DM [22]	Cascada	Medio	Minería de Datos
Business Intelligence RoadMap [11]	Cascada	bajo	Solución de BI
BIEP [23]	Iterativa	alto	Solución de BI
Metodología y modelo de ciclo de vida de una solución de inteligencia de negocios (Gartner) [24]	Iterativo	Bajo	Solución de BI
Pentaho (BI) [25]	Ágil	medio	Solución de BI

Applications”[11], en combinación con el método ágil para el desarrollo de software denominado Programación Extrema [26][27][28]. Como se ilustra en la Figura 2, la metodología propuesta en [11] sigue un modelo de desarrollo en cascada, donde las fases de análisis, diseño, implementación y despliegue, son realizadas de forma secuencial, rígida y sin solapamiento. Pero, proporciona una guía clara de las fases de un proyecto de inteligencia de negocios, especificando algunas actividades con sus respectivos entregables. Adicionalmente, tomando las prácticas de XP se logró dotar a la metodología del nivel de la agilidad demandada por el proyecto.

3. DESARROLLO DE LA SOLUCIÓN

3.1. Análisis

El primer paso de la fase de análisis consiste en la definición de requisitos del proyecto. Para esto, el equipo de trabajo desarrolló entrevistas a los encargados de cada uno de los componentes del PCA, así como, al coordinador general, en donde se identificaron y priorizaron las necesidades a suplir con la solución de BI. En la Tabla 4 se sintetizan los principales requisitos del proyecto.

Tabla 3. Descripción de las fases de la Metodología.

Fase	Objetivo	Resultados	Actividades
Análisis	Describir el negocio y la arquitectura que este posee,	Especificación de: • requisitos. • datos disponibles.	<ul style="list-style-type: none"> Definición de los requisitos del proyecto. Análisis de los datos. Aplicación de prototipos. Análisis de metadatos.
Diseño	Hacer los planos de la solución deseada, de acuerdo a los recursos disponibles: datos, tiempo y personal.	Modelos de: • Bodega de datos. • Metadatos. • Proceso de ETL • Cubos OLAP	<ul style="list-style-type: none"> Diseño de la base de datos. Diseño del repositorio de metadatos. Diseño del proceso de ETL.
Construcción	Realizar lo establecido en los planos.	<ul style="list-style-type: none"> Base de datos. metadatos. Proceso de ETL. Cubos OLAP. Reportes. 	<ul style="list-style-type: none"> Desarrollo del ETL Desarrollo del sistema de metadatos. Desarrollo de la Aplicación.
Despliegue	Poner en producción lo desarrollado.	<ul style="list-style-type: none"> Plan de implementación. Configuración desplegada 	<ul style="list-style-type: none"> Integración y despliegue

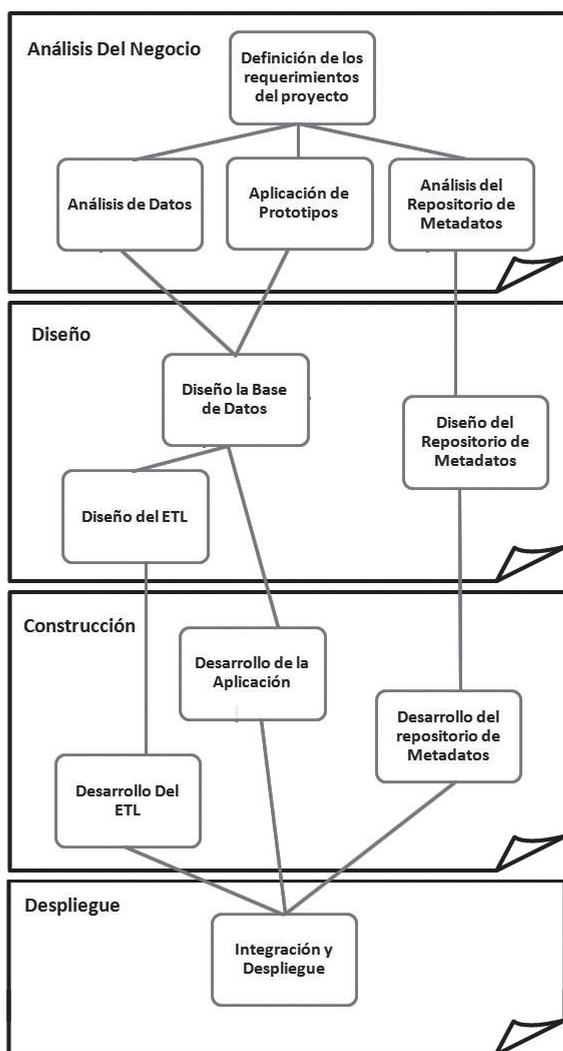


Figura 2. Fases y Pasos de la Metodología propuesta en [11].

