

GENERACIÓN ASISTIDA POR COMPUTADORAS DE MODELOS DE ALMACENES DE DATOS: ANÁLISIS DE LA INFORMACIÓN

COMPUTER-ASSISTED GENERATION OF DATA WAREHOUSE MODEL: ANALYSIS OF INFORMATION

LINDSAY ALONSO GÓMEZ-BELTRÁN

MSC. Jefe Departamento de Computación. Universidad de Camagüey, Cuba. e-mail: lindsay.gomez@reduc.edu.cu

ROSENDO MORENO-RODRÍGUEZ

Ph.D, Universidad Central "Marta Abreu" de Las Villas, Cuba. e-mail: rosendo@uclv.edu.cu

RAMIRO PÉREZ-VÁZQUEZ

Ph.D, Universidad Central "Marta Abreu" de Las Villas, Cuba. e-mail: rperez@uclv.edu.cu

Recibido para revisar Febrero 25 de 2012, aceptado Diciembre 21 de 2012, versión final Enero 8 de 2013

RESUMEN: En este trabajo se presentan las pautas metodológicas para la generación asistida por computadoras de modelos de almacenes de datos (AD), estas pautas metodológicas se dividen en 4 etapas, las 3 etapas primeras son: análisis de la información, modelo conceptual y diseño lógico y una que ocurre dentro de cada una de las etapas anteriores que es denominada etapa de trazabilidad. Estas etapas describen cómo podemos obtener a partir de los sistemas operacionales heredados (E/R) una propuesta de modelado de almacenes de datos. Debido a que el análisis de la información juega un papel fundamental en las etapas tempranas, en este trabajo se profundiza en la primera etapa de estas pautas metodológicas. En la primera etapa se trabaja en dos fases una para la selección de los requisitos de usuario y otra para el trabajo con los esquemas OLTP.

PALABRAS CLAVE: Diseño de almacenes de datos; Modelo de almacenes de datos; Requisito de usuario.

ABSTRACT: In this paper we present the methodological guidelines for computer-assisted generation of data warehouse models (AD), these methodological guidelines are divided into 4 stages, the first 3 stages are: data analysis, conceptual and logical design model and one last occurring within each of the previous stage is called stage of traceability. These stages describe how we obtain from legacy operating systems (E / R) a proposal for modeling data warehouses. Because the analysis of information plays a key role in the early stages, this paper explores the first stage of these methodological guidelines. In the first stage one works in two stages for the selection of user requirements and the other for the work with OLTP schemes.

KEYWORDS: Data warehouse design, Data warehouse model, user requirement.

1. INTRODUCCIÓN

Una de las temáticas que más desarrollo demanda en los momentos actuales en cuanto a las tecnologías software es sin dudas la Ingeniería del Software, y dentro de esta, el desarrollo de métodos y algoritmos apropiados que conlleven a la creación y explotación de herramientas que asistan por medios computacionales al desarrollo de otros sistemas (herramientas CASE).

Dentro del campo de los Sistemas de Información, basados en la explotación de Sistemas de Bases de Datos, en los últimos tiempos se ha desarrollado la

creación y explotación de Almacenes de Datos como uno de los típicos Sistemas de Ayuda a la Toma de Decisiones. Estos sistemas que por concepto, manipulan información histórica recopilada en sistemas de información tradicionales de gestión empresarial, tienen el objetivo de descubrir nuevas informaciones que permitan avanzar en productividad y logros de todo tipo; se caracterizan entre otras cosas -según varios autores entre los que se pueden mencionar a Inmon [1] y Kimball [2]- por tener una estructura de tablas bastante diferente a la heredada de los sistemas tradicionales. Desde la introducción del modelo de datos multidimensional como formalismo de modelado

para Almacenes de Datos, han aparecido en la literatura sobre el tema, distintas propuestas metodológicas para capturar la estructura del almacén de datos. Las soluciones siguen diferentes aproximaciones al diseño: el análisis de los requisitos de usuario, el análisis del esquema de la base de datos operacional o una combinación de ambas aproximaciones [3] (técnicas compuestas).

El problema básico del diseño de un AD consiste en obtener un conjunto de esquemas multidimensionales que permitan satisfacer los requisitos de análisis de los usuarios y que puedan ser mantenidos por las bases de datos operacionales existentes en la organización. La figura 1, muestra las etapas del diseño de un AD.[3]

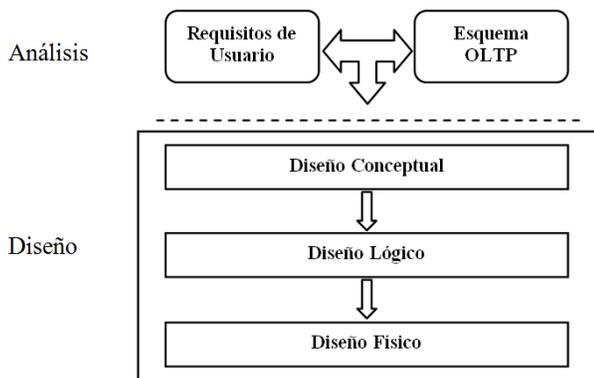


Figura 1. Etapas para el diseño de un AD

En los últimos años se han consagrado muchos esfuerzos por lograr metodologías de modelado multidimensional.

