

**Universidad de las Ciencias Informáticas**

**Facultad 6**



**Título:** Mercado de Datos Series históricas objetivos cubanos del milenio en el área de Información para el Sistema de Información de Gobierno

Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas

**Autora:**

Leisy Basulto Montero

**Tutores:**

Ing. Haymeé Llerena Esperón

Ing. Lully Díaz Montero

**Asesor:**

Ing. Yosvany Arrastia Machin

La Habana 2012

“Año 54 de la Revolución”

## FRASE

*Los ideales que han iluminado mi camino una y otra vez me han infundido valor para enfrentarme a la vida. Han sido la bondad, la belleza, y la verdad.*

*Albert Einstein*



## DECLARACIÓN DE AUTORÍA

Declaro ser autora del presente trabajo Mercado de Datos Series históricas objetivos cubanos del milenio en el área de Información para el Sistema de Información de Gobierno (SIGOB) y reconozco a la Universidad de las Ciencias Informáticas (UCI) los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmo la presente a los \_\_\_\_\_ días del mes de \_\_\_\_\_ del año 2012.

---

Leisy Basulto Montero

Autora

---

Haymeé Llerena Esperón

Tutora

---

Luilly Díaz Montero

Tutor

## **DATOS DE CONTACTO**

Tutora: Ing. Haymeé Llerena Esperón

Email: [hllerena@uci.cu](mailto:hllerena@uci.cu)

Graduada en la Universidad de las Ciencias Informáticas en 2007

Tutor: Ing. Lully Díaz Montero

Email: [ldmontero@uci.cu](mailto:ldmontero@uci.cu)

Graduado en la Universidad de las Ciencias Informáticas en 2011

Asesor: Ing. Yosvany Arrastia Machin

Email: [yarrastia@uci.cu](mailto:yarrastia@uci.cu)

Graduado en la Universidad de las Ciencias Informáticas en 2010

## **AGRADECIMIENTOS**

*A mi mamá, por su total entrega, esfuerzo y experiencia brindada para poder culminar este proyecto, por ser la máxima inspiración de mi carrera, mi modelo a seguir en todos los aspectos de la vida y siempre esperar lo mejor de mí.*

*Al profesor Ing. Yosvany Arrastia Machin, por su paciencia, guía y dedicación durante todo el proceso investigativo, su ayuda incondicional y principalmente por ser mi ángel.*

*A mi papá y hermana, por todo el apoyo y la confianza depositada en mí, por impulsarme a luchar por mis sueños y lo mucho que me han enseñado e.*

*A mis amigos Liván, Grether, Osmil y Chemita, por su eterna amistad y apoyo incondicional, por las noches que se mantuvieron despiertos hasta tarde para ayudarme.*

*A mis grandes amigas Dary y Yeny, que supieron apoyarme cuando más lo necesité, por siempre estar a mi lado y mostrarme lo que valgo.*

*A mis amistades del eterno grupo 2 de la Facultad Regional de Artemisa, que me acompañaron durante 4 años y me enseñaron a ser mejor persona.*

*A mis tutores Ing. Haymeé Llerena Esperón e Ing. Luilly Díaz Montero, por la experiencia transmitida.*

*A mis profesores, por su contribución en mi formación profesional.*

*A la MSc. Lourdes Cereza Tamargo, por sus conocimientos y apoyo profesional brindado.*

*A todos aquellos que de una forma u otra me ayudaron con el desarrollo de este trabajo.*

## **DEDICATORIA**

*A mis padres y hermana, que me han inculcado el amor al estudio y me han brindado todo su amor, esfuerzo y dedicación.*

*A la memoria de mi abuela Esperanza, que donde quiera que esté se encuentre orgullosa de mí.*

## RESUMEN

El presente trabajo de diploma ha sido desarrollado para el área de Información perteneciente a la Oficina Nacional de Estadísticas e Información (ONEI). Esta área presenta deficiencias en el manejo de la información estadística relacionada con los Objetivos de Desarrollo del Milenio (ODM) del país. Por ello, se propone como objetivo general de investigación: desarrollar el Mercado de Datos Series históricas objetivos cubanos del milenio en el área de Información para el Sistema de Información de Gobierno a fin de contribuir a la toma de decisiones. Para dar cumplimiento a lo anteriormente mencionado, en el presente trabajo se fundamentaron las metodologías, herramientas y tecnologías a emplear, realizándose el levantamiento de requerimientos junto con las necesidades de información del cliente. Se definió la arquitectura del Mercado de Datos, identificando los principales subsistemas de: integración, almacenamiento y visualización que la componen. Se diseñaron e implementaron los procesos de Extracción, Transformación y Carga e Inteligencia de Negocio. Mediante los casos de pruebas y las listas de chequeos se logró verificar el correcto funcionamiento de la aplicación. Finalmente, se obtuvo como resultado un Mercado de Datos poblado y funcional, disponible para hacer análisis estadísticos, ofrecer reportes actualizados y contribuir a la toma de decisiones sobre la información que se plantea como objeto de estudio, cumpliendo de esta forma con las necesidades del cliente.

Palabras claves: Almacén de Datos, Mercado de Datos, ONEI, toma de decisiones.

## ÍNDICE

AGRADECIMIENTOS.....	I
DEDICATORIA .....	II
RESUMEN.....	III
INTRODUCCIÓN.....	1
CAPÍTULO 1: FUNDAMENTOS TEÓRICOS .....	5
1.1.    Introducción.....	5
1.2.    Almacenes de Datos .....	5
1.2.1.    Definición de los Almacenes de Datos.....	5
1.2.2.    Características de los Almacenes de Datos.....	6
1.2.3.    Ventajas y desventajas de los Almacenes de Datos .....	6
1.3.    Mercados de Datos .....	7
1.3.1.    Definición.....	7
1.3.2.    Ventajas de los Mercados de Datos.....	8
1.4.    Estructura de los Almacenes de Datos .....	8
1.4.1.    Modelo de datos .....	9
1.4.2.    Modos de almacenamiento.....	11
1.4.3.    Integración de datos .....	13
1.4.4.    Inteligencia de Negocios.....	13
1.5.    Metodología de desarrollo .....	14
1.5.1.    Grupos y flujos de trabajo .....	16
1.6.    Herramientas.....	17
1.6.1.    Herramienta de modelado.....	17
1.6.2.    Sistema Gestor de Bases de Datos .....	18
1.6.3.    Administrador de bases de datos.....	20
1.6.4.    Herramienta para la limpieza de los datos .....	20
1.6.5.    Herramienta para la integración de los datos.....	20
1.6.6.    Herramientas de Inteligencia de Negocio.....	21
1.7.    Conclusiones.....	22
CAPÍTULO 2: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS SERIES HISTÓRICAS OBJETIVOS CUBANOS DEL MILENIO EN EL ÁREA DE INFORMACIÓN .....	24
2.1.    Introducción.....	24
2.2.    Análisis.....	24
2.2.1.    Descripción del negocio.....	24



2.2.2.	Flujo de información.....	25
2.2.3.	Necesidades de información.....	26
2.2.4.	Especificación de requerimientos.....	26
2.2.5.	Modelo de Casos de Uso del Sistema .....	28
2.2.6.	Reportes candidatos.....	30
2.3.	Arquitectura de los subsistemas.....	31
2.4.	Diseño.....	32
2.4.1.	Identificación de dimensiones, hechos y medidas.....	32
2.4.2.	Matriz bus o dimensional .....	34
2.4.3.	Modelo de datos .....	34
2.4.4.	Perfilado de datos.....	34
2.4.5.	Diseño del subsistema de integración.....	35
2.4.6.	Diseño del subsistema de visualización.....	36
2.5.	Esquema de seguridad.....	37
2.6.	Política de respaldo y recuperación.....	38
2.7.	Conclusiones.....	38
<b>CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBAS DEL MERCADO DE DATOS SERIES HISTÓRICAS</b>		
<b>OBJETIVOS CUBANOS DEL MILENIO EN EL ÁREA DE INFORMACIÓN .....</b>		
		<b>40</b>
3.1.	Introducción.....	40
3.2.	Implementación del modelo de datos .....	40
3.2.1.	Estructura de los datos .....	40
3.2.2.	Usuarios y privilegios.....	41
3.3.	Implementación del subsistema de integración .....	41
3.3.1.	Arquitectura de integración.....	41
3.3.2.	Implementación de las transformaciones y los trabajos .....	42
3.3.3.	Control en los procesos de integración de datos .....	43
3.4.	Implementación del subsistema de visualización.....	43
3.4.1.	Implementación de los cubos OLAP .....	44
3.4.2.	Implementación de los reportes candidatos.....	44
3.4.3.	Arquitectura de información .....	45
3.4.4.	Configurar la seguridad de los usuarios.....	47
3.5.	Pruebas.....	47
3.5.1.	Pruebas de software.....	47
3.5.2.	Modelo V .....	47

3.5.3. Herramientas de prueba .....	49
3.5.4. Pruebas realizadas al MD.....	49
3.6. Conclusiones.....	52
CONCLUSIONES .....	54
RECOMENDACIONES .....	55
REFERENCIAS BIBLIOGRÁFICAS .....	56
BIBLIOGRAFÍA.....	58
ANEXOS.....	61
GLOSARIO .....	70

## INTRODUCCIÓN

En la era moderna, con el surgimiento y evolución de las Tecnologías de la Información y las Comunicaciones (TIC), las sociedades se han venido transformando rápidamente. Las empresas necesitan registrar, manejar y controlar sus grandes volúmenes de datos con mayor velocidad y precisión, buscando la forma de poder transformarlos en referencias muy útiles para consultas, análisis y toma de decisiones. Es por ello que la informática busca el mejor tratamiento de la información, requiriendo para ello el uso de técnicas y herramientas innovadoras que representen un salto cuantitativo y cualitativo en la automatización y almacenamiento de los datos.

Cuba al igual que el resto de las sociedades modernas, precisa recopilar y procesar correctamente la información estadística de todos los sectores económicos y sociales del país. Esta labor se encuentra dirigida y centralizada por la Oficina Nacional de Estadísticas e Información (ONEI), entidad rectora del Sistema de Información de Gobierno (SIGOB) en Cuba, la cual es responsable en integrar el trabajo estadístico estatal del país en un potente sistema; proponer, organizar y ejecutar la aplicación de la política estatal en materia de estadística, velando por la autenticidad de su información. Como parte del servicio estadístico que ofrece esta institución, específicamente en el área de Información, se evalúa el estado de cumplimiento de los Objetivos de Desarrollo del Milenio (ODM) del país.

Los ODM tienen su origen en la Declaración de la Cumbre del Milenio firmada por los países miembros de las Naciones Unidas, entre ellos Cuba. En esta declaración quedó expresado un acuerdo internacional que constituye el compromiso de desarrollo humano a alcanzar hasta el año 2015. Estos Objetivos están integrados por metas concretas e indicadores que apuntan a la lucha a erradicar la pobreza, el mejoramiento de la educación y la salud, la promoción de la dignidad humana y la igualdad, alcanzar la paz y la sostenibilidad ambiental a escala global. Para Cuba los ODM constituyen una prioridad, pues permiten medir los logros de desarrollo alcanzados, fundamentalmente los relacionados con la educación y la salud, como resultado del modelo social adoptado y de las políticas sociales puestas en prácticas desde 1959 hasta el presente; representan un desafío de desarrollo a seguir a fin de situar al ser humano como centro de atención. (1)

En el manejo de la información estadística de los ODM que el área de Información realiza, se evidencian como deficiencias las siguientes: la información está almacenada en fuentes de datos que tienen formato *Excel*. Además, esta no presenta datos integrados en una estructura consistente, volviéndose engorrosa su centralización, manipulación e interacción. Anualmente, se generan ficheros no automatizados que se van acumulando con el paso del tiempo, lo que dificulta el proceso de

búsqueda y obtención de la información. Se carece de una aplicación informática que brinde reportes flexibles, lo que dificulta su análisis requiriéndose para ello un alto conocimiento del negocio. Todo lo anterior obstaculiza el manejo de la información, lo que trae como consecuencia frenar el análisis y estudio de la información, por consiguiente, demoras en el proceso de toma de decisiones para los órganos del Estado.

¿Cómo contribuir al proceso de toma de decisiones para las Series históricas objetivos cubanos del milenio en el área de Información del Sistema de Información de Gobierno?, constituye entonces el problema de la investigación a resolver en el presente trabajo, definiendo como objeto de estudio: los Almacenes de Datos, enmarcado en el campo de acción: Mercado de Datos Series históricas objetivos cubanos del milenio en el área de Información para el Sistema de Información de Gobierno.

Una vez planteado el problema, queda definido el objetivo general de esta investigación: desarrollar el Mercado de Datos Series históricas objetivos cubanos del milenio en el área de Información para el Sistema de Información de Gobierno a fin de contribuir a la toma de decisiones.