Tabla 4. Principales requisitos del proyecto

Requisito	Descripción
Soportar la toma de decisiones a través del suministro de información oportuna y contextualizada.	El propósito principal de la solución era explotar la cantidad de datos disponible y que era de cierta forma desaprovechada.
Integrar la solución a la Intranet del proyecto.	La nueva solución debía ser accedida desde la intranet ya existente.
Hacer liberaciones frecuentes de la solución.	Poner al acceso de los usuarios las soluciones una vez hayan sido terminadas y pasado las pruebas.

Alexander Armando Bustamante Martínez,
Ernesto Amaru Galvis Lista, Mayda Patricia González Zabala,
Arys Alberto García Avendaño

Utilizar como unidad mínima de análisis a los estudiantes y núcleos familiares según corresponda.

Contar con valores recientes (menor a 15 días) para el análisis.

Permitir que los diferentes usuarios puedan llegar al nivel de detalle más alto en sus consultas, a saber, del niño y del núcleo familiar.

Las actualizaciones en los datos eran diarias y la coordinación demandó tener máximo un retardo de 15 días en la información analizada.

En la fase de análisis de datos, la metodología plantea actividades orientadas a manejar y garantizar la calidad de datos, es decir, detectar y corregir errores, tratar valores faltantes, darles el formato adecuado para su posterior uso en la bodega de datos, y definir los metadatos necesarios para facilitar la consulta y utilización de la bodega de datos. En este sentido, es necesario decir que la Intranet “Círculos de Aprendizaje” utiliza una única base de datos que constituye la fuente de datos de la solución de BI. En ésta base de datos se almacenan datos referentes a los NNA, sus núcleos familiares, las instituciones educativas receptoras, los círculos encargados de la nivelación académica, el proceso de pre-matricula y matricula, su progreso académico y psicosocial y la asistencia a los círculos. Además se incluyen datos relacionados con los tutores que atienden la operación de los círculos. El análisis realizado abarcó 64 de las 139 tablas de la base de datos de la Intranet, pues se priorizó el análisis sobre los datos de los componentes comunitario y pedagógico. Durante la evaluación de calidad de los datos, se encontraron problemas causados por inconsistencias lógicas, valores nulos y duplicados. En la Tabla 5 se describen las inconsistencias detectadas en la fase de análisis.

Tabla 5. Inconsistencias detectadas en la fase de análisis

Tipo de Inconsistencia	Descripción	Ejemplo
Lógica	Serie de valores que debían seguir una secuencia definida,	Cambios de estado de un niño durante su estadía en un círculo.
Valores Nulos	Valores que eran obligatorios en el formulario y no estaban almacenados	Fecha de nacimiento de los niños.

Valores Duplicados	Registros que aparecían más de una vez innecesariamente	Niños registrados hasta tres veces en el aplicativo.
--------------------	---	--

Por otra parte, respecto del análisis del repositorio de metadatos, se halló pertinente definir únicamente metadatos de consulta que permiten estructurar y analizar los datos contenidos en la bodega de datos, de forma que la herramienta de análisis pueda operar sobre ellos al momento de realizar las consultas. Por tal razón, se analizó una estructura que permitiera definir características de los cubos OLAP como las tablas de dimensiones, las tablas de hechos y los atributos de ambas; así como las operaciones permitidas para su navegación como la obtención de detalles o *Drill Down*, y el resumen o *Roll up*. De forma general, estos metadatos son manejados por las herramientas que permiten crear y gestionar los objetos en la bodega. Sin embargo, se incluyó esta definición para posibilitar la creación de nuevas aplicaciones independiente del esquema utilizado para definir la estructura de los datos. De igual manera, se aplicaron prototipos en la fase de análisis que permitieron identificar requerimientos relacionados con la usabilidad de la solución, el tipo de información que necesitaban acceder, su formato y su sensibilidad. Por ejemplo, se logró identificar que el tipo de gráfico deseado para visualizar el rendimiento

académico de los niños, era un gráfico de barras, con una granularidad de estudiante.

3.2 Diseño

Una vez definidos los recursos de datos disponibles y su estado, se prosiguió a diseñar las estructuras y secuencias necesarias para volverlos útiles para el personal del PCA. El primer paso fue diseñar la bodega de datos utilizando un esquema en estrella, como se muestra en la Figura 3. Esta decisión de diseño se tomó dado que por demanda del cliente, la frecuencia de actualización de los datos debía ser inferior a quince días. Esta restricción implica minimizar el tiempo empleado para la carga de datos, y para la transformación de la estructura relacional a la multidimensional, es decir, cubos OLAP en modo de almacenamiento MOLAP. El diseño de la bodega de datos está compuesto por cincuenta y cinco (55) tablas¹ de las cuales doce (12) son “Tablas de Hechos” y cuarenta y tres (43) son “Tablas de Dimensiones”.

