

Data Mart Para El Sistema De Servicios Sociales Del Conadis

Mosquera L.*; Hallo M.**

*Escuela Politécnica Nacional, Facultad de Ingeniería de Sistemas
Quito, Ecuador; e-mail: eduardo.mosquera; maria.hallo@epn.edu.ec

Resumen: El Consejo Nacional de Discapacidades Conadis, brinda servicios a nivel nacional y requiere conocer indicadores que demuestren el alcance de los servicios y permitan tomar decisiones para el futuro. Para apoyar este requerimiento se desarrolló un prototipo de data mart, para apoyar al proceso de toma de decisiones. Para el desarrollo del sistema, se utilizó como metodología el ciclo de vida de Ralph Kimball, el proceso cubrió las etapas de planificación, recolección de requerimientos, definición de la arquitectura y selección de herramientas, diseño lógico y físico del data mart, diseño y construcción del proceso ETL, diseño y construcción de aplicaciones de usuario final, despliegue del data mart. Todo el software utilizado para la construcción de los diferentes componentes del data mart son de libre difusión; esto incluye la base de datos, el servidor OLAP, la herramienta para ETL y las aplicaciones de usuario final. Entre las aplicaciones de usuario se destaca la aplicación Mapsic, la cual corresponde a una aplicación GIS en web desarrollada con la finalidad de mejorar la interacción de los usuarios con la información proporcionada por el data mart. La metodología fue útil y el sistema desarrollado resolvió los requerimientos urgentes detectados..

Palabras clave: Data mart, Conadis, Data Warehouse, Mapsic

Abstract: The National Council for disabilities Conadis, offers services at national level and require to know indicators showing the range of services to take decisions for the future. To solve this requirement a data mart prototype was developed to help in decision making process. The system was developed using the Ralph Kimball Life cycle methodology. The process include planning, requirements analyse, architecture definition, tools selection, logical and physical data mart design, ETL design, end user applications design and implementation, data mart deployment. The different data mart components are developed using free software such us database, OLAP server, ETL tools and end user applications. One of the end user application developed is Mapsi, a Web GIS application to improve the interaction between the user and the data mart information. The methodology was useful and the developed system solved the urgent requirements detected

Keywords: Data mart, Conadis, Data Warehouse, Mapsic).

1. INTRODUCCIÓN

Un *data mart* es un subconjunto de datos orientado al análisis de un área del negocio.

La propuesta de un *data mart* para el sistema de servicios sociales del CONADIS (Consejo Nacional de Discapacidades), nace de la necesidad de contar con información oportuna y precisa que respalde los procesos de la organización y ayude a la toma de decisiones en el área de las discapacidades, lo cual fue detectado en entrevistas a personeros del Conadis.

En el mundo del *data warehouse* son múltiples las soluciones orientadas al campo empresarial, negocios y ventas; sin embargo los campos de aplicación para esta tecnología son variados, para el caso de este proyecto, el campo de aplicación es el campo social, buscando una mejor comprensión y comunicación entre la organización y la comunidad.

Existen casos de éxito sobre la aplicación de las tecnologías de *data warehouse* en campos de ámbito social, a continuación se mencionan algunos de estos casos:

- El sistema de información demográfica (SID) desarrollado por el Instituto Nacional de Estadísticas de España, es un *data warehouse* que permite la tabulación detallada de variables demográficas hasta el nivel municipal con datos de estadísticas de variaciones Residenciales, sección censal y unidad poblacional[1].
- El *data mart* del *Health and Hospital Corporation of Marion Country* en Oregón - Estados Unidos, que a través de Internet presenta resultados instantáneos a solicitudes personalizadas de información de salud en el condado de Marion. Proporciona todas las estadísticas locales de salud, tales como

nacimientos, defunciones, enfermedades transmisibles, cáncer e información hospitalaria a través de Internet [2].

- El data warehouse del Instituto Nacional de Empleo en Madrid – España, proyecto desarrollado con el propósito de estudiar grandes volúmenes de información desde distintas perspectivas del mercado de trabajo para la toma de decisiones a nivel táctico y estratégico[3].