Derivado de lo anterior, se determinaron los siguientes objetivos específicos:

- Fundamentar la selección de las metodologías, herramientas y tecnologías a utilizar en el desarrollo del Mercado de Datos Series históricas objetivos cubanos del milenio en el área de Información para el Sistema de Información de Gobierno.
- Realizar el análisis y diseño del Mercado de Datos Series históricas objetivos cubanos del milenio en el área de Información para el Sistema de Información de Gobierno.
- Realizar implementación y pruebas del Mercado de Datos Series históricas objetivos cubanos del milenio en el área de Información para el Sistema de Información de Gobierno.

Para el logro de los objetivos específicos propuestos, se plantearon las siguientes tareas de investigación:

- Caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de Almacenes de Datos para definir cuáles serán seleccionadas durante la investigación.
- Levantamiento de requisitos para determinar las necesidades de información.
- Descripción de los casos de uso del Mercado de Datos para establecer cada una de las funcionalidades del sistema.
- Definición de los hechos, las medidas y las dimensiones del Mercado de Datos para determinar los elementos que conforman el modelo lógico de los datos.

- Diseño del modelo de datos para determinar los componentes que integran el modelo físico de los datos.
- Definición de la arquitectura del Mercado de Datos para identificar los subsistemas principales que la estructuran.
- Diseño del subsistema de integración como base para la implementación de sus procesos.
- Diseño del subsistema de visualización para permitir la presentación de la capa de visualización, el diseño de los cubos OLAP y los reportes candidatos.
- Diseño de los casos de pruebas como guía para aplicar las pruebas de sistema.
- Implementación del modelo de datos para definir las estructuras de la base de datos en el momento de ejecutar la carga del Mercado de Datos.
- Implementación del subsistema de integración para poblar el Mercado de Datos, cargando los hechos y las dimensiones correspondientes.
- Implementación del subsistema de visualización para visualizar los reportes que apoyan a la toma de decisiones.
- Aplicación de los casos de pruebas para validar que la aplicación cumple con los requisitos del cliente.
- Aplicación de las listas de chequeo a los artefactos concebidos en la etapa de Extracción, Transformación y Carga para determinar su correcta elaboración.

En el presente trabajo de diploma se utilizaron como técnicas de investigación: la revisión bibliográfica y documental del tema, la observación y entrevistas. Se debatieron los aspectos más cuestionables con profesionales expertos en la materia que se desempeñan actualmente como consultores, investigadores, directivos y especialistas.

La presente investigación se encuentra estructurada en tres capítulos, los cuales son:

Capítulo I: Fundamentos teóricos.

En el primer capítulo se abordan los aspectos teóricos generales que la bibliografía consultada ha proporcionado con relación a los fundamentos teóricos de los Almacenes de Datos. Se reflejan las principales definiciones relacionadas con el desarrollo de los Almacenes de Datos, las principales características del negocio, las herramientas y metodologías que se utilizarán.

Capítulo II: Análisis y diseño del Mercado de Datos Series históricas objetivos cubanos del milenio en el área de Información.

Se presenta una breve caracterización de la institución objeto de estudio. Se definen los requerimientos, casos de uso del sistema y se realizará el diseño de los subsistemas de integración y visualización del Mercado de Datos.

Capítulo III: Implementación y pruebas del Mercado de Datos Series históricas objetivos cubanos del milenio en el área de Información.

En el tercer y último capítulo, se implementa el modelo de datos, los procesos de Extracción, Transformación y Carga de los datos (ETL), así como los reportes candidatos. Se valida el Mercado de Datos a través de las diferentes pruebas aplicadas al software contenidas en el Modelo V, la utilización de las herramientas casos de pruebas y listas de chequeos.

## **CAPÍTULO 1: FUNDAMENTOS TEÓRICOS**

### **1.1. Introducción**

Desde mediados del siglo pasado, el desafío de la mayoría de las organizaciones ha sido convertir los datos en información, en una fuente de conocimientos para poder presentarlos de una forma integrada y consolidada. Por ello, han centrado su inversión en sistemas que logren automatizar los procesos de negocios. Como solución para obtener un sistema capaz de dar soporte a la toma de decisiones estratégicas y tácticas, nacieron las bases de datos multidimensionales, conocidas como Almacenes de Datos<sup>1</sup> (AD).

En el presente capítulo se exponen los aspectos esenciales sobre los AD, los Mercados de Datos<sup>2</sup> (MD), sus ventajas, desventajas, características y los componentes que los integran. Se plasma el resultado de las metodologías existentes y las principales herramientas para el desarrollo de estos.

### **1.2. Almacenes de Datos**

#### **1.2.1. Definición de los Almacenes de Datos**

Existen numerosas definiciones de los AD dadas por autores y estudiosos del tema. El término AD fue mencionado por primera vez por Bill Inmon, quien lo definió como: “una colección de datos orientados a temas, integrados, no volátiles y variantes en el tiempo, organizados para soportar necesidades empresariales”. (2) Ralph Kimball, otro conocido autor en este tema, lo define como: “una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis”. (3)

Un AD constituye un conjunto de datos históricos, internos o externos y descriptivos de un contexto o área de estudio, que se caracteriza por integrar y depurar de forma lógica y versátil la información tomada de una o más fuentes a fin de procesarla, permitiendo su análisis desde diferentes perspectivas y con grandes velocidades de respuesta con la cual se facilita la toma de decisiones en un negocio o empresa. Su creación representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implantar una solución completa y fiable de Inteligencia de Negocios. (4)

Es: “un conjunto de tecnologías de soporte a la toma de decisiones, cuyo objeto es que quien trabaja con los conocimientos (ejecutivo, directivo, analista), pueda tomar decisiones de manera más rápida y eficaz”. (5)

---

<sup>1</sup> En la terminología inglesa es conocido como *Data Warehouse*.

<sup>2</sup> En la terminología inglesa es conocido como *Data Mart*.

En resumen, se asume como AD un sistema para la colección de datos integrados y organizados con respecto a temas en específico, en los que la información va a prevalecer de una manera efectiva a lo largo del tiempo. Su objetivo se centra en satisfacer los requerimientos de información interna de una organización o empresa para una mejor gestión con respecto a la toma de decisiones.

### **1.2.2. Características de los Almacenes de Datos**

Según define el propio Bill Inmon, se caracterizan por ser:

- **Integrado**: la Base de Datos (BD) debe contener los datos de todos los sistemas operacionales de la organización, debiendo integrarse en una estructura consistente. La información debe estructurarse en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.
- **Temático**: los datos deben ser organizados por temas de manera que todos los elementos de datos relativos al mismo evento u objeto queden relacionados entre sí, a fin de facilitar su acceso y comprensión por parte de los usuarios finales.
- **Histórico**: el tiempo es parte implícita de la información contenida en un AD. Los datos se organizan y almacenan en jerarquías en el tiempo. Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones, de esta forma la información almacenada en el AD sirve para realizar análisis de tendencias, análisis comparativos de estados actuales y de períodos anteriores, a diferencia de los sistemas operacionales donde los datos siempre reflejan el estado de la actividad del negocio en el momento presente.
- **No volátil**: la información no se modifica ni se elimina, una vez almacenado un dato este se convierte en información de solo lectura y se mantiene para consultas futuras. La información es por tanto permanente, significando en la actualización del AD la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.
- **Contiene metadatos**: se refiere a datos sobre los datos. Los metadatos son datos altamente estructurados que describen información: el contenido, la calidad, la condición, entre otras características, permitiendo saber la procedencia de la información, su periodicidad de actualización, su fiabilidad, forma de cálculo, entre otras.

### **1.2.3. Ventajas y desventajas de los Almacenes de Datos**

Con la implementación de esta tecnología, la gestión empresarial se ha visto beneficiada con las siguientes ventajas:



- Garantiza una ventaja competitiva, pues permite a los responsables de la toma de decisiones el acceso a datos que pueden mostrar información que antes no estaba disponible, era desconocida o su acceso requería un tiempo de respuesta demasiado alto.
- Facilita la toma de decisiones estratégicas. Los AD integran datos de múltiples sistemas, que pueden incluso ser incompatibles y proporcionan una vista consistente de la organización, permitiendo a sus responsables realizar un análisis más preciso y estable de la información.
- Sirve para trabajar con cambios en la información producidos por las distintas transacciones que se realizan en un tiempo determinado (transacciones del día, mes o año, o sea: transacciones de tiempo real o transacciones *on-line*) a gran escala. El AD es capaz de actualizarse con las distintas bases de datos para poder almacenar esta nueva información, resumirla y así poderle dar las características que esta tecnología brinda.
- Facilidad e inmediatez en el manejo de la información, los usuarios finales acceden más fácil a una gran variedad de datos. El dinamismo de esta tecnología permite formar nuevas relaciones entre tablas, columnas y filas, que puede brindar información nueva al usuario.

No obstante, el uso de ella puede también implicar ciertas desventajas tales como:

- Altos gastos en mantenimiento a medida que transcurre el tiempo.
- Incremento continuo de los requisitos de los usuarios. (6)
- Requieren gran cantidad de espacio en disco.
- Proyectos de larga duración.

A pesar de estos inconvenientes, constituye una herramienta que permite administrar la información de manera correcta, accesible, uniforme y actualizada.

### **1.3. Mercados de Datos**

#### **1.3.1. Definición**

Como en una organización no todos los usuarios trabajan con la misma información y resultaría muy costoso y poco recomendable crear un AD para cada uno, una solución es enfocarse en una unidad de negocios y crear un MD a fin de satisfacer sus requisitos de información. Cada uno focaliza ciertas partes de la información del AD, lo que le da un carácter versátil a esta tecnología, puesto que posibilita trabajar con la información que se requiera. En definitiva, fue Kimball quien determinó que un AD no era más que “la unión de todos los MD de una entidad”. (7)

El diseño de los MD se realiza siguiendo una estructura óptima para analizar los procesos que tienen lugar dentro de un departamento o área, siendo orientado a temas específicos que contienen datos de

una sola línea del negocio. Es un sistema consultado mediante herramientas para el Proceso Analítico en Línea<sup>3</sup> (OLAP) que ofrecen una visión multidimensional de la información. Por lo tanto, se puede afirmar que los MD son un subconjunto específico de datos de un AD centrados en un tema o un área de negocio específico dentro de una organización.

### 1.3.2. Ventajas de los Mercados de Datos

Como ventajas de su construcción se pueden señalar:

- Más simples de implementar que un AD.
- Al ser pequeños conjuntos de datos, resultan más fáciles de manejar, permitiendo realizar consultas más rápidas al reducir la cantidad de datos a recorrer, por consiguiente, se satisfacen rápidamente las necesidades de las unidades de negocio.
- Permite separar los datos en diferentes plataformas de hardware.
- Logra satisfacer necesidades específicas de grupos de usuarios.
- Una organización puede contar con varios MD.
- Estructura los datos para su adecuado acceso por una herramienta.

### 1.4. Estructura de los Almacenes de Datos

Antes de describir la arquitectura del AD, es necesario tener presente la siguiente consideración ya generalizada y presente en la literatura: el término AD, se utiliza indistintamente para hablar tanto de la arquitectura en sí, como para uno de los componentes que la conforman, específicamente el que tiene relación con el almacenamiento físico de los datos. (8)

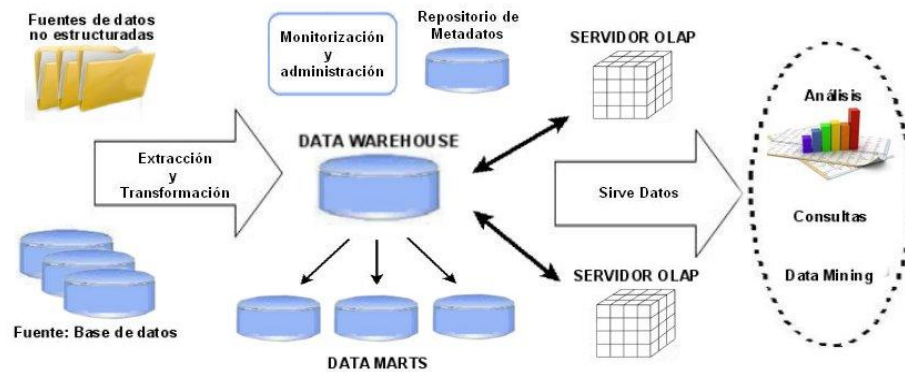
La arquitectura de un AD se constituye de un número de partes interconectadas, es una forma de representar la estructura total de datos, como comunicación, procesamiento y presentación, que existe para los usuarios finales que disponen de una computadora dentro de la empresa. (9) Incluyen:

- Datos operacionales: es referido al origen o fuente de datos para poblar el componente de almacenamiento físico del AD. El origen de los datos son los sistemas transaccionales internos de la organización como también datos externos a esta.
- Extracción de datos: es la elección sistemática de datos operacionales usados para el componente de almacenamiento físico AD.
- Transformación de datos: procesos para sintetizar, transformar y ejecutar cambios en los datos operacionales y para hacerlos coincidir con la estructura del AD.
- Carga de datos: inserción sistemática de datos en el componente de almacenamiento físico.