Por otra parte, para el diseño de la estructura de metadatos se utilizó el diagrama de clases del lenguaje unificado de modelado –UML. En este modelo se representa cada uno de los atributos, los métodos, y las operaciones para la navegación entre los datos. La decisión de utilizar un diagrama de clases de UML para especificar el diseño del repositorio de metadatos pues es sencillo de comprender y facilita la comunicación con los desarrolladores. Esto se presenta en la Figura 4.

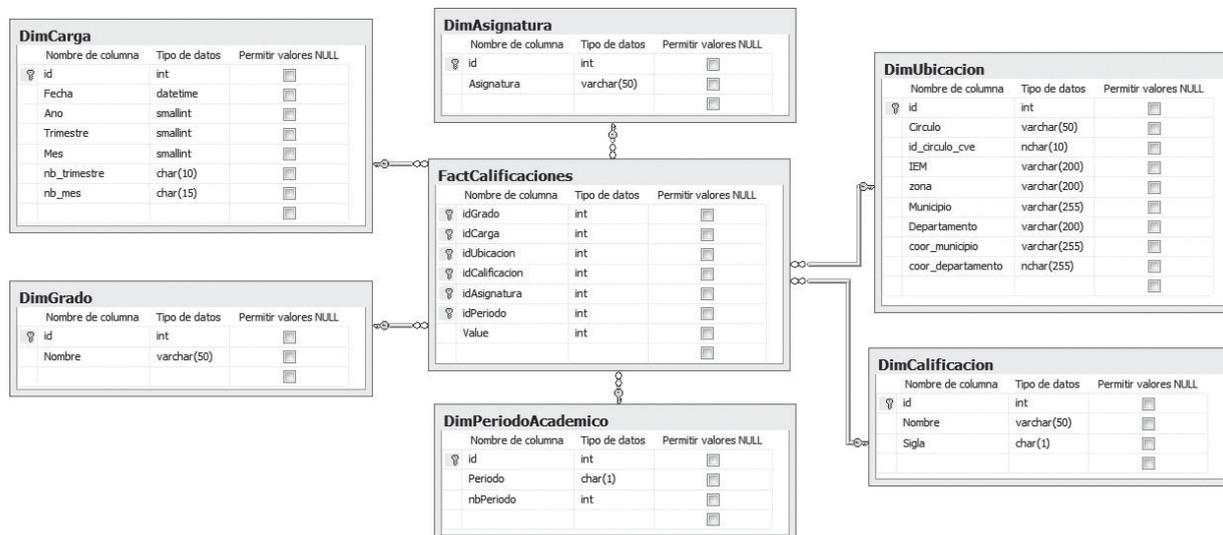


Figura 3. Diseño de la bodega de datos.

¹ Para mayor información consultar [29]