En [3] se propone una metodología del diseño de almacenes de datos a nivel conceptual utilizando un contexto de MDA. En [4] se observa un trabajo profundo sobre MDA para la generación automática de código a partir de modelos UML.

En [5] se presenta una clasificación completa de las metodologías de modelado multidimensional, una de las clasificaciones está dirigida al paradigma del modelado. En donde se proponen cuatro clasificaciones: las metodologías que son manejadas por los requisitos, las metodologías dirigidas por datos, metodologías híbridas, las cuales se pueden clasificar en híbridas intercaladas y en híbridas secuenciales.

El acercamiento dirigido por los requisitos tiene un enfoque para determinar los requisitos multidimensionales de los usuarios, el acercamiento dirigido por datos utilizan un proceso de reingeniería de las fuentes de datos para determinar los conceptos multidimensionales. Las híbridas intercaladas utilizan las metodologías anteriores, de manera que ambas se desarrollan conjuntas y las híbridas secuenciales se utilizan una primero y después la otra de forma secuencial. En [6]. Se plantea una clasificación diferente.

En [7] se describe una herramienta que permite diseñar esquemas de DW a través de transformaciones, permitiendo obtener un modelo lógico de DW y una traza del diseño. Ellos desarrollan el modelo lógico del DW utilizando un grupo de transformaciones de esquemas a las cuales le llaman primitivas, las cuales abstraen y materializan técnicas de diseño, pero ellos no tienen en cuenta el modelo conceptual ni tampoco realizan una propuesta de las tablas de hechos, ni trabajan con base de datos heterogéneas. En [8] el trabajo parte de que el diseñador representa universo del discurso utilizando notación UML y con esto obtiene un esquema UML.

En [9] los autores proponen un levantamiento de requisito utilizando un modelo orientado a meta, utilizan dos perfiles UML y un juego de transformaciones QVT, nosotros coincidimos con los autores en que es importante capturar los requisitos de los usuarios en etapas tempranas.

En [10] se presenta una filosofía de análisis de los requisitos orientada a meta, basada en la metodología Tropos que se orienta en el marco de trabajo de i^* , ellos proponen dos perspectivas diferentes para el análisis de los requisitos: el modelado organizacional y modelado decisional. También proponen que este levantamiento puede emplearse en dos perspectivas cuando es solamente orientado a meta o cuando se emplea una metodología mixta manejada por datos y metas. Consideramos que como utiliza un marco de trabajo basado en i^* se hace un poco difícil de comprender por el usuario final y no propone la documentación del proyecto.

En [11] se desarrolla un diseño lógico del DW aplicando las transformaciones de esquemas, pero la

decisión de qué transformaciones aplicar, y en qué orden hacerlo dependerá de los requerimientos, de las bases de datos fuente y de los criterios de diseño del propio diseñador. Otros trabajos en esta dirección son [12, 13], [14]. También es importante señalar que en [11] no se tienen en cuenta las bases de datos heterogéneas y se necesita un diseño conceptual del DW antes de pasar por el proceso. En [13] se define un armazón metodológica general para el diseño de almacenes de datos. Basado en un modelo de hechos dimensional (DFM), el autor habla de una metodología de forma general, pero no realiza un trabajo profundo en el proceso de semiautomatizar el análisis del esquema de las base de datos ni cómo se trabaja con las bases de datos heterogéneas solo las menciona.

En este artículo nosotros trabajamos el análisis de la información utilizando el paradigma híbrido intercalado, en donde primero utilizamos un análisis de los requisitos de los usuarios para determinar las necesidades del usuario final y luego utilizamos un análisis a través de las fuentes de datos. Se presenta las pautas metodológicas dividiéndola en 4 etapas, las 3 etapas primeras son: análisis de la información, modelo conceptual, diseño lógico, y una cuarta etapa

de trazabilidad. Estas etapas describen cómo podemos obtener a partir de los sistemas operacionales heredados (E/R) una propuesta de modelado de almacenes de datos. Debido a que el análisis de la información juega un papel fundamental en las etapas tempranas, en este trabajo se profundiza en el análisis de la información. Se desarrolla un caso de estudio en donde se detalla cada uno de los elementos fundamentales de esta etapa. Nosotros notamos que en ninguna de las propuestas anteriores se detalla una buena documentación del modelado del AD. En la primera etapa se trabaja en dos fases una para la selección de los requisitos de usuario y otra para el trabajo con los esquemas OLTP. El aporte fundamental de este trabajo se encuentra en lograr una adecuada documentación del proyecto en la primera etapa.