---

<sup>3</sup> En la terminología inglesa es conocido como *Online Analytical Processing*.

- Almacén de datos: almacenamiento físico de los datos.
- Herramientas de acceso al componente de almacenamiento físico AD: herramientas que proveen acceso a los datos.



**Figura 1. Arquitectura de los AD**

#### 1.4.1. Modelo de datos

La modelación puede definirse como una expresión de conceptos que permite a cada parte de la organización contribuir a su propio desarrollo, utilizando para ello métodos que describen sus diferentes partes. Su propósito se centra en proporcionar un registro exacto de determinados aspectos del mundo real en algún contexto particular, brindando al diseñador un instrumento que le permita comprender el comportamiento de los objetos modelados y predecir las consecuencias de cualquier acción o cambio en el ambiente.

Un modelo de bases de datos o modelo relacional es un conjunto de ideas lógicas utilizadas para representar la estructura de datos y las relaciones entre ellos dentro de la BD. (10) Los sistemas de bases de datos relacionales se sustentan en el modelo entidad-relación.

Para el diseño de un AD se emplea la representación de un modelo multidimensional, pudiendo ser considerado como una alternativa al modelo relacional. El modelo multidimensional es una técnica de diseño lógico que busca presentar los datos en un estándar que permita una recuperación adecuada de estos, modelar bases de datos simples y entendibles al usuario final. No está asociado a una representación física de los datos y puede ser soportado encima de un sistema gestor de BD relacional.

La estructura de almacenamiento de este modelo se encuentra diseñada como un cubo o hipercubo multidimensional, donde las aristas representan las dimensiones y las intersecciones entre estas

aristas son los llamados hechos o medidas. La medida o hecho es un dato numérico o fórmula que representa una actividad específica de un negocio, un patrón de interés o un evento dentro de una empresa que necesita ser analizado.

Existen tres clases de medidas:

- Medidas aditivas: pueden ser combinadas a lo largo de cualquier dimensión.
- Medidas semi-aditivas: no pueden ser combinadas a lo largo de una o más dimensiones.
- Medidas no aditivas: no pueden ser combinadas a lo largo de ninguna dimensión.

Las medidas se almacenan en las tablas de hechos y están implícitamente definidas por la combinación de los atributos (datos agregados) de las dimensiones. Una dimensión representa una perspectiva de los datos y es una descripción textual del negocio. Las dimensiones se almacenan en las tablas de dimensiones que son las “compañeras integrales” de las tablas de hechos. Las tablas de dimensiones constituyen las fuentes de casi todos los reportes y consultas que se le pueden realizar a los AD.

El modelo multidimensional tiene como ventajas:

- Garantiza la rapidez en la ejecución de las consultas.
- Se encuentra enfocado en el negocio y sus actividades.
- Permite la adición de dimensiones y hechos previstos, sin que por ello se deba volver a cargar los datos ya almacenados.
- Posibilita agregar nuevos atributos a las dimensiones. A medida que se añaden nuevos requerimientos al sistema, se pueden ir incorporando las modificaciones sin que esto implique demasiados cambios.

Existen diferentes esquemas que contribuyen a la representación lógica de un modelo multidimensional. En el AD pueden coexistir varios tipos de esquemas que pueden integrarse o no, según las necesidades de información del sistema:

- Esquema en estrella: representa una tabla de hechos central relacionada con varias tablas de dimensiones. Es el esquema más conocido y el resto son modificaciones de él. Este se utiliza con frecuencia en los MD, ya que está orientado precisamente a una actividad específica.
- Esquema copo de nieve: se da cuando alguna de las dimensiones se relaciona con una o más tablas de datos. Su objetivo es normalizar las tablas para reducir el espacio de almacenamiento al eliminar la redundancia de datos.

- Esquema constelación: está compuesto por una serie de esquemas de estrella. El esquema de constelación es el modelo que brinda mayor flexibilidad al diseñador de un sistema de información y a su vez, puede ser útil para dar respuesta a las necesidades de integración de aplicaciones, ya que su construcción se debe basar en el aprovechamiento adecuado de las características de cada tipo de modelo.

#### 1.4.2. Modos de almacenamiento

El Almacenamiento de Datos<sup>4</sup> constituye el proceso mediante el cual se facilita la integración de sistemas de aplicación no integrados. Organiza y almacena los datos que se necesitan para el procesamiento analítico e informático sobre una amplia perspectiva de tiempo. Soporta el procesamiento informático al proporcionar una plataforma sólida, a partir de los datos históricos para realizar el análisis.

Una de las técnicas más utilizadas que soporta las actividades de investigación y navegación del usuario final son los sistemas OLAP. Esta tecnología sustenta el estudio del comportamiento del negocio y su proyección. Se caracteriza por un análisis dimensional y dinámico, desde diferentes puntos de vista, de los datos consolidados de la empresa y ayuda al usuario a acelerar la consulta de grandes cantidades de datos y sintetizar la información de la empresa a través de vistas personalizadas, análisis históricos y pronósticos.

Un principio clave del OLAP es que los usuarios deben obtener tiempos de respuesta consistentes para cada vista de datos que requieran. Dado que la información se colecta en el nivel de detalle solamente, el resumen de la información es usualmente calculado por adelantado. Estos valores precalculados son la base de las ganancias de desempeño del OLAP.

Los sistemas OLAP se distinguen de los sistemas de Procesamiento Transaccional en Línea<sup>5</sup> (OLTP). Los primeros constituyen sistemas analíticos que facilitan la creatividad, se centran en la estrategia a largo plazo y la ventaja competitiva, son dirigidos por el negocio, siendo diseñados para dar soporte al procesamiento de consultas *ad hoc*<sup>6</sup>. Los segundos constituyen sistemas operacionales que automatizan el día a día los procedimientos del negocio y están dirigidos por la tecnología siendo diseñados por tanto, para maximizar la capacidad del procesamiento de transacciones. En la siguiente figura se comparan ambos sistemas.

---

<sup>4</sup> En la terminología inglesa es conocido como *Data Warehousing*.

<sup>5</sup> En la terminología inglesa es conocido como *Online Transaction Processing*.

<sup>6</sup> Consultas que permiten al usuario con poca experiencia en SQL tener el mismo acceso a la información de la base de datos.

Concepto	Sistemas Operacionales (OLTP)	Sistemas de Ayuda a la Toma de Decisiones (OLAP)
Datos	Valores actuales	Datos históricos y/o calculados
Organización	Por aplicación	Por áreas de la empresa
Acceso	Muy frecuente (lectura/escritura)	Baja frecuencia
Actualizaciones	Actualizaciones de campos	No se actualiza. Se manipula
Tiempo de respuesta	Medido por el tiempo de la transacción (del orden de segundos)	Medido por el orden de la consulta (del orden de minutos)
Tamaño de la BD	100 MB-GB	100 GB-TB
Usuarios	Miles	Cientos
Unidad de trabajo	Transacciones	Consultas complejas

**Figura 2. Comparación entre OLTP y OLAP**

Se puede definir varios tipos de OLAP en dependencia de las técnicas que se utilicen a la hora de obtener los datos y la forma en la que están estructurados:

- Proceso Analítico Relacional en Línea<sup>7</sup> (ROLAP): son herramientas OLAP que crean vistas multidimensionales extrayendo los datos de BD ordinarias, es decir, relacionales. Estas herramientas simulan los datos multidimensionales usando sofisticadas técnicas de indexación, cachés, metadatos, entre otras. Permiten manejar grandes cantidades de datos, aprovechar las funciones inherentes a la BD relacional, pero el rendimiento puede ser lento y se encuentran limitadas por las funcionalidades de Lenguajes Estructurados de Consultas<sup>8</sup> (SQL).
- Proceso Analítico Multidimensional en Línea<sup>9</sup> (MOLAP): son herramientas que acceden a datos que no están almacenados en registros de tablas, sino que almacenan los datos en arreglos de varias dimensiones, llamados cubos. Estos cubos utilizan índices para optimizar el acceso a los datos. Brindan un excelente rendimiento, permitiendo la ejecución de complicados cálculos. Sin embargo, requieren de una inversión adicional y se encuentran limitadas en la cantidad de datos que pueden manejar.
- Proceso Analítico Híbrido en Línea<sup>10</sup> (HOLAP): son herramientas que permiten un análisis híbrido de la información, ya que combinan los dos tipos anteriores a fin de brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. El análisis HOLAP ayuda a reducir costes de hardware, pues se requiere menos disco que en las bases de datos relacionales. Las respuestas de las consultas sobre las bases de datos multidimensionales son más rápidas que sobre las relacionales. Como aspecto negativo, los datos multidimensionales deben ser cargados antes de ser consultados y actualizados cuando se actualizan los datos de la organización.

<sup>7</sup> En la terminología inglesa es conocido como *Relational Online Analytical Process*.

<sup>8</sup> En la terminología inglesa es conocido como *Structured Query Language*.

<sup>9</sup> En la terminología inglesa es conocido como *Multidimensional Online Analytical Process*.

<sup>10</sup> En la terminología inglesa es conocido como *Hybrid Online Analytical Process*.

Existen discrepancias en el modo de almacenamiento de los cubos multidimensionales, principalmente en los tipos MOLAP y ROLAP. La arquitectura MOLAP tiene el mejor rendimiento de la tecnología relacional pero presenta problemas a la hora de adicionar dimensiones a un esquema ya existente. Se conoce que las capacidades OLAP se soportan mejor contra las bases de datos relacionales y la arquitectura ROLAP por su parte almacena los datos en una BD relacional, fue por tal razón que se decidió utilizar esta a diferencia de las otras explicadas. Además, es capaz de usar datos precalculados si estos se encuentran disponibles, accede directamente a los datos del AD y soporta técnicas de optimización de accesos para acelerar las consultas.

### 1.4.3. Integración de datos

Para poder explotar la información existente en los diferentes sistemas empresariales, hay que extraerla de ellos, transformarla y cargarla al AD.

El procesamiento de datos puede realizarse con tecnologías de Extracción, Transformación y Carga<sup>11</sup> (ETL) que constan de los siguientes procesos de integración: (11)

- Extracción: obtención de la información de las distintas fuentes tanto internas como externas, es decir, consiste en leer los datos de la BD original.
- Transformación: Es el proceso de filtrado, limpieza, depurado, homogenización y agrupación de la información.
- Carga: consiste en tomar los datos procesados, una vez terminada la operación de transformación, organizarlos y actualizarlos en el formato adecuado dentro del AD.

### 1.4.4. Inteligencia de Negocios

Actualmente una de las claves para que las organizaciones alcancen el éxito, es la capacidad para tomar decisiones de forma rápida y precisa. Los sistemas de información tradicionales como los sistemas de Planificación de Recursos Empresariales<sup>12</sup> (ERP), presentan por lo general una estructura inflexible para el logro de este fin, que obstaculiza obtener una debida información, así como extrapolar el conocimiento almacenado en el día a día de las bases de datos. (12) Por tal motivo, las entidades modernas suelen buscar en la Inteligencia de Negocios<sup>13</sup> (BI) la habilidad para transformar los datos en información, y la información en conocimiento, de forma tal que ayude en el proceso de toma de decisiones en los negocios. (13)

---

<sup>11</sup> En la terminología inglesa es conocido como *Extract, Transform and Load*.

<sup>12</sup> En la terminología inglesa es conocido como *Enterprise Resource Planning*.

<sup>13</sup> En la terminología inglesa es conocido como *Business Intelligence*.

BI surge para aprovechar toda la información operacional y convertirla en información estratégica. No constituye una metodología, software, sistema o herramienta específica, es más bien un conjunto de tecnologías que van desde arquitecturas para almacenar datos, metodologías, técnicas para analizar información y software entre otros, con un fin común para el apoyo a la toma de decisiones. (14)

BI es un término "agrupador", pues en él se pueden integrar funciones que tradicionalmente estaban separadas, tales como el acceso de datos, reportes, explotación, pronóstico y análisis. Este se ha convertido en un apoyo indispensable para la toma de decisiones en cualquier nivel de la organización y mucha gente está explotando el potencial estratégico de los datos operativos. Bien utilizada puede ser un arma estratégica de la gente de negocios, sustentada en tecnologías de sistemas. (15)

Para la colección de datos se usan o construyen AD y se manejan con técnicas de análisis y herramientas orientadas al usuario final. Las técnicas de análisis principales son: técnicas OLAP, Sistemas de Información Ejecutiva<sup>14</sup> (EIS), Sistemas de Soporte de Decisiones<sup>15</sup> (DSS), Minería de Datos<sup>16</sup> y Herramientas de Reportes, estas últimas a veces forman parte de las anteriores. (16)

### 1.5. Metodología de desarrollo

Se define como metodología a la ciencia que estudia los métodos del conocimiento. Se refiere a los métodos o procedimientos de investigación que se siguen para alcanzar una gama de objetivos en una ciencia. (14)

“Las metodologías de desarrollo de software son un conjunto de procedimientos, técnicas y ayudas a la documentación para el desarrollo de productos de software”. (17) Existen diversas metodologías para la construcción de un AD. Estas se pueden englobar dentro de dos grandes enfoques: descendente<sup>17</sup> y ascendente<sup>18</sup> que se corresponden con las propuestas por Bill Inmon y Ralph Kimball respectivamente. (16) La metodología propuesta por Inmon fue definida en el año 1992 en el libro “Building the Data Warehouse”. Mientras la propuesta por Kimball fue definida en el año 1998 con el libro: “The Data Warehouse Lifecycle Toolkit”. (18)

Inmon establece que el AD es una parte del todo que conforma un sistema de inteligencia. A la hora de diseñarlo, Inmon defiende un enfoque descendente: los datos son extraídos de los sistemas

---

<sup>14</sup> En la terminología inglesa es conocido como *Execute Information System*.