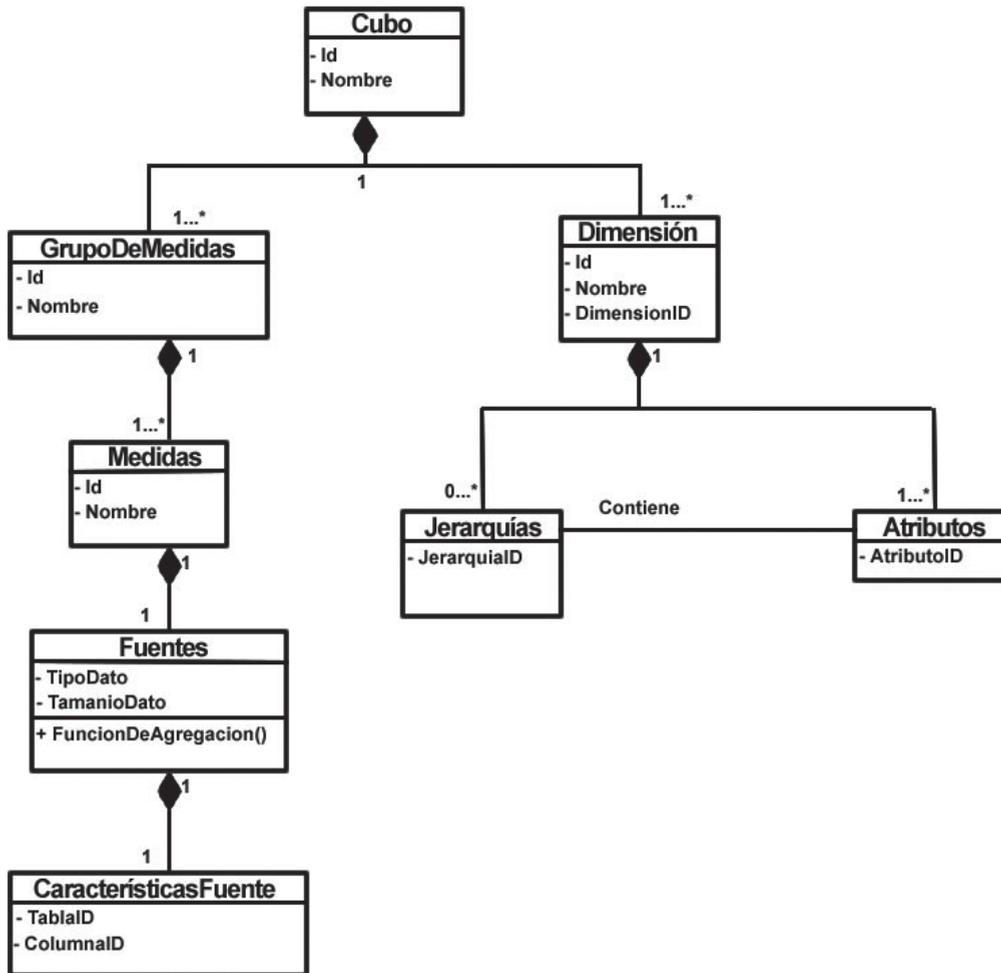


Figura 4. Diseño del Repositorio de Metadatos.

Una vez diseñada la estructura contenedora de los datos, la bodega de datos, y cómo navegarlo, los metadatos. Se debe diseñar el proceso para seleccionar y dejar listos los datos a ser utilizados, es decir, el proceso de ETL. Al momento de diseñar el modelo lógico-conceptual para el proceso de ETL, el equipo de trabajo optó por utilizar el esquema planteado en [24], que usa algunos estereotipos de UML para el diseño de los procesos ETL. Esto expone ventajas pues al ser UML un estándar ampliamente aceptado para el análisis y diseño orientado a objetos, al adaptarlo al modelado de este proceso, la curva de aprendizaje es mucho menor que si se incorpora un nuevo lenguaje. En la Figura 5 se puede apreciar el uso de esta técnica en el diseño del proceso de ETL para la dimensión “Asignatura”, la cual hace referencia a las áreas de formación incluidas en los planes de estudio de lo NNA atendidos en el PCA.

La Tabla 6, representa la documentación para el proceso de carga diseñado para la dimensión Asignatura.

3.3 Implementación

El primer resultado de la fase de implementación fue la bodega de datos. Para esto, a partir del modelo relacional obtenido en el diseño, se generó un script con instrucciones de definición de datos en el lenguaje estructurado de consultas *SQL*. Este script fue ejecutado en el motor de base de datos seleccionado, Microsoft SQL Server 2008, para completar la implementación. La Figura 6, presenta una parte del diagrama de tablas de la bodega de datos en SQL Server 2008.

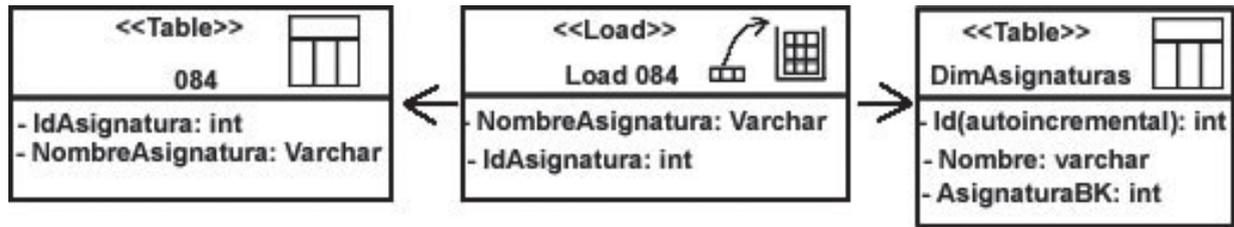


Figura 5. Diseño del Proceso de ETL Dimensión Asignatura.

Tabla 6. Descripción del Proceso de ETL Dimensión Asignatura.

Nombre	Documentación
084	Tabla que contiene el listado de las asignaturas o materias que los estudiantes ven durante un periodo académico y en las cuales deben cumplir logros.
load_084	Variable tabla contiene los datos que van a ser cargados en la dimensión Asignatura
DimAsignaturas	Tabla en el DW que contiene la información referente a Las Asignaturas de los círculos