En este documento se muestran los aspectos y tareas más relevantes que se llevaron a cabo para la construcción de la propuesta de *data mart* para el sistema de servicios sociales del CONADIS.

2. METODOLOGIA

Para el desarrollo del *data mart* se utilizó la metodología de Ralph Kimball[4] con la siguientes etapas:

- Planificación del Proyecto, Gestión del proyecto
- Definición de Requerimientos
- Arquitectura Técnica, selección e instalación de productos.
- Modelo Dimensional, Diseño físico, Diseño ETL
- Especificaciones de Usuario Final, Desarrollo de aplicaciones
- Despliegue, Implantación, mantenimiento

2.1 Planificación del proyecto

El data mart apoyará a los procesos de la Dirección técnica y del Centro de información del CONADIS, procesos relacionados al sistema web SIC, tales como: registro nacional de discapacidades, solicitud y entrega de ayudas técnicas, solicitud y entrega de medicamentos o insumos, solicitud y autorización de vehículos para personas discapacitadas.

Las aplicaciones desarrolladas permiten realizar consultas interactivas y distribuir en el un mapa indicadores tales como, el número de personas con discapacidades, de tal forma que se puedan realizar programas convenientes de apoyo social de acuerdo a diversos factores.

Para el desarrollo del *data mart* se usó únicamente productos de software libre, manteniendo así los lineamientos gubernamentales para las administraciones públicas del Ecuador[5].

2.2 Definición de requerimientos

En esta etapa se definen los requerimientos que soportan todas las demás etapas del ciclo de vida del data mart, así: se definen requisitos para la arquitectura y selección de productos, requerimientos de análisis e información para el modelado dimensional, también se determinan especificaciones para las aplicaciones de usuario final.

Para llevar a cabo la definición de requerimientos, se recolectó información mediante las siguientes técnicas:

- Entrevistas al personal de la organización
- Análisis de documentos y sistemas fuente

La información obtenida incluye el ambiente de hardware y software bajo el cual trabaja la organización y estructura de los sistemas de información que manejan.

Se analizó una serie de consultas SQL (Structured Query Language) que normalmente se utilizan para obtener información de la base de datos del sistema web SIC, estas consultas generan reportes totalizados de personas discapacitadas, ayudas técnicas, medicamentos.

En base a estas consultas, se detectaron como indicadores claves:

- Número de personas discapacitadas
- Cantidad de ayudas técnicas solicitadas y entregadas
- Cantidad de medicamentos solicitados y entregados
- Cantidad de vehículos autorizados

Se detecto además datos descriptivos bajo los cuales se categoriza y totaliza la información, datos como: provincia, cantón, género, discapacidad, edad, tipo de ayuda técnica, tipo de medicamento, entre otros.

Se analizó la base de datos del sistema web SIC para identificar el origen de la información anteriormente descrita y buscar otros posibles indicadores y datos descriptivos; como resultado se obtuvo los siguientes indicadores adicionales:

- Porcentaje de discapacidad de una persona
- Edad de la discapacidad de una persona
- Monto y subvención de ayudas técnicas
- Monto y subvención de medicamentos

- Cantidad de tiempo que lleva autorizar un vehículo

2.3 Arquitectura técnica del data mart.

En esta etapa del ciclo de vida se busca consolidar tres factores:

- Los requerimientos de la organización.
- El ambiente actual de hardware y software.
- Los componentes requeridos para la implementación del *data mart*

Para esta tarea se consideraron dos escenarios:

Back Room.- Encapsula todos los procedimientos y recursos necesarios para llevar a cabo el tratamiento de los datos, básicamente los procesos ETL(Extract, Transform and Load).

Front Room.- Encapsula todos los procedimientos y recursos necesarios para presentar los datos ya tratados al usuario final.

En medio de estos dos escenarios se ubica el servidor de presentación y los metadatos; la figura 2 muestra la ubicación y relación que existe entre los componentes que conforman la arquitectura del *data mart*.