<sup>15</sup> En la terminología inglesa es conocido como *Decision Support Systems*.

<sup>16</sup> En la terminología inglesa es conocido como *Data Mining*.

<sup>17</sup> Tendencia conocida en la terminología inglesa como *Top-Down*.

<sup>18</sup> Tendencia conocida en la terminología inglesa como *Bottom-Up*.



operacionales por los procesos ETL y cargados en las áreas temporales<sup>19</sup>, existiendo además existen los llamados metadatos que documentan de una forma clara y precisa el contenido del mismo. Una vez realizado este proceso, los MD departamentales obtienen la información de él y con las consiguientes transformaciones, organizan los datos en las estructuras particulares requeridas por cada uno de ellos. (19) Es decir, los MD se crearán después de haber terminado el AD de la organización. Al tener este enfoque global, resultaría más difícil desarrollar un proyecto sencillo, pues se está intentando abordar el “todo”, a partir del cual luego se irá al “detalle”.

Kimball defiende un enfoque ascendente a la hora de diseñar el AD. Esta característica le hace más flexible y sencillo de implementar, pues permite construir un MD como primer elemento del sistema de análisis y luego ir añadiendo otros que comparten las dimensiones ya definidas o incluyen otras nuevas. En este sistema los procesos ETL extraen la información de los sistemas operacionales y los procesan igualmente en el área temporal, realizando posteriormente el llenado de cada uno de los MD de una forma individual, aunque siempre respetando la estandarización de las dimensiones.

La metodología propuesta por Kimball conlleva a una solución completa en una cantidad de tiempo relativamente pequeña. Los AD obtenidos son fácilmente consultables, tanto para los desarrolladores como para los usuarios finales. La relación directa entre los hechos y dimensiones confieren al usuario la posibilidad de construir consultas muy sencillas, la mayoría de las veces sin tener a mano la documentación de los metadatos. Es excelente para los primeros pasos de la implantación de BI a un cliente, cuando la complejidad del AD no es excesivamente grande y donde la infraestructura del BI se encarga de los datos procedentes de un número limitado de fuentes.

Se decidió utilizar en esta investigación el Modelo para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio, del Centro de Tecnologías y Gestión de Datos (DATEC) de la Universidad de las Ciencias Informáticas (UCI). Este modelo toma como base la metodología elaborada por Ralph Kimball, en la cual se crean los conceptos de hechos y dimensiones, que proporcionan mayor agilidad en el proceso de desarrollo y facilitan la toma de decisiones; propone ir construyendo el AD a través de la construcción de los MD departamentales, lo que coincide con la división lógica de las entidades; permite presentar al cliente los resultados parciales en cortos plazos, además de existir abundante documentación sobre esta metodología. (20)

---

<sup>19</sup> Espacio orientado a almacenar la información proveniente del sistema operacional que será el punto de partida de los procesos ETL en el AD. En la terminología inglesa es conocida como *stage*

Las principales desventajas que presenta la metodología propuesta por Kimball, que hacen que el modelo propuesto por DATEC solo la tome como guía en el proceso de su confección, son las siguientes: presenta un grupo de roles sin especificar las competencias y responsabilidades a desarrollar por cada uno dentro del proyecto, plantea un gran número de actividades y artefactos que pueden extender los tiempos de desarrollo sobre todo si se cuenta con pocos recursos humanos, se encuentra estructurada para el desarrollo de proyectos-productos donde un proyecto desarrolla un producto determinado y no establece el análisis de diferentes criterios de diseño en el levantamiento de requisitos que permita la construcción más adecuada del AD. (21)

La nueva metodología se ajusta a las condiciones y características de producción de DATEC y de la UCI. Como complemento a la misma y fortaleciendo la etapa del levantamiento de requisitos, se tomó lo planteado por Leopoldo Zenaido Zepeda Sánchez en su tesis de doctorado, orientando el trabajo a los Casos de Uso. (21)

### **1.5.1. Grupos y flujos de trabajo**

El modelo propuesto establece cinco grupos de trabajo:

- Grupo de Análisis
- Grupo de Almacenes
- Grupo de ETL
- Grupo de BI
- Grupo de Dirección

Dentro de un proyecto, cada uno de estos grupos ejecuta tareas bien determinadas y delimitadas de acuerdo a sus responsabilidades.

El proceso de desarrollo de este modelo se integra por 7 flujos de trabajos:

- Estudio preliminar o planeación: se efectúa el diagnóstico del usuario final, la planeación del proyecto, se definen los objetivos, el alcance preliminar, los costos estimados, entre otras actividades.
- Requerimientos: se lleva en dos direcciones, la primera mediante la identificación de las necesidades de información y reglas del negocio; y la segunda, con un levantamiento detallado de las fuentes de datos a integrar. Posteriormente a esto, se definen los requerimientos.
- Arquitectura y diseño: se determinan las estructuras de almacenamiento, se diseñan las reglas de ETL, se establece la arquitectura de información que administrará el desarrollo de la solución.

- **Implementación:** se diseña físicamente el repositorio de datos, se elaboran las estructuras y el área temporal de almacenamiento, se ejecutan las reglas de ETL y se configuran e implementan las herramientas de BI para la obtención de los elementos que se acordaron con el cliente final.
- **Prueba:** en este flujo se llevan a cabo las pruebas al sistema desde las pruebas de unidad hasta las de aceptación con el usuario final.
- **Despliegue:** se realiza un despliegue piloto donde se configuran los servidores y se instalan las herramientas, cargándose una muestra de los datos a fin de demostrar que el sistema funciona. Posterior a la aceptación del usuario final, se ejecuta la carga de los datos, la capacitación y la transferencia tecnológica.
- **Soporte y mantenimiento:** luego de la implantación de la solución, se brindan los servicios de soporte en línea, vía telefónica, *web* u otras según se haya acordado con el usuario y las condiciones de soporte que se hayan establecido.

Grupos/Flujos	Estudio Preliminar	Requerimientos	Arquitectura y Diseño	Implementación	Prueba	Despliegue	Soporte y Mantenimiento
Análisis	Responsable	Responsable	Participa	No Participa	Responsable	No Participa	No Participa
Almacén	Participa	No Participa	Responsable	Responsable	Participa	Responsable	Participa
ETL	Participa	No Participa	Responsable	Responsable	Participa	Responsable	Participa
BI	Participa	No Participa	Responsable	Responsable	Participa	Responsable	Participa
Dirección	Responsable	Responsable	Responsable	Responsable	Responsable	Responsable	Participa

Leyenda:

- Responsable
- Participa
- No Participa

**Figura 3. Participación y responsabilidad de cada grupo de trabajo en los diferentes flujos de trabajo**

## 1.6. Herramientas

Con el objetivo de ofrecer un acercamiento a la automatización del diseño, construcción, implementación y mantenimiento del MD Series históricas objetivos cubanos del milenio en el área de Información, se disponen de diferentes herramientas.

### 1.6.1. Herramienta de modelado

Las herramientas de Ingeniería de Software Asistida por Computadora<sup>20</sup> (CASE), son un conjunto de programas y ayudas para el desarrollo de programas informáticos que son empleadas por los

<sup>20</sup> En la terminología inglesa es conocido como *Computer Aided Software Engineering*.

diseñadores y analistas durante todas las fases del ciclo de vida del desarrollo de un software, proveyéndolos de métodos y técnicas a fin de garantizar el desarrollo del software.

Visual Paradigm 8.4: es la herramienta profesional de Lenguaje de Modelado Unificado<sup>21</sup> (UML) que se utilizará en la construcción del AD para la entidad objeto de estudio por las siguientes razones:

- Es fácil de instalar y actualizar.
- Soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue.
- Permite modelar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación.
- Permite la exportación de *script* de manera íntegra.
- Genera código para un gran número de lenguajes de programación entre los que se encuentra Java y permite integración con varias herramientas de Java.
- Generación de BD. Transformación de diagramas de entidad-relación en tablas de BD.
- Ingeniería inversa: código a modelo, código a diagrama, desde sistemas gestores de bases de datos (DBMS) existentes a diagramas de entidad-relación.
- Generador de informes para generación de documentación.
- Existe compatibilidad entre ediciones.
- Ayuda a una más rápida construcción de aplicaciones de calidad a un menor coste.
- Soporta aplicaciones *web*.
- Editor de detalles de casos de uso, entorno todo en uno para la especificación de los detalles de los casos de uso, incluyendo la especificación del modelo general y de las descripciones de los casos de usos.
- Dispone de múltiples versiones dependiendo de la necesidad.
- Disponibilidad en múltiples plataformas.
- Proporciona un entorno amigable e interactivo al usuario.

### **1.6.2. Sistema Gestor de Bases de Datos**

Un Sistema Gestor de Bases de Datos (SGBD) es un conjunto de programas que se encargan de manejar la creación y todos los accesos a las bases de datos. Se compone de un lenguaje de definición de datos, de un lenguaje de manipulación de datos y de un lenguaje de consulta. El objetivo primordial de un SGBD es proporcionar un entorno que sea a la vez conveniente para ser utilizado al

---

<sup>21</sup>

En la terminología inglesa es conocido como *Unified Modeling Language*.

extraer y almacenar información de la BD. (22) Estos programas se encargan de la integridad y la seguridad de los datos, garantizando la interacción con el sistema operativo.

PostgreSQL 9.1: es el SGBD a utilizar en el desarrollo del MD, el cual es un potente Sistema de Base de Datos Relacional libre para grandes bases de datos. Su código fuente se encuentra disponible (*Open Source*), lo que garantiza independencia tecnológica. Emplea un modelo cliente/servidor y usa multiprocesos que proporcionan la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto y le permite al sistema seguir funcionando.

Este SGBD proporciona ventajas dadas por las características que lo definen:

- Funciona en la mayoría de los sistemas operativos actuales.
- Soporta casi toda la sintaxis SQL.
- Posee puntos de recuperación a un momento dado, *tablespaces*, replicación asincrónica, transacciones jerarquizadas y copia de seguridad en línea.
- Posibilita una alta concurrencia de usuarios accediendo a la vez al sistema, permitiendo que mientras unos escriban en una tabla, otros accedan a la misma tabla sin necesidad de bloqueos.
- Aproxima los datos a un modelo objeto-relacional y es capaz de manejar complejas rutinas y reglas.
- Soporta operadores, funciones, métodos de acceso y tipos de datos definidos por el usuario.
- Incluye características avanzadas tales como las uniones (*joins*) SQL92.
- Soporta integridad referencial, la cual es utilizada para garantizar la validez de los datos de la BD. El usuario ve con transparencia los detalles del almacenamiento físico de los datos mediante varios niveles de abstracción de la información.
- Presenta un complejo sistema que maneja grupos, usuarios y permisos para las diferentes actividades que se pueden realizar dentro del mismo, proveyendo de esta manera al usuario la seguridad que sus datos no podrán ser accedidos, ni manipulados por quien no tenga permiso para ello. Permite además, realizar copias de seguridad de la información almacenada y a partir de estas, restaurar los datos.
- Funciona con grandes cantidades de datos, manteniendo la integridad de estos, así como el manejo de grandes volúmenes de información.
- Hace más sencillo el análisis de datos.
- Es diseñado para ambientes de alto volumen.
- Herramientas gráficas de diseño y administración de BD.

### 1.6.3. Administrador de bases de datos

Para la administración de bases de datos se puede disponer como herramienta: PgAdmin.

PgAdmin III, 1.14.1: es una herramienta de código abierto para la administración de bases de datos PostgreSQL y derivados (EnterpriseDB Postgres Plus Advanced Server y Greenplum Database). Presenta las siguientes características que la hacen una de las más populares para administrar las bases de datos en PostgreSQL:

- Ha sido diseñada a fin de responder a las necesidades de los usuarios, desde escribir simples consultas SQL hasta desarrollar bases de datos complejas.
- La interfaz administrativa gráfica soporta todas las características de PostgreSQL y hace simple la administración.
- Es un software libre, encontrándose disponible en más de una docena de lenguajes y para varios sistemas operativos.
- Editor de código procedural.