2. EL DESARROLLO DE LAS ETAPAS.

En este apartado se presenta mediante la tabla 1 la relación de cada una de las etapas, cuáles son las entradas de información necesarias y cuáles son sus salidas, así cuales son los actores involucrados en cada una de las etapas.

Tabla 1. Pautas Metodológicas.

Etapas	Entradas	Salida	Involucra
Análisis de requisitos de información	1. Requisitos de Usuario 2. Esquema OLTP	Esquema refinado	Diseñador, gerente de sistema de información y usuario final.
Modelo conceptual.	Esquema refinado	Modelo conceptual	Sistema, diseñador.
Diseño lógico	Modelo conceptual	Esquema lógico del AD	Sistema, diseñador.
Trazabilidad	Todas las etapas.	Traza del modelo.	Sistema

2.1. Análisis de la información. (Análisis de requisitos de información)

En la etapa se utiliza un análisis dirigido por los requisitos del usuario, para con ello determinar las necesidades del usuario final y que tiene que ver mucho con la efectividad de la información que se le brinda al usuario, con la idea de lo que necesita el usuario final se analizan los sistemas OLTP de las bases de datos fuentes, buscando que las necesidades del usuario puedan ser

mantenidas por las bases de datos operacionales existentes en la organización.

Sintetizando, nuestro trabajo consta de 4 etapas y en la primera etapa (análisis de la información) se definen dos fases.

- Fase 1 - Especificación de requisitos de usuario: consiste en identificar las necesidades de análisis de los usuarios.
- Fase 2 – Analizar los modelos de las base de datos operacionales: con el objetivo de buscar la información que mantenga las necesidades del usuario final dando como resultado un esquema refinado.

En la literatura encontramos que esta primera fase cuyo objetivo es obtener los requisitos de información que tienen los usuarios para el apoyo a la toma de

decisiones, i.e. medidas interesantes y el contexto para su análisis. Puede realizarse utilizando diferentes métodos, se señalan dos trabajos que utilizan diferentes métodos y que consideramos son los más importantes. En [3] se adapta un método de elicitación de requisitos basado en metas y en [15] se utiliza una aproximación basada en objetivos (mediante el uso de i^*).

En la primera fase consideramos trabajar con la propuesta utilizada en [3] adoptando un método de *elicitación de requisitos basado en metas*, donde los requisitos de usuario son recogidos por medio de entrevistas. El propósito de las entrevistas es obtener información acerca de las necesidades de análisis de la organización. El aporte fundamental de la primera etapa se encuentra en la reestructuración de la elicitación de requisito, para contribuir a una óptima documentación del proyecto de creación de almacenes de datos.

2.1.1. Fase 1 - Especificación de requisitos de usuario.

En esta sección, se muestra el modelo de requisitos para capturar los requisitos de un AD. Este modelo se fundamenta en la filosofía de las metodologías basadas en metas, la cual ha sido ampliamente aceptada para la captura de requisitos en sistemas tradicionales. Este modelo de requisitos permite capturar los conceptos del modelo multidimensional a partir de las tareas que el usuario pretende realizar al interactuar con el AD.

Modelo de requisitos.

El análisis de los requisitos inicia con una lista de los usuarios y las tareas de análisis que pretende realizar al interactuar con el AD. Las tareas son integradas en metas para formar un Árbol de Refinamiento de Metas (ARM). Se identifican los conceptos relacionados con el dominio del problema.

Teniendo en cuenta el trabajo de [3] y [16] el esquema del diseño propuesto es dividido en tres etapas: **a) definición de la misión**, **b) identificación de las metas de análisis** y **c) especificación de los requisitos**.

Los usuarios del AD definen la misión el cual refleja el objetivo principal que debe ser satisfecho y que la integración entre los usuarios y el sistema deben de dar respuesta a este propósito.

a) Definición de la misión.

La misión, se refiere a una descripción breve sobre el propósito del sistema que se pretende construir, así como de los servicios que éste proporcionará a sus futuros usuarios [16].

Existen dos razones para enunciar la misión del AD. En primer lugar, los usuarios, diseñadores, programadores y todos los involucrados en el desarrollo del sistema tendrán en mente el objetivo final del sistema. En segundo lugar, fijar el objetivo final del sistema evitará diferencias de interpretación sobre el propósito del mismo. El enunciado de la misión, debe estar compuesto por la **descripción** del sistema, el **propósito** y en algunos casos planteamos la necesidad tener presente la **responsabilidad**.