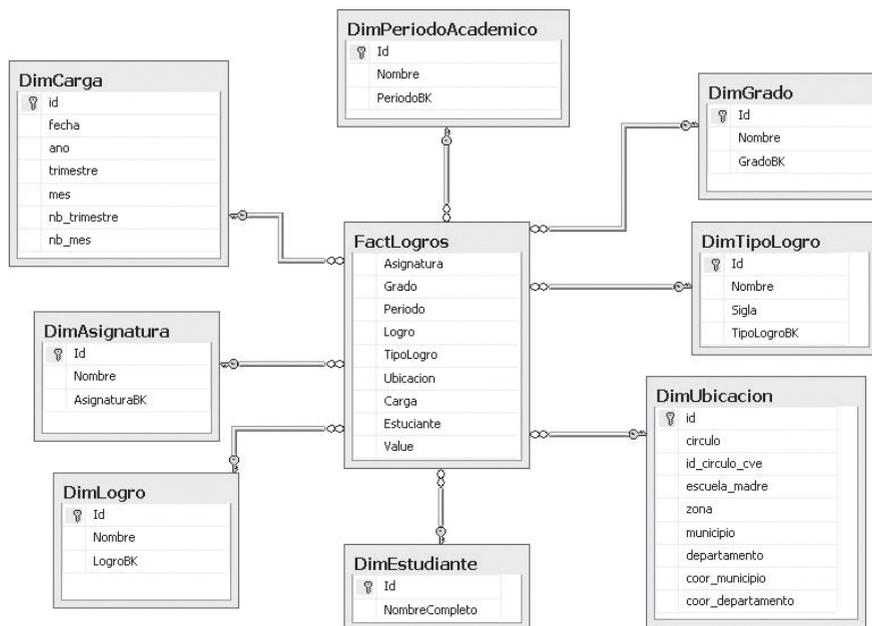


Figura 6. Segmento de la Bodega de Datos

Posteriormente, para dotar a la solución de una estructura que permitiera navegar a través de los datos almacenados, se implementó el repositorio de metadatos. Para esto, se utilizó el lenguaje de marcado extendido XML en el cual se definieron las etiquetas de acuerdo al bosquejo del diagrama de clases. Con esto, la herramienta de análisis de datos seleccionada, Microsoft Analysis Services, puede reconocer la estructura de los cubos y habilitar las operaciones de

consulta especificadas. El esquema XML se presenta en la Figura 7 .

El siguiente modelo a implementar, fue el del proceso ETL. Para esto se utilizó la herramienta Microsoft Integration Services por ofrecer distintas facilidades y características útiles en el desarrollo del Proyecto, entre las que se pueden mencionar: dispone de componentes pre construidos, los cuales permiten realizar tareas

evitando su desarrollo; y su perfecta integración con la Suite de Microsoft SQL Server, la cual fue escogida para la implementación de la bodega de datos.

La estructura base del proceso de carga, se puede apreciar en la Figura 8, un paquete padre, que se encarga de manejar el proceso de carga. Su función es delegar el proceso de ciertas operaciones a paquetes hijos y recibir el resultado de éstas. La descripción de este paquete se puede apreciar en la Tabla 7.

Adicionalmente, en la Figura 8, se puede apreciar la precondición que existe para la ejecución de los tres paquetes de hechos, ubicados en la parte inferior de la figura, que consiste en la ejecución del paquete que carga todas las dimensiones. Debido a que los componentes incluidos en la bodega de datos en la mayoría de los casos no comparten sus dimensiones, se facilitó dividir la carga en paquetes que pueden ejecutarse en paralelo. Por otro parte, los datos compartidos fueron cargados en un único flujo de datos denominado, “datos Comunes”, lo cual se muestra en la Figura 9.

```
<?xml version="1.0" encoding="utf-8"?>
<Cubo>
  <ID></ID>
  <Nombre></Nombre>
  <Dimensiones>
    <Dimension>
      <ID></ID>
      <Nombre></Nombre>
      <DimensionID></DimensionID>
      <Atributos>
        <Atributo>
          <AtributoID>
          </AtributoID>
        </Atributo>
        <Jerarquias>
          <Jerarquia>
            <JerarquiaID></JerarquiaID>
          </Jerarquia>
        </Jerarquias>
      </Atributos>
    </Dimension>
  </Dimensiones>

  <GrupoDeMedidas>
    <GrupoDeMedida>
      <ID></ID>
      <Nombre></Nombre>
      <Medidas>
        <Medida>
          <ID></ID>
          <Nombre></Nombre>
          <Fuente>
            <FuncionAgregacion></FuncionAgregacion>
            <TipoDato></TipoDato>
            <TamañoDato></TamañoDato>
            <CaracteristicasFuente>
              <TablaID></TablaID>
              <ColumnaID></ColumnaID>
            </CaracteristicasFuente>
          </Fuente>
        </Medida>
      </Medidas>
    </GrupoDeMedida>
  </GrupoDeMedidas>
</Cubo>
```

Figura 7. Implementación del Repositorio de Metadatos.