### 1.6.4. Herramienta para la limpieza de los datos

- DataCleaner 1.5.4: es una aplicación de código abierto para el perfilado, la validación y comparación de datos que ayuda a la administración y supervisión de la calidad de los datos con el objetivo de garantizar que la información sea útil y aplicable a la situación del negocio. El DataCleaner puede acceder y analizar prácticamente cualquier AD. Es un software fácil de utilizar, genera sofisticados informes y gráficos que les permite a los usuarios comprobar la calidad de los datos existentes, la investigación estadística, la preparación para el proceso de ETL y otras actividades.

### 1.6.5. Herramienta para la integración de los datos

- Pentaho Data Integration 4.2.1: es una herramienta de código abierto y fácil lectura para los XML que recogen transformaciones, tareas programadas y un repositorio relacional de metadatos ETL. Puede funcionar sobre varias plataformas a través de un sistema que soporte Java 1.4 Runtime Environment o una versión superior. Exige solamente alrededor de 128 MB de RAM. Provee un JDBC<sup>22</sup> que permite la conexión con cualquier BD sin tener que instalar un cliente adicional. Se integra con ficheros de Microsoft Office, *web services* y cubos MOLAP. Es aplicable a diversos tipos de bases de datos (SQL server, PostgreSQL, MySQL, Microsoft Access, entre otros). Posee facilidad para la importación y exportación de datos de un formato

---

<sup>22</sup> En la terminología inglesa es conocido como *Java Database Connectivity*, permite la ejecución de operaciones sobre bases de datos desde el lenguaje Java.

a otro cualquiera. Su principal fortaleza es la posibilidad que brinda de ser extensible mediante *pluggins*.

#### 1.6.6. Herramientas de Inteligencia de Negocio

- Pentaho Schema Workbench 3.2.1: es un entorno visual para el desarrollo y prueba de cubos OLAP Mondrian. Provee un mecanismo para buscar datos con rapidez y tiempo de respuesta uniforme independientemente de la cantidad de datos en el cubo o la complejidad del procedimiento de búsqueda. Permite la ejecución de consultas de Expresiones Multidimensionales de Microsoft<sup>23</sup> (MDX) contra el esquema y la BD, así como la navegación por la BD subyacente. (23) Entre sus características se destacan:
  - Permite crear todos los objetos que soporta Mondrian: esquema, cubo, dimensiones, métricas.
  - Tiene dos áreas: la zona en la que se muestra la estructura jerárquica del esquema OLAP y la zona de edición de las propiedades de cada elemento.
  - Realiza consultas MDX contra el esquema creado (requiere conocer la sintaxis del lenguaje).
  - Consulta la BD que sirve de origen para el esquema de Mondrian.
  - Publica directamente el esquema en el servidor de Pentaho. (24)
- Pentaho Report Designer 3.6.1: es una herramienta independiente que forma parte de la unidad de reportes de Pentaho (Pentaho Reporting). Este diseñador de reportes ofrece un entorno gráfico familiar, con herramientas intuitivas, fáciles de utilizar, una estructura de reporte bastante acertada y flexible para darle libertad al diseñador de generar reportes que se adapten totalmente a su necesidad. Por otra parte, simplifica el proceso de generación de reportes, posibilitando la creación de informes sofisticados. Incluye formatos de salida como: *Adobe PDF*, *HTML*, *Microsoft Excel*. Finalmente, publica directamente los reportes diseñados en el servidor de Pentaho.
- Apache Tomcat 6.6: es una herramienta utilizada como servidor *web* autónomo en entornos con alto nivel de tráfico y alta disponibilidad. Presenta como características:
  - Los usuarios disponen de libre acceso a su código fuente.
  - Está integrado en la implementación de referencia *Java 2 Enterprise Edition* (J2EE) de *Sun Microsystems*.

---

<sup>23</sup>

En la terminología inglesa es conocido como *Microsoft's Multidimensional Expressions*.

- Funciona en cualquier sistema operativo que disponga de la máquina virtual Java.
  - Es una implementación completamente funcional de los estándares de JSP y *servlets*.
  - Puede especificarse como el manejador de las peticiones de JSP y *servlets* recibidas por servidores *web* populares.
- Pentaho BI Server 3.8.0: funciona como una red basada en sistemas de gestión de informe, el servidor de integración de aplicaciones y un motor de flujo de trabajo ligero (secuencias de acción). Está diseñado para integrarse fácilmente en cualquier proceso de negocio. (25)
  - Mondrian OLAP Server 3.0.4: es un servidor OLAP de código abierto que gestiona la comunicación entre la aplicación OLAP y la BD. Se ejecuta sobre un servidor *web*. Permite crear cubos de información para análisis multidimensional. Es desarrollado en *Java/Servlets/JSPs* que posibilita ser instalado en servidores de aplicaciones como *JBoss*. Entre sus principales características se encuentran:
    - Facilidad para el análisis de grandes volúmenes de información que se encuentren almacenados en bases de datos que soporten JDBC.
    - Soporta el lenguaje MDX.
    - Soporta los APIs: Java OLAP (JOLAP) y *XML for Analysis application Programming*.

## 1.7. Conclusiones

En este capítulo fueron expuestos los conceptos teóricos básicos del proceso de desarrollo de un AD. A través de la revisión bibliográfica y documental del tema, se logró fundamentar las metodologías, tecnologías y herramientas que se utilizarán en la construcción del MD objeto de estudio. Todo lo anterior permitió determinar como conclusiones:

- Fue definida como metodología el Modelo para el Desarrollo de Soluciones de AD y BI aplicada en DATEC, debido a sus ventajas que permitirán ganar en rapidez, organización, calidad y obtener resultados parciales para satisfacer al cliente hasta lograr la solución final. Además, esta se ajusta a las condiciones y características de producción de este centro y de la UCI, tomando como guía la metodología de Kimball.
- Como herramienta profesional UML se seleccionó el Visual Paradigm 8.4, ya que posibilitará modelar todos los tipos de diagramas necesarios para la aplicación a desarrollar.
- Como SGBD relacional se utilizará el PostgreSQL 9.1 y como herramienta para la administración de los datos el PgAdmin III 1.14.1, pues proporcionarán el entorno conveniente para la extracción y almacenamiento de la información de la BD, así como garantizarán la integridad y seguridad de los datos.



- Las herramientas seleccionadas DataCleaner 1.5.4 y Pentaho Data Integration 4.2.1, ayudarán a la supervisión de la calidad de los datos que serán cargados al MD con el fin de garantizar una información útil y aplicable.
- Las herramientas elegidas Pentaho Schema Workbench 3.2.1, Pentaho Report Designer 3.6.1, Apache Tomcat 6.6, Pentaho BI Server 3.8.0 y Mondrian OLAP Server 3.0.4, garantizarán una correcta implementación del subsistema de visualización de los datos para la solución final.

## **CAPÍTULO 2: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS SERIES HISTÓRICAS OBJETIVOS CUBANOS DEL MILENIO EN EL ÁREA DE INFORMACIÓN**

### **2.1. Introducción**

En este capítulo para una mayor comprensión del negocio, se detallan las principales características de las Series históricas objetivos cubanos del milenio en el área de Información. Se abordan los aspectos esenciales del análisis y el diseño de la solución, específicamente se definen los requisitos funcionales, no funcionales y de información, los Casos de Uso del Sistema (CUS) y los reportes candidatos. Se determinan las dimensiones del MD, los hechos y las medidas asociadas a estos. Se confecciona la matriz bus y el modelo de datos. Por último, se describe la arquitectura del Mercado de Datos.

### **2.2. Análisis**

#### **2.2.1. Descripción del negocio**

La ONEI es el órgano rector de las estadísticas en Cuba, es la encargada de gestionar y guardar toda la información estadística nacional. Tiene como misión: garantizar la producción de estadísticas de calidad a través del Sistema Nacional ejerciendo una adecuada dirección, ejecución y control de la captación de las cifras económicas y sociales, así como su adecuada difusión de acuerdo al comportamiento de la economía y las demás necesidades del país en información estadística. Como visión, se fija construir un sistema estadístico profesional capaz de responder con calidad y oportunidad las necesidades de información estadística para enfrentar las metas del desarrollo económico, social y su adecuado reflejo intersocial, conscientes de ser útiles y necesarios a la sociedad.

Esta organización también ofrece un servicio estadístico para evaluar el estado de cumplimiento de los Objetivos de Desarrollo del Milenio (ODM) del país. La recolección de las series de datos estadísticos se ejecuta con una periodicidad anual, las cuales son conocidas como series históricas.

El objetivo de estas series históricas se centra en divulgar la información actualizada de Cuba desde el año 1990 hasta el 2010. Las mismas recogen los ocho ODM identificados en la Asamblea General de las Naciones Unidas en septiembre del 2000, en la cual se aprobó la Declaración del Milenio. Esta declaración aborda integralmente y con un mayor alcance los acuerdos suscritos en las cumbres mundiales de las Naciones Unidas celebradas en la década de los años noventa. En esta se acordó además, el nivel que debían alcanzar para el 2015 con relación al 1990, la revisión periódica del avance de los mismos, identificándose para ello sus indicadores. Estos objetivos identificados apuntan

a la lucha contra la pobreza y el hambre, la reversión del deterioro ambiental, el mejoramiento de la educación y la salud, la promoción de la igualdad entre los sexos y la cooperación internacional.

Los objetivos son:

1. Erradicar la pobreza extrema y el hambre.
2. Lograr la enseñanza primaria universal.
3. Promover la igualdad entre los géneros y el empoderamiento de la mujer.
4. Reducir la mortalidad de los niños menores de 5 años.
5. Mejorar la salud materna.
6. Combatir el VIH/SIDA, el paludismo y otras enfermedades.
7. Garantizar la sostenibilidad del medio ambiente.
8. Fomentar una alianza mundial para el desarrollo.

Los ODM se encuentran divididos en 16 metas y presentan 135 indicadores en total. Dispersos en las distintas metas consideradas en Cuba, existen 78 indicadores adicionales junto a los 41 indicadores oficiales utilizados en el país. Se incluyen también 16 indicadores de apoyo, que si bien no pueden consignarse en una meta determinada, contribuyen a la evaluación de los postulados que consignan determinados objetivos. La información estadística que aparece abarca los datos a nivel de Cuba y siempre que ha sido posible, se ha procurado información de los años 1990, 1995, 2000 y del 2005 al 2010.

### **2.2.2. Flujo de información**

La fuente fundamental de los datos procede del Sistema de Información Estadístico Nacional (SIEN). Para la mayoría de los indicadores, la ONEI realizó un trabajo de captación de datos y elaboración de las fichas técnicas en los ministerios de Salud Pública (MINSAP), Educación (MINED), Educación Superior (MES), del Trabajo y Seguridad Social (MTSS), del Interior (MININT), de la Agricultura (MINAG), de la Informática y las Comunicaciones (MIC), de Relaciones Exteriores (MINREX), de Ciencia Tecnología y Medio Ambiente (CITMA), así como en el Instituto Nacional de Recursos Hidráulicos (INRH), Instituto Cubano de Radio y Televisión (ICRT), Tribunal Supremo Popular y Fiscalía General de la República, entre otros. Toda la información procedente de estos organismos, es recopilada de los modelos que forman parte del Sistema de Información Estadística Complementaria (SIEC).

En el Centro de Estudios de Población y Desarrollo (CEPDE) correspondiente a la entidad objeto de estudio, han sido desarrollado los objetivos no. 1, 2, 4, 5 y 6; el primero, por el área de Desarrollo

Social y los otros, por el Departamento de Estadísticas Sociales. En cuanto a los restantes objetivos, los no. 3 y 8 son desarrollados por la propia Dirección de Información, mientras que el no. 7 por el Grupo de Medio Ambiente perteneciente a la Dirección de Industrias y Medio Ambiente. Por último, la Dirección de Información es la responsabilizada en recopilar, procesar y elaborar toda esta información de forma homogénea obteniendo como resultado final las series estadísticas, donde incorpora además la procesada por esta área.

### **2.2.3. Necesidades de información**

Después de haber realizado un estudio de las Series históricas objetivos cubanos del milenio en el área de Información, se pudo determinar que los especialistas que trabajan con ellas, se enfocan en el análisis de cómo están estructurados todos los indicadores que describen cada uno de los ODM, qué nivel deben presentar por año y si existe un avance de estos con respecto a los de años anteriores. Por lo anteriormente explicado, se definió como necesidad de información el análisis y la difusión de los indicadores de los objetivos cubanos del milenio.

### **2.2.4. Especificación de requerimientos**

La especificación de requerimientos es una descripción completa del comportamiento del sistema que se va a desarrollar para las Series históricas objetivos cubanos del milenio en el área de Información. En estas, se describen los Requisitos de Información (RI) agrupados por las necesidades de información del cliente, los Requisitos Funcionales (RF) orientados a las funcionalidades del sistema y los Requisitos no Funcionales (RNF) que constituyen cualidades que el sistema debe cumplir.