Descripción.- Consiste en una breve descripción del sistema que se pretende desarrollar.

Propósito.- Consiste en la definición del objetivo o el propósito del sistema. Su descripción debe de contener preferentemente solo una oración.

Responsabilidad.- Cada responsabilidad es una funcionalidad general del sistema y se describe a lo sumo por dos o tres oraciones.

Las técnicas que pueden utilizarse durante esta etapa son: entrevistas, reuniones en grupo, cuestionarios o inmersión en el negocio del cliente, entre otras. Para una mejor ilustración ver el caso de estudio de la sección 3.1.2.

b) Identificación de las metas de análisis

El objetivo principal de esta etapa, consiste en identificar las metas y tareas que el usuario llevará a cabo con el AD. En un ambiente de AD estas tareas, se relacionan con los procesos de negocios a ser analizados.

Siguiendo los lineamientos de las metodologías orientadas a metas, se parte de un conjunto de tareas específicas desempeñadas por los usuarios final del AD y posteriormente realizamos la integración de las tareas hasta llegar a tareas más generales (metas). En [15] se

propone realizar un árbol de refinamiento de funciones, este se usa para una descomposición jerárquica de funciones comerciales.

Para lograr identificar las metas se proponen dos pasos: 1) definición de los procesos de negocio. 2) identificación de las metas de análisis.

Paso 1. Definición de los procesos de negocio.

El primer paso, tiene por objetivo definir los procesos de negocio que van a ser analizados [2].

El análisis de los procesos de negocio, proporcionará al usuario final el conocimiento necesario para la toma de decisiones estratégicas que permitan mejorar dicho proceso de negocio. Para identificar estos procesos de negocio, se pueden utilizar técnicas tradicionales de obtención de requisitos como son los grupos de trabajos o las entrevistas.

Paso 2. Identificar las metas de análisis.

El objetivo del paso 2 consiste en describir el funcionamiento del AD en relación con las tareas de análisis que los usuarios finales pretenden realizar al interactuar con él AD. En [3] proponen la construcción de un Árbol de Refinamiento de Metas (ARM). Para lo cual, utilizan los conceptos de meta y tarea empleados por las metodologías orientadas a metas [17].

La construcción del ARM puede ser realizada empleando técnicas de refinamiento o de abstracción [18] [19]. Cada una de estas técnicas están relacionadas con dos enfoques distintos, el *top-down* y el *bottom-up*. A continuación se explica brevemente cada una de ellas:

Refinamiento de metas: utiliza un enfoque *top-down* para la identificación de las metas. Se parte de las metas del más alto nivel de la organización hasta llegar a las tareas individuales realizadas por los actores del proceso de negocio.

Abstracción de metas: utiliza un enfoque *bottom-up* para identificar las metas. Se inicia con las tareas de bajo nivel de los actores organizacionales y se procede a encontrar las metas del negocio con las que éstas se corresponden.

Para identificar las metas, es necesario precisar primero quiénes son los usuarios responsables de llevar a cabo las tareas.

Sin importar la estrategia seguida para la construcción del ARM, la estructura de éste se compone, por un nodo raíz que representa la meta de más alto nivel, nodos intermedios que representan metas y por último las tareas.

En la figura 2, se muestra un ARM y la notación gráfica que se emplea para la construcción del ARM.

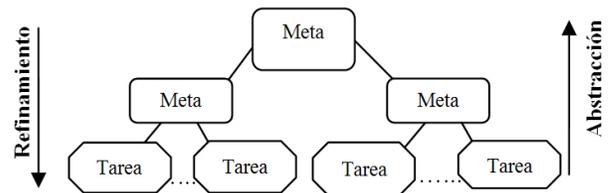


Figura 2. Notación gráfica del ARM

Siguiendo las pautas señaladas en [16] [20], el ARM se convierte en el punto de partida para la construcción de los casos de uso y los diagramas de actividad ya que por cada meta se construirá un caso de uso y un diagrama de actividad.

En todos los trabajos [3, 16, 20, 18] se habla que Sin importar la estrategia seguida para la construcción del ARM, la estructura de éste se compone, por un nodo raíz que representa la meta de más alto nivel, nodos intermedios que representan metas y por último las tareas. Proponemos que en un ARM se puedan analizar más de un proceso de negocio y de esta forma se pueda obtener un modelado mucho más completo dando lugar posteriormente al diagrama de constelación de hechos.

c) Especificación de los requisitos

Esta actividad, consiste en identificar los requisitos de usuario final a partir de las metas identificadas en el ARM. Esta etapa, se realiza en tres pasos: 1) *descripción de las metas en casos de uso*, 2) *descripción de la interacción del usuario con el sistema en diagramas de actividad* y 3) *descripción de los requisitos de información*.