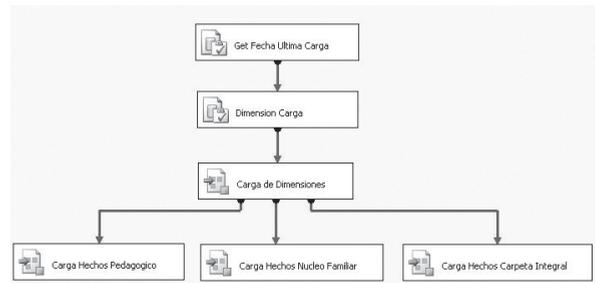


Figura 8 Implementación del paquete principal de ETL

Tabla 7. Descripción del Paquete Principal de ETL

Nombre Paquete	Descripción
MasterPackage.dtsx	Su función es coordinar la ejecución de los demás paquetes.
CargaDimensiones.dtsx	Su función es realizar la carga de todas las dimensiones utilizadas en el almacén de datos.
CargaHechosPedagogico.dtsx	Su función es realizar la carga de las tablas de hechos que nos permiten analizar en el componente pedagógico.
CargaHechosNucleo.dtsx	Su función es realizar la carga de las tablas de hechos que nos permiten analizar las características básicas de los núcleos familiares.
CargaHechosIntegral.dtsx	Su función es realizar la carga de las tablas de hechos que nos permiten analizar a los niños en factores como su asistencia.

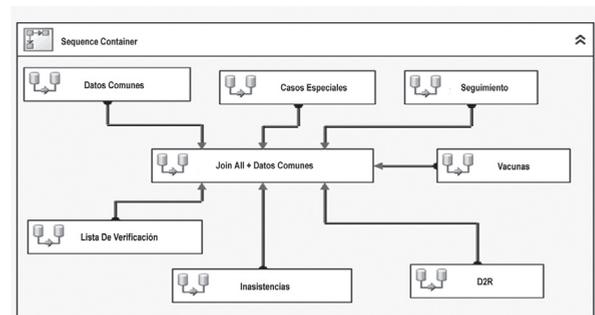


Figura 9. Paquete de ETL para la carga de datos comunes

En este punto se completa la implementación de la bodega de datos y las rutinas que posibilitan llenarla. El paso siguiente fue construir estructuras multidimensionales

para definir perspectivas y agregaciones sobre los datos, basados en la forma de almacenamiento escogida. Por tanto, en la construcción de los cubos OLAP se crearon vistas sobre los datos y se crearon columnas que solo son procesadas en la implementación del cubo.

Al finalizar la creación del cubo, se procedió a configurar las estructuras de dimensiones, seleccionando los atributos y estableciendo sus propiedades de implementación. Del mismo modo, se especificaron las jerarquías dentro de las dimensiones. Esto se ilustra en la Figura 10.

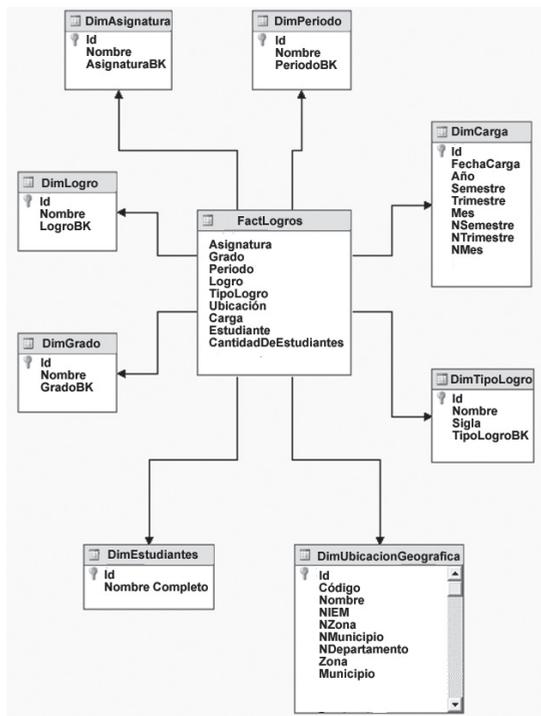


Figura 10. Estructura de un cubo implementado en SQL Analysis Services.

Además, se definieron las agregaciones o resúmenes de datos pre-calculados preparados para dar respuestas a las consultas antes de que sean formuladas. En este punto es importante destacar que entre mayor número de agregaciones localice la herramienta de análisis, mayor será el rendimiento en el tiempo de respuesta para el usuario.

Una vez construidos todos los cubos, se procedió a su explotación a través de reportes. Para esto, se utilizó la herramienta Microsoft SQL Server Reporting Services. Esta herramienta hace uso del lenguaje para la programación de reportes denominado Lenguaje de Expresiones Multidimensionales o MDX. Este lenguaje permite crear de forma simple, consultas complejas sobre los cubos, obteniendo tiempos de respuesta considerablemente bajos. Adicionalmente, el hecho de haberse convertido en un lenguaje estándar, utilizado por una gran cantidad de los proveedores de herramientas para desarrollar sistemas OLAP, contribuye a mantener cierto grado de independencia con respecto a la plataforma.