#### Requisitos de información:

Para poder obtener los RI de las series históricas, se tuvo en cuenta que estos deben describir qué información debe ser almacenada en el sistema y así satisfacer la necesidad de información del cliente. Se definieron a partir de los elementos que se encontraban disponibles en las fuentes de datos. También se tuvo en cuenta que los mismos deben constituir la entrada fundamental para todos los procesos de BI y sus futuros reportes. Conforme con la investigación del negocio y la necesidad de información anteriormente descrita, se definieron como RI:

- RI1: Obtener el valor de indicador de los objetivos cubanos del milenio por indicador y por año.
- RI2: Obtener las notas técnicas de todos los indicadores de los objetivos cubanos del milenio.

#### Requisitos funcionales:

El sistema que se va a implementar para el MD debe contar con ciertas funcionalidades, capacidades o condiciones que satisfagan las necesidades o expectativas de los especialistas. Este debe permitir al

cliente que pueda consultar los distintos reportes que visualizarán todos los indicadores de los objetivos cubanos del milenio. Debe además permitir que se puedan autenticar varios usuarios, administrar los mismos y los diferentes roles que se les asignarán a cada uno de ellos. Para el desarrollo del MD se definieron los siguientes RF:

- RF1: Autenticar usuario.
- RF2: Adicionar roles.
- RF3: Eliminar roles.
- RF4: Adicionar usuarios.
- RF5: Eliminar usuarios.
- RF6: Insertar reportes.
- RF7: Modificar reportes.
- RF8: Eliminar reportes.
- RF9: Extraer información.
- RF10: Realizar transformación y carga.
- RF11: Abrir navegador OLAP.
- RF12: Mostrar editor MDX.
- RF13: Mostrar Padres.
- RF14: Ocultar repeticiones.
- RF15: Intercambiar ejes.
- RF16: Configurar impresión.
- RF17: Exportar a PDF.
- RF18: Exportar a Excel.
- RF19: Suprimir filas.
- RF20: Detallar miembros.
- RF21: Entrar en detalles.
- RF22: Mostrar datos de origen.

#### Requisitos no funcionales:

El sistema, además de presentar funcionalidades que satisfagan las necesidades del cliente, debe tener propiedades, cualidades o características que lo harán usable, rápido, confiable y seguro. En el análisis del negocio se identificaron 16 RNF, los cuales se clasificaron en requisitos de usabilidad, de soporte, de restricciones del diseño, para la documentación de usuarios en línea y ayuda del sistema, de interfaz, legales, de derecho de autor y otros.

- Usabilidad:
  - RNF1: Cumplir con las pautas de diseño de las interfaces.
  - RNF2: Mostrar los mensajes, títulos y demás textos que aparezcan en la interfaz del sistema en idioma español.
  - RNF3: Agilizar el acceso a los reportes del almacén de datos mediante la distribución de la información por áreas de análisis.
- Confiabilidad:
  - RNF4: Garantizar la persistencia de la información.
- Soporte:
  - RNF5: Lograr la homogeneidad de la estructura de los elementos definidos en el almacén.
- Restricciones de diseño:
  - RNF6: Utilizar los lenguajes de programación definidos durante la investigación.
  - RNF7: Utilizar el Sistema Gestor de Base de Datos definido durante la investigación.
  - RNF8: Utilizar la herramienta de integración de datos definida durante la investigación.
  - RNF9: Utilizar las herramientas para la implementación de la capa de inteligencia de negocios definidas durante la investigación.
- Requisitos para la documentación de usuarios en línea y ayuda del sistema:
  - RNF10: Confección de un manual de usuario.
- Interfaz:
  - RNF11: Acceso al sistema.
- Interfaces de usuario:
  - RNF12: Garantizar una interfaz amigable al usuario.
- Interfaces de hardware:
  - RNF13: Definir las interfaces de hardware que soportará el sistema.
- Interfaces de software:
  - RNF14: Instalar en las estaciones de trabajo el software necesario para el correcto funcionamiento del sistema.
- Requisitos legales, de derecho de autor y otros:
  - RNF15: Entregar el sistema a la ONEI.
  - RNF16: Requerimientos legales, de derecho de autor y otros.

### **2.2.5. Modelo de Casos de Uso del Sistema**

Con el objetivo de que los especialistas puedan consultar la información de los indicadores de los objetivos cubanos del milenio y una vez realizado el levantamiento de requisitos, se definen cuáles

serán los Casos de Uso del Sistema (CUS). Estos últimos constituyen las acciones que el sistema debe permitir, la interacción del usuario con este para obtener un objetivo específico.

Para la realización del MD fue necesario identificar los CUS que se dividieron en Casos de Uso de Información (CUI) y en Casos de Uso Funcionales (CUF). Los CUI se agruparon por el tipo de información que se maneja en las Series históricas objetivos cubanos del milenio en el área de Información y en dependencia de las necesidades de información de los usuarios, por lo que fue identificado un solo CUI. La realización de los CUF está basada en las operaciones ETL que se le realizarán a los datos fuentes, en la administración de usuarios, roles y reportes. En resumen, se determinaron siete CUF. En la siguiente tabla se detallan los CUS identificados y una breve descripción de estos:

<b>CUI</b>	<b>Descripción</b>
Presentar información de los indicadores de los objetivos cubanos del milenio.	Permite a los usuarios visualizar los valores de los indicadores de los objetivos cubanos del milenio desde el año 1990 hasta el 2010.
<b>CUF</b>	<b>Descripción</b>
Autenticar usuario.	Permite a los usuarios, mediante una contraseña que les proporcionará los permisos relacionados a ellos, poder interactuar con la aplicación.
Administrar usuario.	Da la posibilidad de adicionar un nuevo usuario o eliminar uno ya existente en el sistema.
Administrar rol.	Da la posibilidad de añadir un nuevo rol o eliminar uno ya existente en el sistema.
Administrar reporte.	Da la posibilidad de insertar, modificar o eliminar un reporte en el sistema.
Visualizar reporte.	Permite visualizar los reportes existentes en el sistema, así como las opciones que se presentan para que los usuarios puedan modificarlos si lo desean.
Realizar la extracción de los datos.	Permite realizar la extracción de la información contenida en las fuentes de información para ser cargadas al MD.
Realizar la transformación y carga de los datos.	Permite transformar la información que fue extraída y finalmente la carga en el MD.

**Tabla 1: CUS y su descripción**

Actores del sistema:

Para poder identificar los actores del sistema se tuvo en cuenta que estos pueden ser las personas o programas que interactuarán con este, por lo que se procede a describir seguidamente:

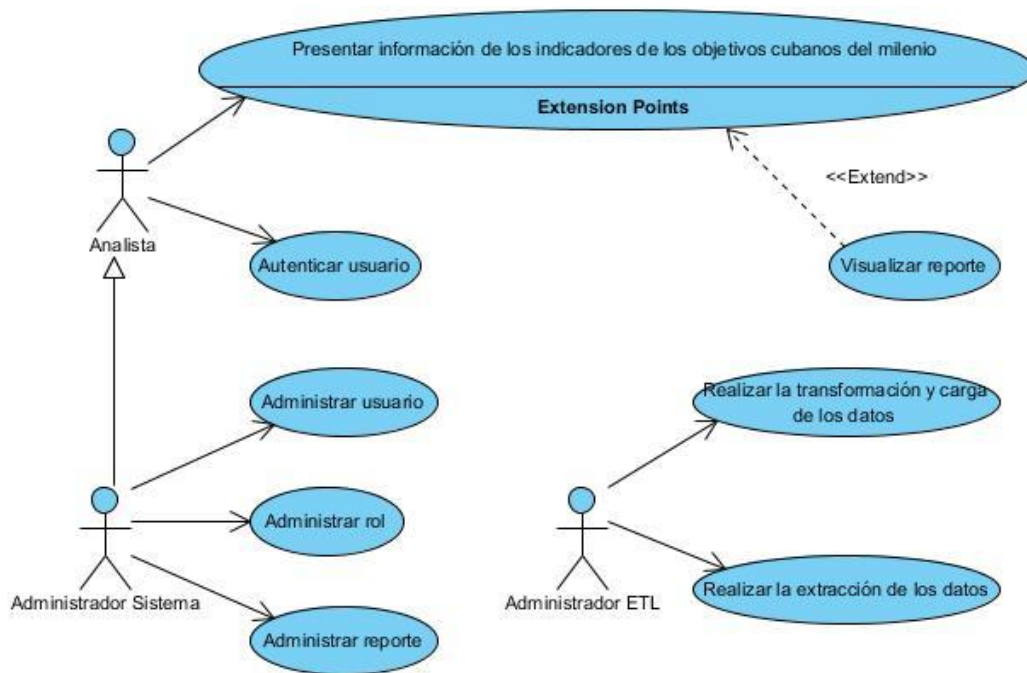
<b>Actor</b>	<b>Objetivo</b>
--------------	-----------------

Analista	Es el encargado de analizar los datos.
Administrador del sistema	Administra los roles, los usuarios y los reportes; además, de analizar los datos contenidos en los reportes.
Administrador ETL	Realiza los procesos ETL.

**Tabla 2. Actores del sistema y su descripción**

Diagrama de CUS:

Después de haber identificado los CUS y los actores del sistema, se debe hacer una representación gráfica de la relación existente entre ellos, para poder especificar la comunicación y el comportamiento del sistema que se va a desarrollar, al mostrar cómo reacciona a eventos que se producen en su ámbito o en él. Esta representación gráfica también es conocida como diagrama de CUS y a continuación se muestra el diseñado para el MD:



**Figura 4. Diagrama de CUS**

**2.2.6. Reportes candidatos**

De acuerdo con el pedido de información identificado en el análisis del negocio, se identificaron dieciséis reportes candidatos para cada uno de los ODM. Estos reportes son:

1. Valor de indicador por indicadores de la pobreza extrema y hambre.
2. Valor de indicador por indicadores de la enseñanza primaria universal.
3. Valor de indicador por indicadores de la igualdad de género y el empoderamiento de la mujer.
4. Valor de indicador por indicadores de la mortalidad de los niños menores de 5 años.



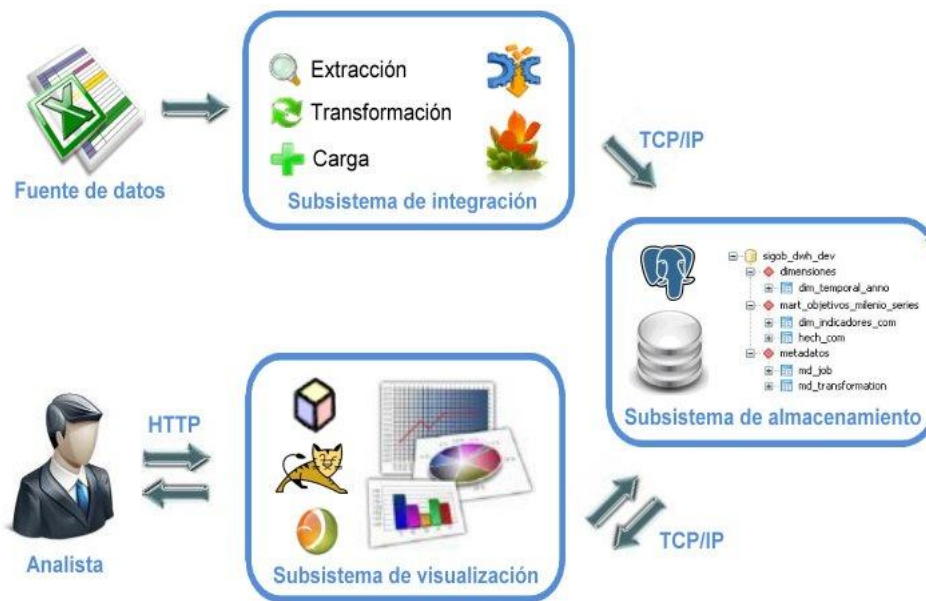
5. Valor de indicador por indicadores de la salud materna.
6. Valor de indicador por indicadores del VIH/SIDA, el paludismo y otras enfermedades.
7. Valor de indicador por indicadores del medio ambiente.
8. Valor de indicador por indicadores de la alianza mundial.
9. Notas Técnicas del Objetivo 1.
10. Notas Técnicas del Objetivo 2.
11. Notas Técnicas del Objetivo 3.
12. Notas Técnicas del Objetivo 4.
13. Notas Técnicas del Objetivo 5.
14. Notas Técnicas del Objetivo 6.
15. Notas Técnicas del Objetivo 7.
16. Notas Técnicas del Objetivo 8.

### **2.3. Arquitectura de los subsistemas**

El MD Series históricas objetivos cubanos del milenio en el área de Información consta de tres subsistemas: el de integración, de almacenamiento y visualización. En el subsistema de integración es donde se realizan todos los procesos ETL en los cuales se extrae, se limpia e integra toda la información almacenada a través de transformaciones y trabajos. Esta integración es llevada a cabo mediante la herramienta Pentaho Data Integration, debido a que es muy segura para este tipo de procesos.

En el subsistema de almacenamiento es donde se guarda toda la información que ha sido cargada desde todas las fuentes de datos. Utilizando la herramienta PostgreSQL, esta información es almacenada en la BD en las tablas de hechos y dimensiones que tendrá el sistema.