Paso 1. Descripción de las metas en casos de uso: Los casos de uso, permiten representar la funcionalidad del sistema y de aquí que hayan sido adoptados por

numerosos métodos. Un caso de uso, representa la interacción entre el sistema y los actores que harán uso de la funcionalidad del sistema. Esta interacción, es usualmente dividida en frases de lenguaje natural (descripción del caso de uso).

Paso 2. Descripción de la interacción del usuario con el sistema en diagramas de actividad.

Con la finalidad de describir la interacción existente entre los usuarios y el AD, utilizamos diagramas de actividad.

Esta descripción, parte de las acciones que el usuario, necesita realizar al interactuar con el AD para lograr con éxito cada una de las metas. Durante esta actividad, se realiza un diagrama de actividad por cada meta identificada en el ARM.

El diagrama obtenido mostrará los datos necesarios y producidos por cada actividad. Similar a [21] [22] los datos son representados como Elementos de Información (EI) que fluyen entre el usuario y el sistema. Durante la interacción se identifican dos tipos de EI:

El de salida. - El sistema, proporciona información a los usuarios. Estos elementos, se usan para dar respuesta a una consulta.

El de entrada. - El sistema, espera información por parte del usuario. Esta información, es utilizada por el sistema para realizar correctamente una actividad.

Paso 3. Definición de los requisitos de información.

Para definir los requisitos de información, en [3] se propone analizar cada diagrama de actividad y documentar los conceptos (dimensiones, niveles y medidas) que guiarán la selección y el refinamiento de un esquema multidimensional candidato.

Se debe de resumir los requisitos de información relacionados con cada una de las metas identificadas y los dos tipos de EI (*Input*, *Output*). Esta información, debe de ser relacionada con los conceptos del modelo multidimensional de la siguiente forma:

1. Cada EI *Input*, debe ser relacionada con el concepto de dimensión o de jerarquía.
2. Cada EI de tipo *Output*, debe ser relacionada con el concepto de medida.

2.1.2. Caso de Estudio.

Se trabaja con un problema Hospitalario en donde se enfrenta la construcción de un AD para el Hospital Clínico “San Cecilio” de Granada, España; las actividades del Hospital que se desean manipular con el AD son las de las Intervenciones Quirúrgicas y Urgencias. Es importante señalar que este AD fue analizado [23] con la perspectiva del trabajo con atributos textuales y nosotros lo realizamos con una visión del diseño dimensional del AD.

a) Definición de la misión del caso de estudio.

La misión de la organización se compone por la definición y el propósito. Las técnicas que se utilizaron fueron las entrevistas y la inmersión en el negocio del cliente. Obteniéndose los siguientes elementos:

Definición: El sistema proporcionará los medios necesarios para analizar el Hospital Clínico “San Cecilio” de Granada, España.

Propósito: El objetivo del sistema, será proporcionar la información relativa a la conducta de las actividades de Intervenciones Quirúrgicas y Urgencias.

b) Identificación de las metas de análisis del caso de estudio.

El objetivo principal de esta etapa, consiste en identificar las metas y tareas que el usuario llevará a cabo con el AD. En un ambiente de AD estas tareas, se relacionan con los procesos de negocios a ser analizados.

Para lograr esto se proponen dos pasos:

Paso 1. Definición de los procesos de negocio.

En nuestro ejemplo los procesos de negocio que se pretenden analizar son: 1. Intervenciones Quirúrgicas y 2. Urgencias.

Paso 2. Identificar las metas de análisis.

Para la construcción del ARM, utilizamos la técnica de refinamiento de meta, utiliza un enfoque top-down para la identificación de las metas. En [3] se

utiliza un enfoque bottom-up. En nuestra propuesta consideramos que es necesario que estos pasos queden bien documentados, por lo que exponemos que se lleve una tabla que contenga toda la información, en la Tabla 2 se muestra cómo quedaría el caso de estudio.

Tabla 2. Resumen de las etapas a y b.