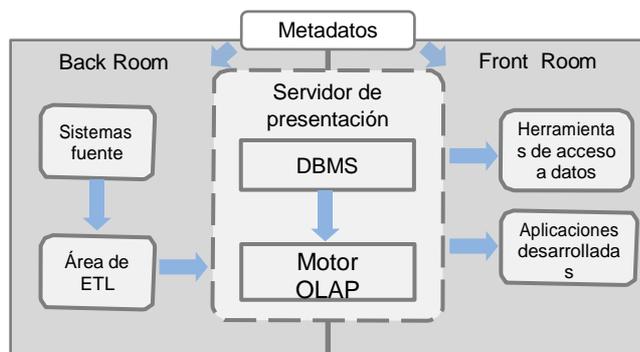


Figura 1. Arquitectura de un *data mart*. Extraído de [6]

La arquitectura presenta dos tipos de componentes fundamentales, servicios y repositorios de datos.

Servicios.- Funciones necesarias para llevar a cabo las tareas requeridas en el *data mart*.

Repositorios de datos.- Lugares temporales o permanentes en los cuales se almacenan los datos.

Los servicios para el *Back Room* son los siguientes:

Servicio de extracción de datos.- Es el encargado de obtener los datos de los sistemas fuente.

Servicio de transformación de datos.- Cambios que sufren los datos necesarios para el rendimiento del *data mart*, entre estas actividades están: generación de claves subrogadas, mapeo de claves de un sistema a otro, desnormalización de tablas, conversión de los tipos de datos, cálculo de valores.

Servicio de carga de datos.- Es el encargado de almacenar los datos en los sistemas destino.

Servicio de control de tareas.- Ayuda a controlar los errores y a manejar las excepciones mediante la definición y programación de tareas, monitoreo y registro de eventos.

Los repositorios de datos para el *Back Room* son los siguientes:

Sistemas fuente.- Fuente de datos que alimenta el *data mart*, sobre este repositorio de datos ocurre el servicio de extracción de datos.

Área de ETL.- Repositorio temporal de datos, aquí tiene lugar el proceso de transformación de datos y el servicio de control de tareas, el cual consiste en la calendarización y monitoreo de los procesos ETL.

DBMS.- Sistema de administración de bases de datos que administra la base de datos del *data mart*.

Los servicios para el *Front Room* son los siguientes:

Servicio de navegación sobre el data mart.- El proceso de navegación a través de la información es uno de los procesos más importantes del *Front Room*, permite al usuario mediante una herramienta del sistema buscar algún dato que le sea importante o necesite saber.

Servicio de acceso y seguridad.- Estos servicios también pueden ser llamados de autorización y autenticación, y ambos se refieren a las conexiones que los usuarios necesitan para acceder a las bases de datos. La autenticación se refiere a alguna técnica para verificar que la persona es quien dice ser, este servicio puede ser controlado por otro sistema.

Servicio de administración de consultas.- La administración de consultas le permite al usuario una interacción con las bases de datos.

Servicio de reportes estándar.- El servicio de reportes estándar le permite al usuario generar reportes con formatos estándar.

Los repositorios de datos para el *Front Room* son los siguientes:

Motor OLAP.- Genera los cubos y dimensiones con la información almacenada en la base de datos del *data mart* y procesa las consultas analíticas, responde al

Servicio de administración de consultas MDX (Multidimensional Expressions).

Herramientas de acceso a datos.- Herramientas que permiten a los usuarios interactuar con los datos almacenados en el *data mart*, mediante consultas MDX y SQL, además deben proporcionar los servicios de navegación sobre el *data mart* y servicio de reportes estándar.

Aplicaciones desarrolladas.- Para este caso se planteó la construcción de una aplicación web que permita al usuario interactuar con los datos almacenados en el *data mart* mediante consultas SQL, valiéndose de un componente GIS (Geographic Information System) para representar la información de manera gráfica en un mapa.

Los metadatos del *Back Room* están relacionados a los procesos ETL, guían la extracción, limpieza y carga de datos; por lo tanto entre estos metadatos esta la información acerca de los sistemas fuente y el área ETL.