Por último, en el subsistema de visualización es donde se muestra toda la información de forma multidimensional al cliente. Se revelan además, todos los reportes que les permitan a este realizar un análisis de toda la información procesada.



**Figura 5. Arquitectura de subsistemas**





## 2.4. Diseño

Después de haber concluido con el análisis del MD, se procede a diseñar el mismo. En esta fase se identifican las dimensiones, hechos y medidas que servirán para la confección de la matriz bus y del modelo de datos del MD de las Series históricas objetivos cubanos del milenio en el área de Información.

### 2.4.1. Identificación de dimensiones, hechos y medidas

Para identificar las tablas de dimensiones se tuvo en cuenta que estas constituyen las fuentes de todos los reportes identificados en el MD y las consultas que se les pueden realizar a este. Son las que describirán los objetos relevantes para las series y las que caracterizarán los hechos que se identificarán posteriormente. Las dimensiones identificadas durante la fase de diseño son:

















- `dim_temporal_anno`: una dimensión año debido a que los indicadores de las series se ejecutarán con una periodicidad anual.

dimensiones.dim_temporal_anno	
 <code>dim_temporal_anno_id</code>	<code>int4</code>
 <code>anno_codigo</code>	<code>varchar(4)</code>
 <code>anno_nombre</code>	<code>varchar(4)</code>
 <code>anno_numero</code>	<code>int4</code>

**Figura 6. dim\_temporal\_anno**

- `dim_indicadores_com`: una dimensión que almacene cada uno de los indicadores de las Series históricas objetivos cubanos del milenio en el área de Información, divididos por objetivos, por





metas y por tipo de indicador. Además, consta de una descripción de cada uno de los indicadores y la fuente de donde fueron extraídos los valores.

mart_com.dim_indicadores_com	
 <b>dim_indicadores_com_id</b>	<b>int4</b>
 objetivo	varchar(80)
 codigo_meta	int4
 meta	varchar(200)
 codigo_tipo	int4
 tipo_indicador	varchar(35)
 codigo_nombre	int4
 nombre_indicador	varchar(200)
 codigo_categoria	int4
 categoria_indicador	varchar(200)
 codigo_detalle	int4
 detalle_indicador	varchar(200)
 nombre_meta	varchar(200)
 indicador	varchar(200)
 definicion	varchar(500)
 fuente	varchar(500)

**Figura 7. dim\_indicadores\_com**

Para poder identificar la tabla de hechos del MD se tuvo en cuenta que esta debe ser la tabla primaria en el modelo multidimensional desarrollado y que debe además contener los valores de todos los indicadores de las series y los identificadores de las dimensiones asociadas a ella. Después de analizado esto, se identificó como tabla de hecho para las Series históricas objetivos cubanos del milenio en el área de Información la siguiente:

- hech\_com

mart_com.hech_com	
 <b>dim_indicadores_com_id</b>	<b>int4</b>
 <b>dim_temporal_anno_id</b>	<b>int4</b>
 valor_indicador	float4 

**Figura 8. hech\_com**

Para identificar las medidas que se almacenarán en la tabla de hechos se tuvo en cuenta que estas deben ser atributos numéricos y las variables de salida del diseño. La medida identificada para la tabla de hecho diseñada es:

- valor\_indicador: medida que almacena el valor de todos los indicadores de las Series históricas objetivos cubanos del milenio en el área de Información.

### 2.4.2. Matriz bus o dimensional

Después de haber identificado la tabla de hechos y las tablas de dimensiones se diseñó la matriz bus para el MD, también conocida como matriz dimensional. En ella se representó la relación existente entre estas tablas, a fin de mostrar cómo quedará estructurada la BD relacional.

Dimensiones/Hecho	hech_com
dim_temporal_anno	X
dim_indicadores_com	X

Tabla 3. Matriz bus

### 2.4.3. Modelo de datos

Después de haber identificado las dimensiones, la tabla de hecho, la medida y la matriz bus, se procede a modelar la BD utilizando un diseño lógico conocido como modelo de datos. Para el diseño de este se utilizó el esquema de estrella, debido a que existe una tabla de hecho central que tiene relación con las distintas dimensiones existentes.

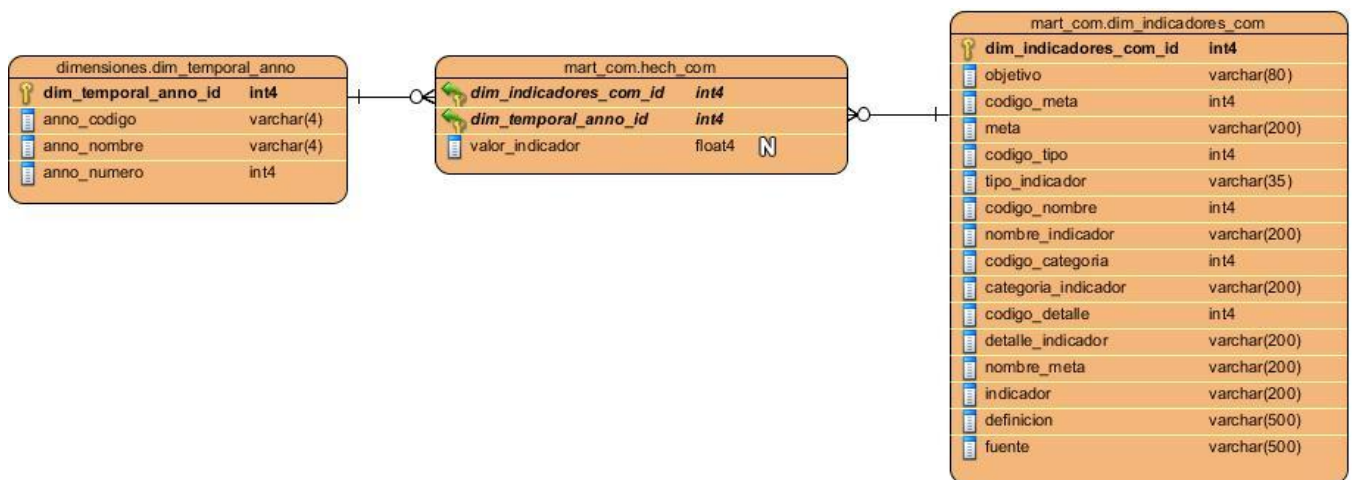
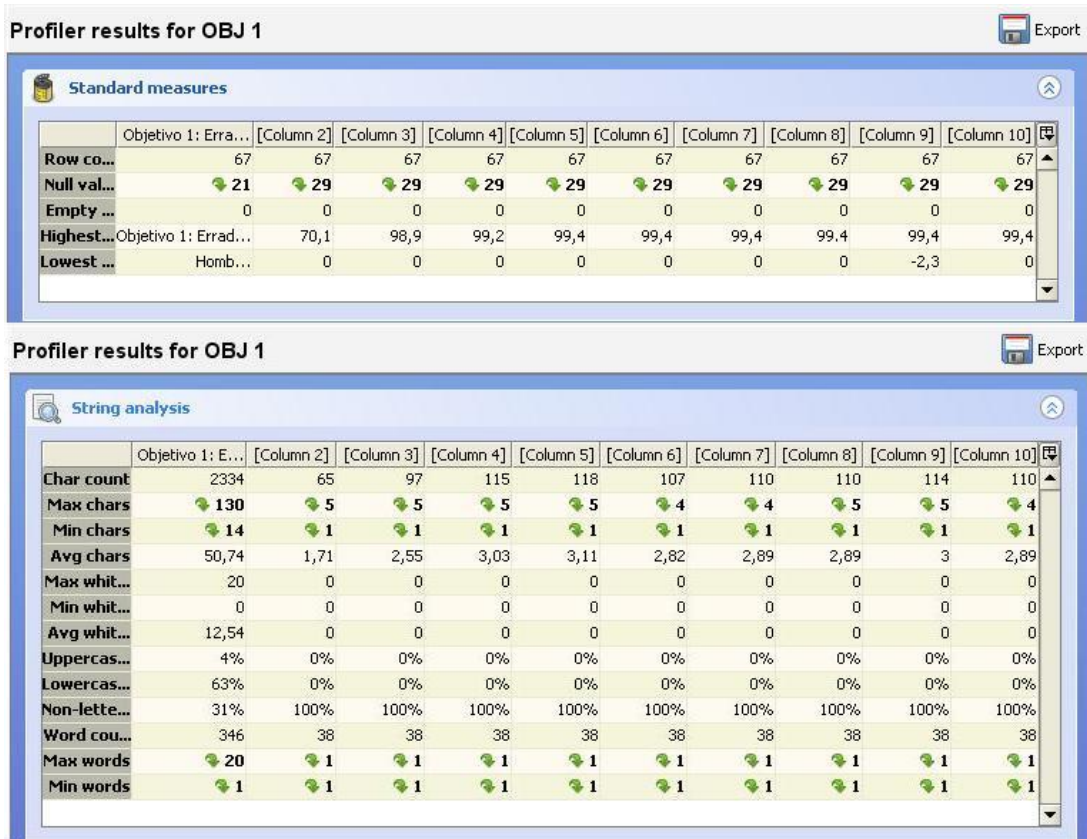


Figura 9. Modelo de datos

### 2.4.4. Perfilado de datos

Con el objetivo de efectuar un adecuado diseño del subsistema de integración, es necesario conocer el estado en el que se encuentra la información procedente de la fuente de datos, su calidad y su estructura. Es preciso realizar un análisis de la misma para lograr su mejor entendimiento y poder definir las reglas de transformación que se tendrán en cuenta para el diseño del subsistema y se llevarán a cabo en la implementación del mismo. Para ello fue imprescindible realizar un perfilado de los datos utilizando la herramienta DataCleaner, donde se obtuvieron resultados sustanciales, mostrándose en la figura no. 14 el perfilado realizado a la hoja 1 del AEC ODM Libro completo. El perfilado de las hojas restantes y de la fuente de datos ficha se encuentran disponibles en el artefacto "Perfil de los Datos". Con la realización del perfilado a las fuentes de datos, se obtuvieron los

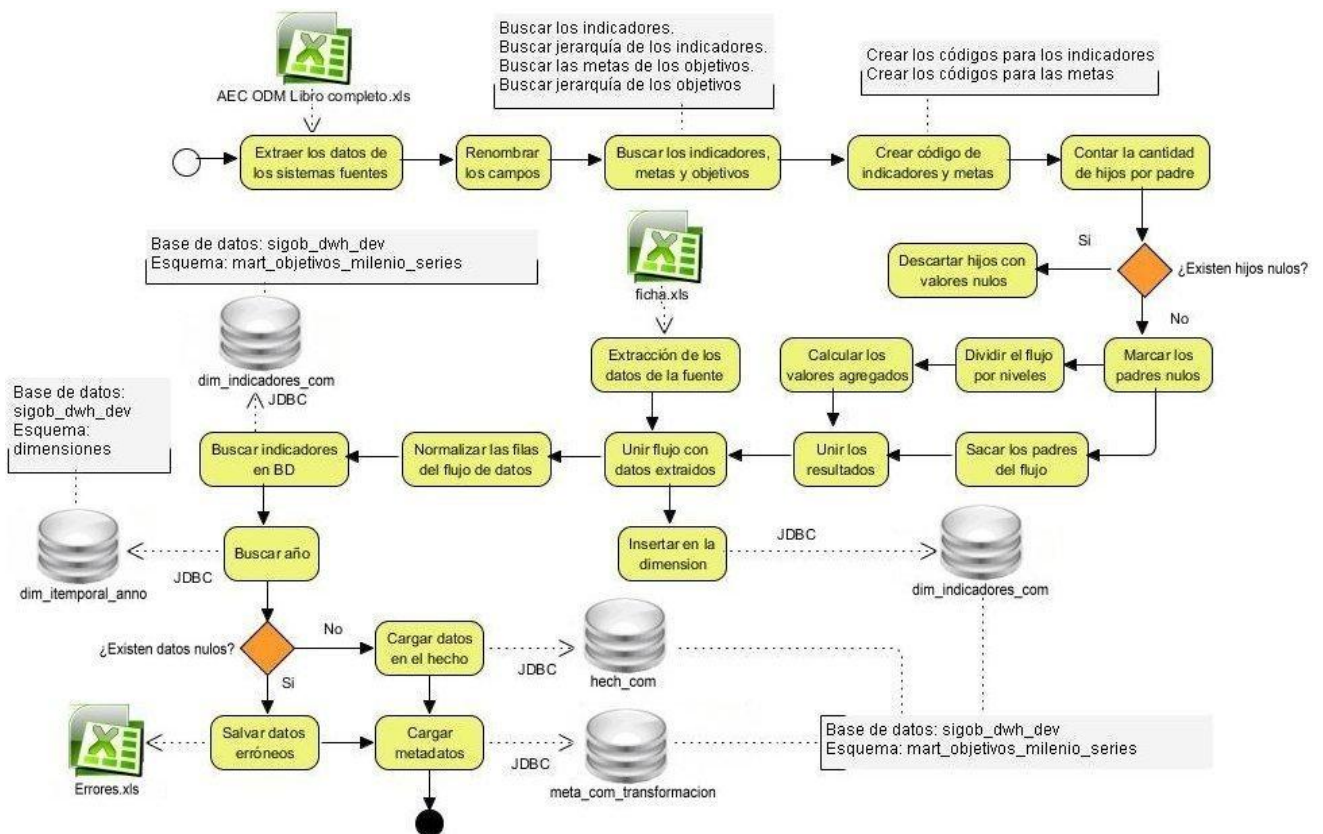
siguientes resultados: se analizaron dos fuentes de datos en formato *Excel*, la primera que tenía ocho hojas y la segunda una hoja. De ella se comprobaron 5508 valores, quedando identificados 3183 valores únicos, 1739 valores nulos y 565 valores duplicados.