3.4 Integración y despliegue

Para terminar, una vez diseñados, implementados y probados los reportes, la tarea siguiente fue disponerlos en ambiente de producción. Para el despliegue de los reportes e indicadores, se integró con la Intranet Círculos de Aprendizaje, lo que posibilitó el acceso a los datos contenidos en los cubos OLAP desde un mismo sitio, la Intranet del proyecto. En la Figura 11, se puede apreciar uno de los reportes, un gráfico de barras, el cual muestra el número de niños, por calificación obtenida en los círculos de la institución educativa C.E.B de la ciudad de Barranquilla. Reportes de este tipo, algunos más dinámicos, y otro en Excel, permitieron a la coordinación del proyecto obtener los datos que necesitaban para soportar sus decisiones en un menor tiempo. Además el hecho de que los usuarios ya estaban entrenados para usar la intranet del PCA, facilitó su aceptación.

4. CONCLUSIONES

La implementación de la solución de BI permitió apreciar como el uso de la TI, y en especial la BI pueden apoyar los procesos de gestión y toma de decisiones en contextos como el ámbito educativo y contribuyen a fortalecer las iniciativas orientadas a mejorar las condiciones de vida de la población infantil en estado de vulnerabilidad. En el proyecto PCA se evidenció, mediante el uso de la solución desarrollada

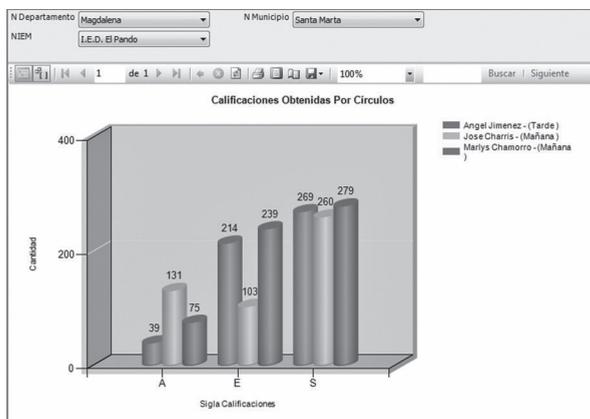


Figura 11. Ejemplo de reporte construido utilizando SQL Server Reporting Services

en decisiones como la asignación de tutores y compra de útiles

En relación con la metodología base empleada, se encontró que permitió la identificación, de procesos, actividades y tareas propias de una solución de BI, pero no suplió carencias en cuanto a las tareas y técnicas de construcción. Así mismo, no se pudo adaptar completamente al enfoque ágil que demandó el proyecto. No obstante, este sentido, la incorporación de los prácticas ágiles para el desarrollo de software expuestas en la metodología eXtreme Programming (XP) permitieron superar algunas falencias detectadas.

De igual manera, pese a que la inexperiencia del equipo, contribuyó a tomar la decisión de combinar las metodologías, es un hecho la necesidad de una metodología que propicie a equipos de desarrollo inmaduros, emergente de instituciones de educación superior, entrar en el campo de la inteligencia de Negocios.

Por otra parte, el que no exista un modelo estándar para el diseño de los procesos ETL lleva a que se presenten dificultades en esta etapa del desarrollo. Dificultades como: desorientación por parte de equipos inexpertos, aumenta el riesgo de una mala elección de la técnica para modelar, retrasa la construcción de herramientas que permitan el modelado de estos procesos y la integración con los IDEs existentes en el mercado.

Finalmente, la interacción con el equipo del PCA y la disposición de recursos fueron factores fundamentales en lograr el alcance de los objetivos propuestos. Por otra parte, el desarrollo del proyecto aportó en el fortalecimiento del trabajo que viene adelantando la Vicerrectoría de Extensión y el Programa de Ingeniería de Sistemas de la Universidad del Magdalena.