Los metadatos del *Front Room* son más descriptivos, ayudan a las herramientas de consulta, nos permite entender de donde vienen los datos; estos metadatos cubren información acerca del *data mart* y herramientas de acceso a datos.

Selección de productos

En esta etapa del ciclo de vida, se busca elegir herramientas y aplicaciones de software apropiadas para la implementación de la arquitectura propuesta.

Los productos seleccionados son de software libre, para seguir los lineamientos de la organización y además satisfacen los servicios tanto para el *Back Room* como para el *Front Room*.

La tabla 1 resume los productos de software elegidos para cada componente de la arquitectura del Data mart.

Tabla 1. Productos seleccionados para el Data mart	
Back Room	
DBMS	PostgreSQL 8.4
Área ETL	Pentaho data intregation – Kettle 3.2
Front Room	
Motor OLAP	Mondrian 3.1.6
Acceso a Datos	Openi 2.0 RC2
Otros productos necesarios	
Servidor de aplicaciones java	Tomcat 6.0
Servidor web	Apache 2.2
Servidor GIS	Mapserver 3.0

2.4 Modelado dimensional

Cada modelo dimensional esta compuesto por una tabla cuya clave es compuesta, llamada tabla de hechos, y un grupo de tablas pequeñas llamadas tablas de dimensiones.

Cada tabla de dimensión tiene una clave primaria singular que corresponde exactamente con uno de los componentes de la clave compuesta de la tabla de hechos, esta estructura adopta el nombre de modelo en estrella.

En base a la información descriptiva de las consultas SQL analizadas en la etapa de requerimientos, se determinaron las dimensiones agrupando los atributos descriptivos que están relacionados[7]. Se definieron un total de 13 dimensiones.

Para la conformación de las tablas de hechos se agruparon las medidas asociadas a los indicadores claves detectados en la etapa de requerimientos, la agrupación se realizó en función del proceso asociado al sistema web SIC. Se determinó las siguientes tablas de hechos:

- Tabla de hechos de personas discapacitadas
- Tabla de hechos de ayudas (agrupa ayudas técnicas y medicamentos)
- Tabla de hechos de autorización de vehículos

Para cada tabla de hechos se determinó la granularidad, se seleccionaron las dimensiones y medidas. Las figuras 3, 4 y 5 muestran los diagramas de las Tablas de hechos conformadas[8].

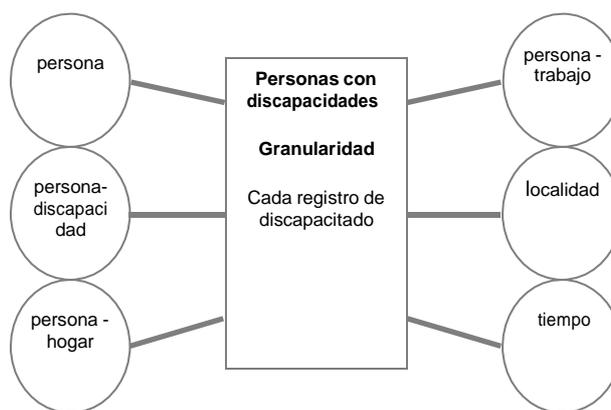


Figura 2. Diagrama de la tabla de hechos de personas con discapacidades.

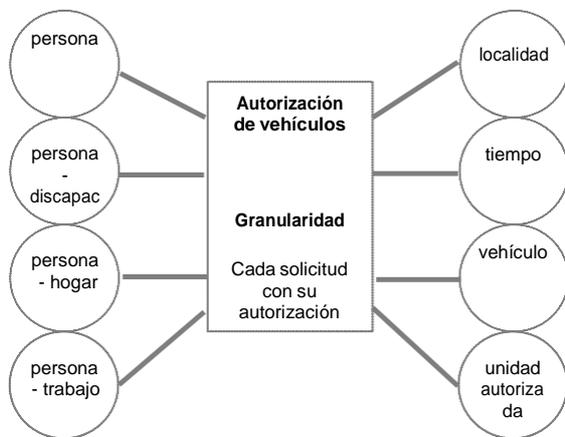


Figura 3. Diagrama de la tabla de hechos de autorización de vehículos.