Misión	El sistema analizará el Hospital. Con el objetivo de proporcionar información sobre la conducta de las actividades de Intervenciones Quirúrgicas y Urgencias.		
Procesos de Negocio	Metas	Tarea	Usuario
Intervenciones Quirúrgicas	Meta 1: Calidad de la planificación de las intervenciones	T1: Analizar las intervenciones suspendidas con la intención de determinar las causas.	Gerente de intervenciones
	Meta 2: Efectividad de los tratamientos en las intervenciones	T2: Analizar las intervenciones realizadas con el fin de ver el comportamiento de este con el tratamiento.	Equipo de cirujanos
		T3: Analizar las intervenciones realizadas con el fin de ver el comportamiento de las anestésicas utilizadas.	
		T4: Analizar las intervenciones realizadas con el fin de ver la relación de esta con el diagnóstico.	
		T5: Analizar las intervenciones realizadas con el fin de ver el comportamiento de los implantes.	
Urgencias	Meta 3: Calidad de los servicios de urgencia	T6: Analizar el Tiempo que demora la atención de las urgencias con el fin de determinar las causas.	Gerente de Urgencia
	Meta 4: Efectividad y eficiencia de las urgencias.	T7: Analizar las urgencias con el fin de determinar el número de enfermos por las diferentes localidades.	
		T8: Analizar las urgencias para determinar las causas.	
		T9: Analizar las urgencias para determinar las áreas más afectadas.	
Árbol de Refinamiento de Metas			

c) Especificación de los requisitos del caso de estudio.

Esta actividad, consiste en identificar los requisitos de usuario final a partir de las metas identificadas en el ARM. Esta etapa, se realiza en tres pasos:

- 1) Descripción de las metas en casos de uso
- 2) En la figura 3 se muestran la descripción de las metas 2 en caso de uso.
- 3) Descripción de la interacción del usuario con el sistema en diagramas de actividad.

Caso de Uso	Meta 2: Efectividad de los tratamientos en las intervenciones
Actores	Equipo de cirujanos
Propósito	Permite a los usuarios analizar el comportamiento de las intervenciones y su relación con el diagnóstico.
Resumen	El análisis, consiste en el estudio de las diferentes intervenciones, este estudio le permitirá al usuario ver el comportamiento de las intervenciones, mostrando la cantidad de intervenciones. El análisis podrá realizarse por diagnóstico, sexo, la propuesta de tratamiento y tipo de anestesia.

Figura 3. Descripción de la meta 2.

En la figura 4 se muestra el diagrama de actividades para el caso de uso: Meta 2. Los datos son representados como elementos de información (EI). A partir de esta información, el sistema realizará algunas de las siguientes acciones: A) cantidad de intervenciones por cada tipo de anestesia que contiene un tipo determinado de propuesta de tratamiento. B) obtener la cantidad de intervenciones que contienen en su propuesta algún tipo tratamiento determinado por sexo.

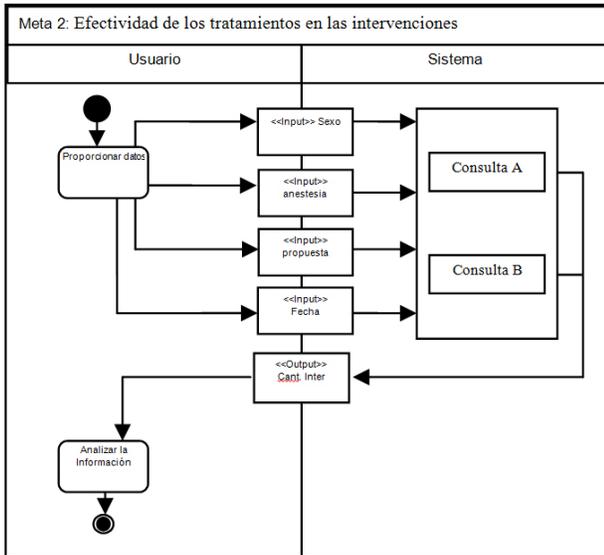


Figura 4. Diagrama de actividades para la meta 2.

En esta meta existen otras tareas pero por motivos de espacio solo representamos dos de estas.

Paso 3. Definición de los requisitos de información del caso de estudio.

Siguiendo la guía definida previamente, la información necesaria para el análisis de los datos relacionados con la Meta 2 se muestra en la figura 5:

Del análisis de estas metas se puede concluir que el modelo multidimensional será necesario contar con la siguiente información:

Dimensiones: Causa, Fecha, Sexo, Provincia, Procedencia, Asistencia, propuesta, anestesia, (Diagnóstico, Juicio, categoría y Área) Son las dimensiones que se obtienen con el análisis de las restantes metas.

Medidas: Cantidad de urgencias, Cantidad de intervenciones.

Meta 2	Efectividad de los tratamientos en las intervenciones
Cantidad de intervenciones	OutPut
Sexo	Input
anestesia	Input
propuesta	Input
Fecha	Input

Figura 5. Requisito de información para la meta 2.

2.1.3. Fase 2 – Analizar las base de datos operacionales:

Para la segunda fase nosotros proponemos utilizar una herramienta desarrollada en Eclipse, que ayuda al diseñador a crear el modelo del AD de forma semiautomática, nosotros proponemos que esta herramienta sea semiautomática debido a que a nuestro entender el diseño del AD no es siempre el mismo y no puede ser meramente una selección automática de algunos rasgos de los sistemas operacionales.