5. REFERENCIAS

- [1] Ministerio de Educación Nacional, «CVN - MÁS EDUCACION PARA LA POBLACIÓN DESPLAZADA». [Online]. Available: <http://www.mineduacion.gov.co/cvn/1665/article-100922.html>. [Accessed: 15-feb-2012].
- [2] Ministerio de Educación Nacional, «Lineamientos de política para la atención educativa a la población afectada por la violencia». .
- [3] Ministerio de Educación Nacional, «PLAN NACIONAL DE DESARROLLO EDUCATIVO INFORME DE GESTIÓN». jun-2008.
- [4] «Población Vulnerable - Fundación Escuela Nueva». [Online]. Available: <http://www.escuelanueva.org/pagina/index.php?codmenu=554&idioma=1>. [Accessed: 15-feb-2012].
- [5] Universidad del Magdalena, «Objetivo - Ciclos de Aprendizaje». [Online]. Available: <http://ciclos.unimagdalena.edu.co/>. [Accessed: 15-feb-2012].
- [6] Cardona, Sonia, «La Inteligencia de negocios y su aplicación en algunas empresas del área metropolitana de Medellín». Universidad EAFIT, 2005.
- [7] «The Gartner Fellows: Howard Dresner's Biography». [Online]. Available: http://www.gartner.com/research/fellows/asset_79427_1175.jsp. [Accessed: 15-sep-2011].
- [8] W. W. Eckerson, Performance Dashboards: Measuring, Monitoring, and Managing Your Business, 2.a ed. Wiley, 2010.
- [9] R. Kimball, M. Ross, B. Becker, J. Mundy, y W. Thornthwaite, Kimball's Data Warehouse Toolkit Classics: The Data Warehouse Toolkit, 2nd Edition; The Data Warehouse Lifecycle, 2nd Edition; The Data Warehouse ETL Toolk, 2.a ed. Wiley, 2009.
- [10] R. Kimball, The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses. John Wiley & Sons, 1996.
- [11] L. T. Moss y S. Atre, Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. Addison-Wesley Professional, 2003.
- [12] W. H. Inmon, Building the Data Warehouse, 4.a ed. Wiley, 2005.
- [13] R. Kimball y J. Caserta, The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleanin, 1.a ed. Wiley, 2004.
- [14] Edgar Codd, «Providing OLAP to user-analysts: An IT mandate». E F Codd and Associates, 1993.
- [15] U. M. Fayyad, G. Piatesky-Shapiro, P. Smyth, y R. Uthurusamy, Advances in Knowledge Discovery and Data Mining. AAAI Press, 1996.
- [16] R. Kimball, M. Ross, W. Thornthwaite, J. Mundy, y B. Becker, The Data Warehouse Lifecycle Toolkit, 2.a ed. Wiley, 2008.
- [17] «White Paper DSDM and Data Warehousing – DSDM Atern». [Online]. Available: <http://www.dsdm.org/dsdm-atern/dsdm-v4-2-white-papers/white-paper-dsdm-and-data-warehousing>. [Accessed: 14-sep-2011].

- [18] «Data Warehousing y metodología Hefesto | Dataprix». [Online]. Available: <http://www.dataprix.com/es/data-warehousing-hefesto>. [Accessed: 14-sep-2011].
- [19] S. Luján-Mora y J. Trujillo, «A Data Warehouse Engineering Process», in *Advances in Information Systems*, vol. 3261, T. Yakhno, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 14–23.
- [20] «MBD1.0-METODOLOGÍA DE DESARROLLO DE BODEGAS DE DATOS PARA MICRO, PEQUEÑAS Y MEDIANAS EMPRESAS», 27-oct-2010. [Online]. Available: <http://revistas.uis.edu.co/index.php/revistausingenierias/article/view/1059>. [Accessed: 14-sep-2011].
- [21] «SAS | SEMMA». [Online]. Available: <http://www.sas.com/offices/europe/uk/technologies/analytics/datamining/miner/semma.html>. [Accessed: 14-sep-2011].
- [22] «Metodología CRISP-DM para minería de datos | Dataprix». [Online]. Available: http://www.dataprix.com/modelo_crisp-dm. [Accessed: 14-sep-2011].
- [23] E. Herrera, «“Metodología para el desarrollo de un sistema de inteligencia de negocios basada en el proceso unificado”», Universidad Nacional De Colombia, Bogotá, 2011.
- [24] Martínez Jaime, «La inteligencia de negocios como herramienta para la toma de decisiones estratégicas en las empresas. Análisis de su aplicabilidad en el contexto corporativo colombiano.», Universidad Nacional De Colombia, Bogotá, 2010.
- [25] «Agile BI -- TDWI -The Data Warehousing Institute». [Online]. Available: <http://tdwi.org/portals/agile-bi.aspx>. [Accessed: 14-sep-2011].
- [26] K. Beck y C. Andres, *Extreme Programming Explained: Embrace Change*, 2.a ed. Addison-Wesley Professional, 2004.
- [27] N. Roodyn, *eXtreme .NET: Introducing eXtreme Programming Techniques to .NET Developers*. Addison-Wesley Professional, 2004.
- [28] R. Jeffries, *Extreme Programming Adventures in C#*, 1.a ed. Microsoft Press, 2009.
- [29] L. Benavides, A. Bustamante, y A. García, «Desarrollo de una solución de Inteligencia De Negocios para apoyar a la toma de decisiones en el Proyecto Círculos De Aprendizaje», Universidad del Magdalena, Magdalena, Colombia, 2009.

6. AGRADECIMIENTOS

Los autores expresan su agradecimiento a la Vicerrectoría de Extensión de la Universidad del Magdalena por el apoyo y la confianza depositada para desarrollar con éxito este proyecto.