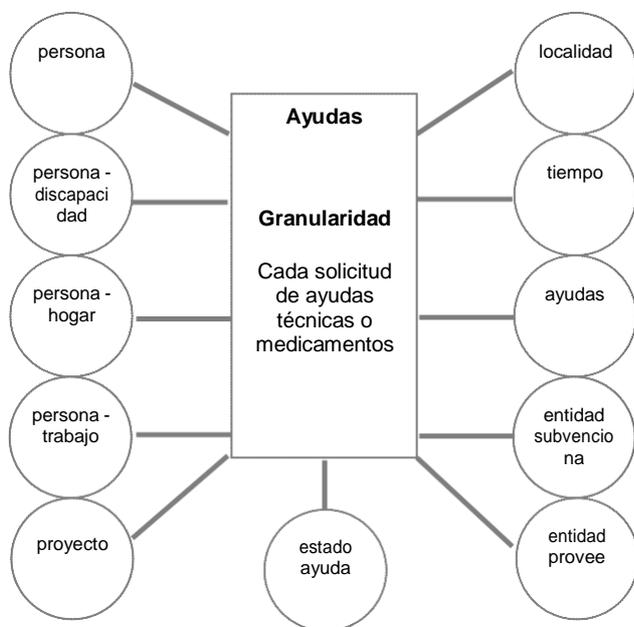


Figura 4. Diagrama de la tabla de hechos de ayudas.

2.5 Diseño físico

En esta etapa del ciclo de vida se contemplan los pasos para convertir el diseño lógico en una base de datos física, las dimensiones y tablas de hechos se definen como tablas de la base de datos.

Como primera tarea se elaboró estándares para nombrar los objetos de la base de datos y del ETL.

Mediante el uso de una herramienta de modelado de datos, se paso el modelo lógico a un modelo físico

respetando los estándares definidos para los objetos de la base de datos, tipos de datos, etc.

La herramienta permitió generar el código necesario para crear la base de datos sobre PostgreSQL.

Como segunda tarea se elaboró el plan de índices, este plan consideró los siguientes índices:

Índices para las Tablas de hechos.- Corresponde a un índice compuesto por todas las claves dimensionales de la Tabla de hechos, también se definieron índices para cada clave dimensional por separado. Estos índices tienen como propósito mejorar el rendimiento de las consultas, el planificador de consultas del DBMS es el encargado de utilizar convenientemente estos índices.

Índices para las tablas de Dimensiones.- Se creó un índice para la clave subrogada de la tabla, índices para los atributos e índices compuestos para las jerarquías entre atributos utilizando las claves primarias asociadas tomadas de la base de datos fuente. Estos índices tienen como propósito mejorar el rendimiento de las consultas, especialmente cuando se declaran filtros y ordenamientos.

Índices para el proceso ETL.- Para optimizar la carga y actualización de datos en las tablas de Dimensiones y especialmente en las Tablas de hechos, se definieron índices simples y compuestos por las claves operacionales en las tablas de Dimensiones para mejorar el proceso de obtención de claves subrogadas.

Para asegurar el correcto funcionamiento del planificador de consultas, los índices y estadísticas deben estar actualizados, esta operación se definió posteriormente como parte del proceso ETL.

Como tercera y última tarea se construyeron los cubos OLAP (Online Analytical Processing), Como fue determinado en la selección de productos, el motor OLAP a ser utilizado es Mondrian, por lo tanto la construcción de cubos se la realizó mediante el esquema de etiquetas XML propuesto por Pentaho – Mondrian. La figura 6 muestra la estructura básica del esquema[9].

```
<Schema>
  <Cube>
    <Dimension>
      <Hierarchy>
        <Level/>
      </Hierarchy>
    </Dimension>
    <DimensionUsage/>
    <Measure />
  </Cube>
</Schema>
```

Figura 5. Estructura básica del esquema Mondrian

2.6 Diseño y construcción de los ETLs

En esta etapa se desarrolló el *Back Room*, especialmente en el área ETL donde se definen los procesos de extracción, limpieza, transformación y carga de datos desde la base de datos fuente hacia la base de datos del *data mart*.