La herramienta trabaja con los esquemas de la(s) base(s) de dato fuente(s), esta soporta los gestores de base de datos Oracle, PostgreSQL, SQL Server, Microsoft Access y MySQL. La herramienta le permite al diseñador integrar las base de datos, el diseñador debe de seleccionar qué información de la base de dato fuente es de interés y que cumpla con las necesidades de información que se obtuvieron en la fase 1.

La herramienta cuando se conecta a la base de datos extrae la información que necesita y son guardados en dos fichero XML uno con los datos del esquema y otro con las relaciones existentes.

Luego de cargar toda la información en los XML, se le brinda al diseñador la posibilidad de seleccionar la información que es de interés para el AD. En nuestra propuesta al culminar esta selección queda un esquema refinado de los OLTP.

El objetivo general de esta etapa es que el diseñador y el usuario final puedan determinar la información que describa los eventos que ocurren dinámicamente en las empresas y que presentan la forma estructural de las empresas. En la figura 6 se muestra una imagen de la utilización de la herramienta en el caso de estudio.

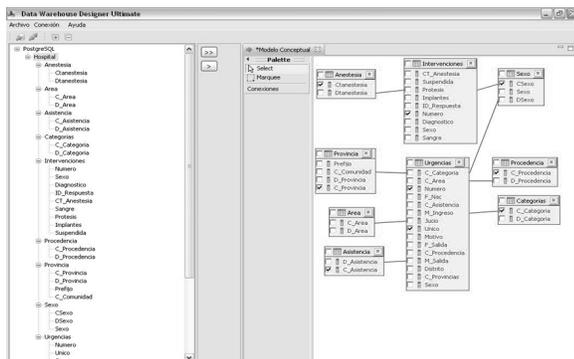


Figura 6. Imagen de la herramienta.

El resultado de esta etapa es la obtención de un esquema refinado de las bases de datos fuente en donde se integran todos los sistemas de información existente en la empresa, en la figura 7 se muestra este esquema refinado el cual es la base para las siguientes etapas.

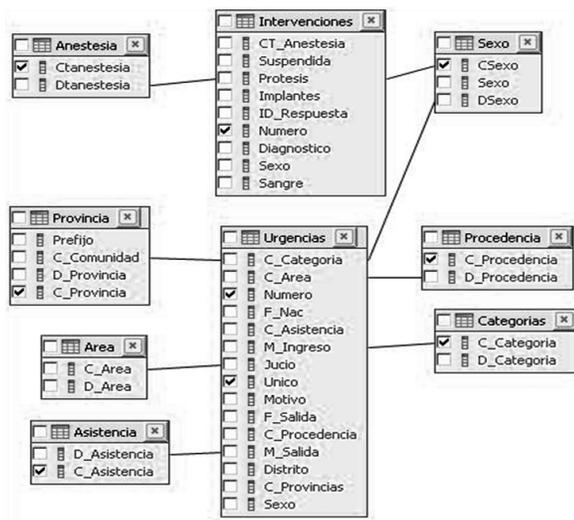


Figura 7. Esquema refinado.

Al culminar esta etapa el sistema proporciona dos informaciones detalladas que ha nuestro entender es un aspecto importante para lograr una documentación apropiada del diseño del AD, una información tiene que ver con la información de interés de los OLTP que seleccionó, en la tabla 3 se muestra cuál es la estructura de esta información.

Tabla 3. Información de interés seccionada.

Base de datos fuente	Tabla	Atributo	TD
PostgreSQL/Hospital	Anestesia	Ct anestesia	numer
		Dt anestesia	text
PostgreSQL/Hospital	Sexo	C Sexo	numer
		D sexo	text
		Sexo	boole

Otro aspecto que entendemos que es importante, está relacionado con las transformaciones que se realizan a la información de los OLTP, estas transformaciones pueden ser: cambiar el nombre a una tabla o a un atributo, eliminación de algún atributo o tabla del esquema refinado y atributos que relacionan dos OLTP (para el trabajo con base de datos heterogénea).

3. CONCLUSIONES

En este trabajo se mencionaron las pautas metodológicas necesarias para obtener un modelo asistido por computadora de AD, en este trabajo solo se explica la primera etapa análisis de la información, presentando dos fases fundamentales: Fase 1 - Especificación de requisitos de usuario. En donde se trabaja con una metodología basada en metas, la cual ha sido ampliamente aceptada para la captura de requisitos. Fase 2 – Analizar los modelos de las base de datos operacionales. En esta fase proponemos la utilización de nuestra herramienta para trabajar con las base de datos fuentes.

En futuros trabajos presentaremos la implementación de las restantes etapas de la pauta metodológica en la herramienta creada por nosotros. También se presentará un trabajo en donde se manipulan diferentes OLTP. También estamos trabajando para incorporar a la herramienta el soporte a otras fuentes de datos.