Como primera tarea se elaboró el plan ETL[10], este plan detalla la forma como se realizó la carga de datos desde los sistemas fuente hacia el *data mart*. Se plantearon tres procesos y son los siguientes:

Carga inicial de datos.- Este proceso carga datos sobre el *data mart* por primera vez, el primer trabajo consiste en la creación de la base de datos, el segundo trabajo consiste en llenar la dimensión tiempo, como tercer paso se ejecutan los trabajos para llenar el resto de dimensiones y finalmente se ejecutan los trabajos para llenar las tablas de hechos.

Debido a que se trata de la carga inicial, los registros que se cargan a las tablas de hechos y dimensiones comprenden desde los más antiguos hasta los ingresados en el sistema operacional la fecha en que se ejecuta la carga inicial.

Carga incremental de datos.- Este proceso se ejecuta periódicamente y permite mantener actualizadas las dimensiones cargando nuevos registros, así como actualizando los datos ya registrados de acuerdo a la política de cambio que se planteo para cada atributo.

En el caso de las tablas de hechos, se cargan los nuevos registros comprendidos entre la última ejecución de la carga incremental o inicial hasta la fecha anterior a la nueva ejecución.

Actualización de datos.- Este proceso permite actualizar las dimensiones cargando nuevos registros así como actualizando los antiguos.

En el caso de las tablas de hechos, aunque es poco común realizar actualizaciones sobre los registros, este proceso permite actualizar los registros comprendidos entre dos fechas, una inicial y una final suministradas por el usuario.

Como siguiente tarea se determinó el flujo de datos base tanto para dimensiones como tablas de hechos, este flujo muestra las etapas por las cuales pasan los datos desde que son extraídos del sistema operacional hasta que son cargados al *data mart*. También se identificaron las transformaciones y cálculo de datos, muchas de estas operaciones se realizan como parte de la extracción de datos.

El flujo de datos base para la carga y actualización de las dimensiones[11] consta de los siguientes pasos:

- 1) *Extracción de datos de las tablas origen:* Mediante código SQL se seleccionan las columnas que proveen los datos, los nombres de las columnas se las redefine conforme al nombre de las columnas de la tabla de dimensión. En este paso se lleva a cabo una primera parte de las transformaciones.
- 2) *Transformaciones y cálculos:* Los registros entran a este paso y se llevaron a cabo las transformaciones y cálculos definidos, de ser necesario este paso añade columnas a los registros.
- 3) *Asignación de la clave subrogada:* En base a las claves operacionales se asigna un número secuencial único al registro, esta es la clave subrogada.
- 4) *Carga y actualización de datos:* Los registros del paso anterior son cargados a la tabla de dimensión, los registros nuevos son insertados mientras que los registros alterados son actualizados.

El flujo de datos base para la carga y actualización de las tablas de hechos consta de los siguientes pasos:

- 1) *Extracción de datos de las tablas origen:* Mediante código SQL se seleccionan las claves operacionales correspondientes a las dimensiones asociadas a la tabla de hechos, se seleccionan las columnas que proveen los datos para las medidas, los nombres de estas columnas se las redefine para que coincidan con las medidas de la tabla de hechos. En este paso se lleva a cabo una primera parte de las transformaciones esencialmente el cálculo de algunas medidas. Los registros seleccionados están dentro de un rango de fechas, las cuales dependen del tipo de proceso ETL.
- 2) *Transformaciones y cálculos:* Se llevaron a cabo las transformaciones y cálculos definidos, de ser necesario este paso añade columnas a los registros y realiza validaciones para comprobar la integridad de los cálculos, en caso de existir datos incoherentes estos son exportados a un archivo para su posterior análisis.
- 3) *Obtención de claves dimensionales:* Se busca reemplazar las claves operacionales con sus respectivas claves dimensionales, las cuales se añaden como nuevas columnas a los registros. Las claves operacionales son removidas de los registros.

- 4) *Validación de claves:* Las claves obtenidas en el paso anterior son sometidas a una verificación para asegurar que no hayan datos nulos, en caso de hallarse registros con claves nulas estos son exportados a un archivo para su posterior análisis. Los registros que pasan la verificación son ordenados en base a las claves dimensionales.
- 5) *Extracción de datos de la tabla de hechos:* Mediante código SQL se seleccionan todas las columnas de la tabla y se las ordena por las claves dimensionales, los registros seleccionados están dentro de un rango de fechas, las cuales dependen del tipo de proceso ETL.
- 6) *Comparación y Marcado de filas:* Los registros de los pasos 4 y 5 son comparados en base a las claves dimensionales y dependiendo del resultado de la comparación se marca cada registro como: idéntico, alterado, nuevo, borrado; esta información se almacena en una nueva columna la cual se añade a los registros.
- 7) *Actualización de Registros:* Los registros del paso anterior son conmutados dependiendo de la marca, los nuevos son insertados, los alterados son actualizados y los borrados son eliminados de la tabla de hechos.

2.7 Construcción de las aplicaciones de usuario

Al iniciar esta etapa del ciclo de vida, nos concentraremos en el *Front Room*.

La principal actividad de desarrollo que se lleva a cabo en el *Front Room* es la creación de un primer grupo de reportes, los cuales adoptan el nombre de plantillas de la aplicación de usuario [10], un segundo esfuerzo está enfocado en la creación de la aplicación Mapsic.

El usuario interactúa con el Data mart mediante los elementos del *Front Room*, es decir mediante herramientas de acceso a datos y aplicaciones desarrolladas sobre el Data mart. Como fue determinado en la selección de productos para el *Front Room*, la herramienta de acceso a datos es el sistema de reportes web Openi, adicionalmente se planteó el desarrollo de una pequeña aplicación GIS para realizar consultas interactivas.

Como primera tarea se elaboraron las plantillas de la aplicación de usuario, estos reportes fueron construidos tomando como referencia los reportes observados en la sección de estadísticas del sitio web del CONADIS, se

definieron los siguientes reportes como plantillas iniciales:

Análisis cruzado

- Causa de deficiencia por edades
- Distribución local de PCD (personas con discapacidades) por situación laboral
- Distribución local de ayudas técnicas

Distribución por categoría

- Distribución de PCD por deficiencia principal
- Tiempo promedio para autorizar un vehículo

Promedio por categoría

- Porcentaje de PCD por género

Tendencias por categoría

- PCD registradas por año
- Monto total en ayudas entregadas al año
- Vehículos autorizados al año

Como segunda tarea se desarrolló la aplicación Mapsic, esta aplicación web tiene como propósito permitir a los usuarios realizar consultas interactivas, obtener información sobre el número y porcentaje de personas discapacitadas y su distribución provincial y cantonal en el Ecuador.

Para el desarrollo se siguieron las etapas de requerimientos, análisis, diseño, construcción y pruebas, de entre estas etapas se resaltó el diseño de la arquitectura, que corresponde a un diseño en cuatro capas. La tabla 2 muestra en detalle en que consiste cada capa.

Tabla 2. Capas de la aplicación - Mapsic	
Capa	Descripción
Presentación	Interfaz de usuario estilo web, utilizando xhtml, css y javascript. Presenta ayudas visuales tales como: resaltado de filas y columnas, ordenamiento por columnas, en los reportes. Paneles expandibles e indicadores de carga.
Comunicación	Las llamadas al servidor son asincrónicas, el refrescamiento del contenido de la pagina web es mediante Ajax.
Lógica	Procedimientos para la generación de reportes, procedimientos para la generación de mapas (GIS). Definición de constantes y variables de sesión.
Datos	Procedimientos para acceder y realizar consultas a bases de datos. Procedimientos para acceder a shape files (mapas vectoriales). Definición de metadatos

2.8 Despliegue

El despliegue es la convergencia de tecnología, datos y aplicaciones; es la integración de los componentes que se han ido desarrollando en el transcurso del ciclo de vida del *data mart*, con el fin de que los usuarios puedan empezar a sacar provecho.