REFERENCIAS

- [1] Iimon, W. H., Building the Data Warehouse, Wiley Publishing, 2005.
- [2] Kimball, R. and Ross, M., The Data Warehouse Toolkit. The Complete Guide to Dimensional Modeling, Robert Ipsen, 2002.
- [3] Sánchez, L. Z., Metodología para el Diseño Conceptual de Almacenes de Datos[Tesis doctoral].Universidad Politécnica De Valencia,2008.
- [4] Arango, F., Gómez, M. C. and Zapata, C. M., Transformación del modelo de clases UML a oracle9i@ bajo la directiva MDA: un caso de estudio, Dyna. Medellín, Nro.149, pp. 165-179,2006.
- [5] Romero, O. and Abelló, A., A Survey of Multidimensional Modeling Methodologies, International Journal of Data Warehousing & Mining,, Nro.1-23,2009.
- [6] List, B., Bruckner, R. M., Machaczek, K. and Schiefer, J.A., Comparison of Data Warehouse Development Methodologies Case Study of the Process Warehouse.13th International Conference on Database and Expert Systems Applications (DEXA 2002),Aix-en-Provence, France,2453,2002
- [7] Marotta, A., Data Warehouse Design and Maintenance through Schema Transformations[Master Thesis]. Universidad de la República.,2000.
- [8] Akoka, J., Wattiau, I. C. and Prat, N., Dimension Hierarchies Design from UML Generalizations and Aggregations, Springer, Nro.500, pp. 442-455,2001.
- [9] Pardillo, J., Mazón, J. N. and Trujillo, J., An MDA Approach and QVT Transformations for the Integrated Development of Goal-Oriented Data Warehouses and Data Marts, Journal of Database Management (JDM), Nro.22, pp. 43-68, 2011.
- [10] Giorgini, P., Rizzi, S. and Garzetti, M., GRAnD: A Goal-Oriented Approach to Requirement Analysis in Data Warehouses, Decision Support Systems, Nro.45, 18,4,2008.
- [11] Peralta, V. and Marotta, A., Hacia la Automatización del Diseño de Data Warehouses.CLEI,2002
- [12] Road, H., Data Modeling Techniques for Data Warehousing, IBM Corporation, International Technical Support Organization, San Jose, California, 1998.
- [13] Golfarelli, M. and Rizzi, S., A Methodological Framework for Data Warehouse Design, ACM First International Workshop on Data Warehousing and OLAP, Washington, D.C., USA., 1998.
- [14] Moody, D. L. and Kortink, M. A. R., From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design,International Workshop on Design and Management of Data Warehouses (DMDW'2000),2000
- [15] Mazón, J. N. and Trujillo, J.,Desarrollo de modelos multidimensionales de almacenes de datos basado en MDA: del análisis de requisitos al modelo lógico,Zaragoza, España,2007
- [16] Pelozo, C. E., I.A Requirements Engineering Approach for Object-Oriented Conceptual Modeling[Doctor].Valencia Polytechnic University,2003.
- [17] Lamsweerde, A., V.Goal-Oriented Requirements Engineering: A Guided Tour.5th IEEE International Symposium on Requirements Engineering (RE'01),Toronto,August 2001,2001
- [18] Antón, A., Goal Based Requirements Analysis.2nd International Conference on Requirements Engineering (ICRE '96),Colorado, USA, April, pp. 15 - 18,1996
- [19] Estrada, H., Pastor, O., Martinez, A. and Torres, J., Using a Goal-Refinement Tree to Obtain and Refine Organizational Requirements. .Computational Science and its Applications (ICCSA'04),Assisi, Italy,3046 May pp. 14-17,2004
- [20] Valderas, P., Fons, J. and Pelechano, V., Using Task Descriptions for the Specification of Web Application Requirements.Workshop en Ingeniería de Requisitos (WER), Porto, Portugal,Junio, pp.13-14,2005
- [21] Stefanov, V., List, B. and Korherr, B., Extending UML 2 Activity Diagram with Business Intelligence Objects.7th International Conference on Data Warehousing and Knowledge Discovery (DaWaK'05) Copenhagen, Denmark,3589,August, pp. 22 - 26 2005,2005
- [22] Molina, J. G., Ortín, M. J., Moros, B., Nicolás, J. and Toval, A., Towards Use Case and Conceptual Models through Business Modeling.19th International Conference on Conceptual Modeling (ER'00),Utah, USA,1920,October, pp. 9 – 12 2000,2000
- [23] Ávila, E. T., Data Warehousing con procesamiento de datos textuales[TESIS DOCTORAL].Universidad de Granada,2010.