Esta etapa también contempla capacitación para los usuarios y la preparación del ambiente para poder integrar cada uno de los componentes del *data mart*.

Entre las tareas llevadas a cabo en esta etapa final está la elaboración y ejecución de pruebas para validar el funcionamiento de los componentes del *data mart*, como son los procesos ETL, los reportes iniciales y la aplicación Mapsic.

3. CONCLUSIONES

El ciclo de vida propuesto por Kimball facilita las tareas de desarrollo de un *data warehouse*, en este caso las del *data mart*, define los productos que se deben obtener tras finalizar cada etapa, además de presentar plantillas, formatos y diagramas para la documentación.

El *data mart* mejora el trabajo de los usuarios poniendo a su disposición la información que necesitan cuando la necesitan, por lo tanto es importante seleccionar herramientas apropiadas que se ajusten a las necesidades de la organización.

Una de las actividades más importantes dentro del desarrollo del *data mart* es la definición del proceso ETL, especialmente el aseguramiento de la calidad de los datos, por lo tanto es importante contar con sistemas fuentes confiables que validen y controlen el registro de datos, estos mecanismos facilitan en gran medida el tratamiento de los datos.

El esfuerzo aplicado en estructurar consultas SQL para el análisis de datos es alto, especialmente en bases de datos transaccionales esta tarea es compleja; ahora mediante el *data mart* el cual presenta una base de datos apropiada para realizar análisis y consultar grandes cantidades de datos, el estructurar consultas en SQL involucra menos esfuerzo, sin embargo es más eficaz utilizar MDX mediante los cubos OLAP.

El trabajo ha permitido elaborar y validar una solución de apoyo a la toma de decisiones para el CONADIS utilizando software libre.

REFERENCIAS

- [1] Plan de actuación del Instituto Nacional de Estadísticas para el año 2009, <http://www.mzes.uni-mannheim.de/eda/e/series/2242129-4/2009.pdf> (accesado noviembre 2013).
- [2] Missouri Department of Health, World Access to Indy's Health, <http://www.hhcdata mart.com/> (accesado julio 2013).
- [3] INEM, Data warehouse del Instituto Nacional de Empleo en Madrid, http://foil.oit.or.cr/documentos/index.php?option=com_k2&view=foildoc&id=224:sistema-de-informacion-data-warehouse&Itemid=55 (accesado julio 2013)
- [4] KIMBALL Ralph, *et al.* The Data Warehouse Lifecycle Toolkit. 1a ed. New York: Wiley. 1998. Capítulo 2, p. 33 – 36.
- [5] CORREA Rafael, Decreto 1014, Registro Oficial 322, Quito: Editora Nacional, 2008.
- [6] KIMBALL Ralph, *et al.* The Data Warehouse Lifecycle Toolkit. 1a ed. New York: Wiley. 2001. Capítulo 8, p. 327 – 330, Capítulo 9, p. 336 – 364, Capítulo 10, p. 375 – 386.
- [7] KIMBALL Ralph, ROSS Margy. The Data Warehouse Toolkit. 2a ed. New York: Wiley. 2002. Capítulo 6, p. 154-157.

- [8] KIMBALL Ralph, *et al.* The Data Warehouse Lifecycle Toolkit. 1a ed. New York: Wiley. 1998. Capítulo 5, p. 194 – 197.
- [9] Mondrian Documentation, How to Design a Mondrian Schema.
<http://mondrian.pentaho.org/documentation/schema.php>. 2011-06-24.
- [10] KIMBALL Ralph, *et al.* The Data Warehouse Lifecycle Toolkit. 1a ed. New York: Wiley. 1998. Capítulo 16, p. 612 – 615.
- [11] KIMBALL Ralph, CASERTA Joe. The Data Warehouse ETL Toolkit. 1a ed. New York: Wiley. 2004. Capítulo 5.