**Figura 10. Perfilado de los datos**

#### 2.4.5. Diseño del subsistema de integración

Para el diseño del subsistema de integración de datos es preciso hacer un análisis profundo de los datos y de las especificaciones reflejadas en el artefacto “Reglas del negocio y transformación”. Esta actividad mencionada con anterioridad, posibilita crear las reglas de transformaciones que complementan a las reglas del negocio y describen las especificaciones que van a presentar el subsistema de integración para lograr la satisfacción del cliente. Atendiendo a los resultados arrojados por el proceso previo realizado al diseño del subsistema de integración de datos, se muestra en la figura no. 10 una imagen donde quedan reflejadas las especificaciones que van a presentar los procesos de integración de datos.



**Figura 11. Diseño del subsistema de integración**

#### 2.4.6. Diseño del subsistema de visualización

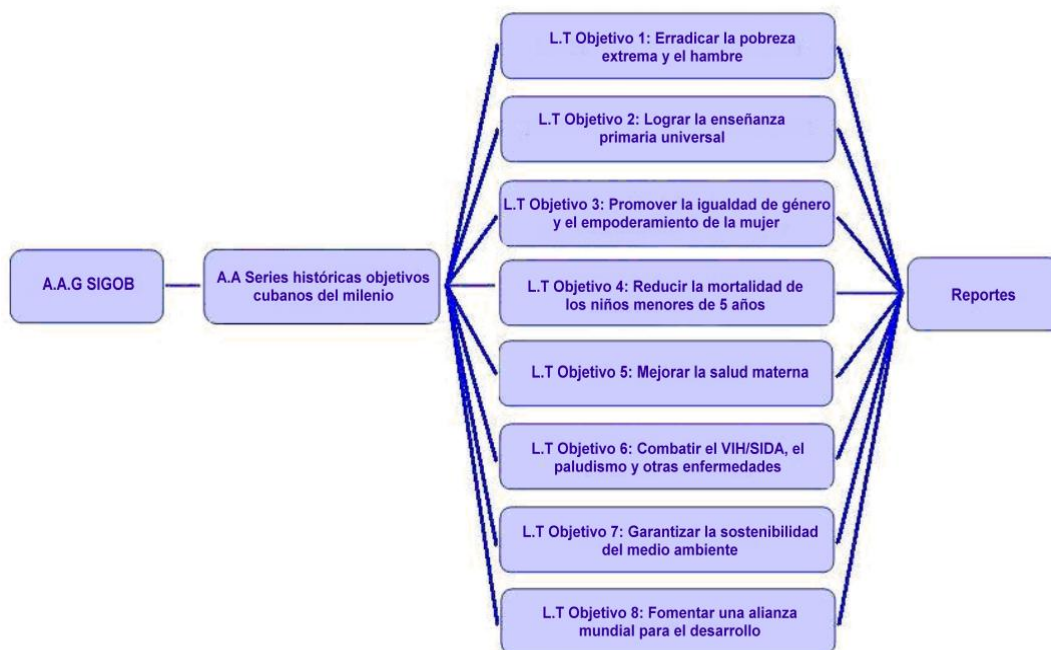
Para diseñar el subsistema de visualización de los datos, se deben agrupar las necesidades del cliente en carpetas y ficheros. Con el fin de mantener su seguridad, se puede asignar el acceso a un usuario o a un rol determinado. Se debe primero almacenar la información concerniente a las Series históricas objetivos cubanos del milenio en el área de Información. Estos contendrán los reportes construidos para poder visualizar los indicadores de los ODM. Seguidamente, se abordan cada uno de los elementos que componen la estructura del mapa de navegación:

Área de Análisis General (A.A.G): permite almacenar la información concerniente a la ONEI. Se definió como A.A.G a SIGOB.

Área de Análisis (A.A): son las áreas que se encuentran contenidas dentro del A.A.G y constituyen las secciones en las que está dividido el AD. En el diseño del MD se definió como A.A a las Series históricas objetivos cubanos del milenio.

Libro de Trabajo (L.T): son los libros que se encuentran contenidos dentro del A.A, constituyen las categorías en los que se agrupan los reportes. En el diseño del MD se definieron ocho L.T que se muestran en la figura 11.





**Figura 12. Mapa de navegación**

## 2.5. Esquema de seguridad

La seguridad para el MD Series históricas objetivos cubanos del milenio en el área de Información está determinada en gran medida por los niveles de acceso al sistema. Dicha seguridad se rige fundamentalmente por los roles y permisos que los usuarios poseen en su interacción con la BD y la aplicación. Para el acceso al MD se definió un usuario por cada uno de los roles existentes en el sistema con el objetivo de garantizar un control de acceso, operando cada usuario en el sistema según los permisos que se le definieron al rol.

Roles	Permisos	
	Lectura	Escritura
Sobre la base de datos		
Administrador del sistema	<b>x</b>	<b>x</b>
Administrador ETL	<b>x</b>	<b>x</b>
Analista	<b>x</b>	
Sobre la aplicación		
Administrador del sistema	<b>x</b>	<b>x</b>
Analista	<b>x</b>	

**Tabla 4. Roles y permisos en la BD y la aplicación**

## 2.6. Política de respaldo y recuperación

La política de respaldo y recuperación del MD Series históricas objetivos cubanos del milenio en el área de Información, juega un papel fundamental para que esté preparado a fin de recuperarse ante la pérdida de cualquier información. Con el objetivo de contrarrestar dicha posible pérdida y poder poner en marcha nuevamente el MD, se definieron los tres puntos siguientes:

- Periodicidad de salvos del sistema: estas se harán anuales a toda la información contenida en el MD, verificando siempre la existencia de una copia de toda la información almacenada.
- Tablas involucradas: las tablas que se involucran en la realización de las salvos son: hech\_com. En caso de que cambie alguna dimensión en un determinado período del tiempo, se realizará una salva de la misma.
- Salvos existentes: aunque no existan salvos en esta área se tiene previsto la realización de reemplazos de *backups* cada dos años. También se realiza una salva obligatoria, por lo regular en el mes de septiembre.

Además, se hará una copia total de la BD cada vez que se realicen las cargas de los ficheros fuentes, en este caso se ejecutará anualmente. Se guardarán las salvos en discos duros por un período de dos años.

## 2.7. Conclusiones

En este capítulo se abordó el análisis y diseño del MD que permitió cumplir con el segundo objetivo específico propuesto en la investigación, presentando como conclusiones:

- Se determinaron las necesidades de información de las series históricas objetos de estudio. Se realizó el levantamiento de requisitos en las que se describieron dos RI, 22 RF y 16 RNF. Una vez realizado el levantamiento de requisitos, se definieron los CUS que permitirán la interacción del usuario con el sistema, identificándose un solo CUI.
- El modelo de datos se diseñó partiendo de la representación gráfica del esquema en estrella, al ser definida una sola tabla de hecho relacionada con las dos tablas de dimensiones que fueron identificadas.
- La arquitectura del MD elaborada permitió identificar como subsistemas que la componen los de integración, almacenamiento y visualización implicados en la elaboración de la aplicación.
- En el diseño del subsistema de integración se realizó el perfilado de los datos, identificándose una regla de transformación: las llaves de las tablas de dimensiones no deben tomar valores nulos. Seguidamente, se realizó el diseño de las transformaciones lo que permitirá servir como guía para la implementación de dicho subsistema.



- En el diseño del subsistema de visualización fueron diseñados el mapa de navegación y los reportes candidatos, esto garantizará tener un mejor punto de vista de cómo quedarán estructuradas las necesidades del cliente, agrupadas en carpetas y ficheros para una posterior implementación del subsistema.

## CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBAS DEL MERCADO DE DATOS SERIES HISTÓRICAS OBJETIVOS CUBANOS DEL MILENIO EN EL ÁREA DE INFORMACIÓN

### 3.1. Introducción

En este capítulo se presenta el procedimiento aplicado para la etapa de implementación del MD. Se desarrollan los procesos ETL, al igual que el subsistema de visualización de los datos (BI), teniendo en cuenta los requerimientos del cliente. Se diseñan los cubos OLAP, se implementan los reportes candidatos, se desarrolla la estructura de navegación y se incluye la configuración de la seguridad de los usuarios. Finalmente, se ejecutan las pruebas que valida el MD y se exponen los resultados obtenidos en cada una de ellas luego de ser aplicadas.

### 3.2. Implementación del modelo de datos

La forma en que se encuentran estructurados y organizados los datos, hace que el modelado de ellos sea uno de los componentes más importantes en el inicio del desarrollo de cualquier proyecto. Por tal razón, se precisa una correcta organización de estos, pues esta estructura facilita la manipulación de ellos, constituye por tanto, el éxito y la calidad de la aplicación.

#### 3.2.1. Estructura de los datos

Después de diseñado el modelo dimensional, este es transformado en un modelo físico del cual es generado el *script* de la BD, donde se evidencian las relaciones que existen entre las diferentes tablas y a la vez, determina si el proyecto cumplirá o no con su propósito. Para ello la implementación de las tablas del modelo de datos se dividió en tres esquemas: dimensiones, mart\_objetivos\_milenio\_series y metadatos. En el esquema dimensiones se almacenarán las dimensiones comunes o compartidas en el AD, en el esquema mart\_objetivos\_milenio\_series, los hechos y las dimensiones específicas del MD y en el esquema metadatos las tablas para capturar los metadatos de los procesos ETL.

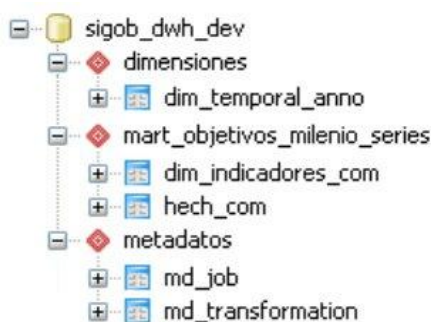


Figura 13. Estructura física de la BD

### 3.2.2. Usuarios y privilegios

Para establecer seguridad a la BD, se definieron los usuarios y sus privilegios a fin de agruparlos según las necesidades de permisos y accesos que necesitará cada rol para ejercer su función como trabajador del sistema:

- Administrador del sistema: tiene acceso a la BD en su totalidad, tanto a su administración y configuración. Controla el derecho de accesos del resto de los usuarios al sistema.
- Analista: tiene permisos de solo lectura en la BD, debido a que únicamente consulta la información contenida dentro de esta.
- Administrador de ETL: es el responsable de realizar los procesos ETL que interactúan con la BD.

### 3.3. Implementación del subsistema de integración

Para implementar el subsistema de integración es necesario desarrollar las tareas relacionadas con los procesos ETL. Para ello los datos extraídos de las fuentes de información, son transformados y limpiados, según las reglas definidas en el negocio y posteriormente, cargados en el sistema. La realización de estos pasos tiene como objetivo transferir los datos desde las aplicaciones de producción a los sistemas de BI.

#### 3.3.1. Arquitectura de integración

Para la integración de los datos se requiere de la definición previa de una arquitectura, la cual constituye la guía para la construcción del MD. Esta se encuentra estructurada de la siguiente forma:

- Fuente de datos: son los archivos fuentes con formato *Excel* que contienen la información que será extraída y modificada de acuerdo a las necesidades de información del cliente.
- Área temporal: es el área intermedia que separa la fuente de datos del MD, donde los datos serán integrados, limpiados y transformados, siendo conocida en la terminología inglesa como *Staging Area*.
- Mercado de Datos: es donde serán cargados los datos ya integrados para un futuro análisis.



**Figura 14. Arquitectura del subsistema de integración**

### 3.3.2. Implementación de las transformaciones y los trabajos

A fin de extraer la información de cada fichero de la fuente de datos para ser cargadas en cada tabla de dimensión y hecho de la BD, se procede a implementar las transformaciones. Estas constituyen la ejecución de un conjunto de pasos enlazados por medio de saltos que conforman caminos por donde fluye la información.

Para poder cargar la información de los ficheros fuente AEC ODM Libro completo y ficha, hacia la tabla de dimensión `dim_indicadores_com` y la tabla de hechos `hech_com`, se implementó como transformación:

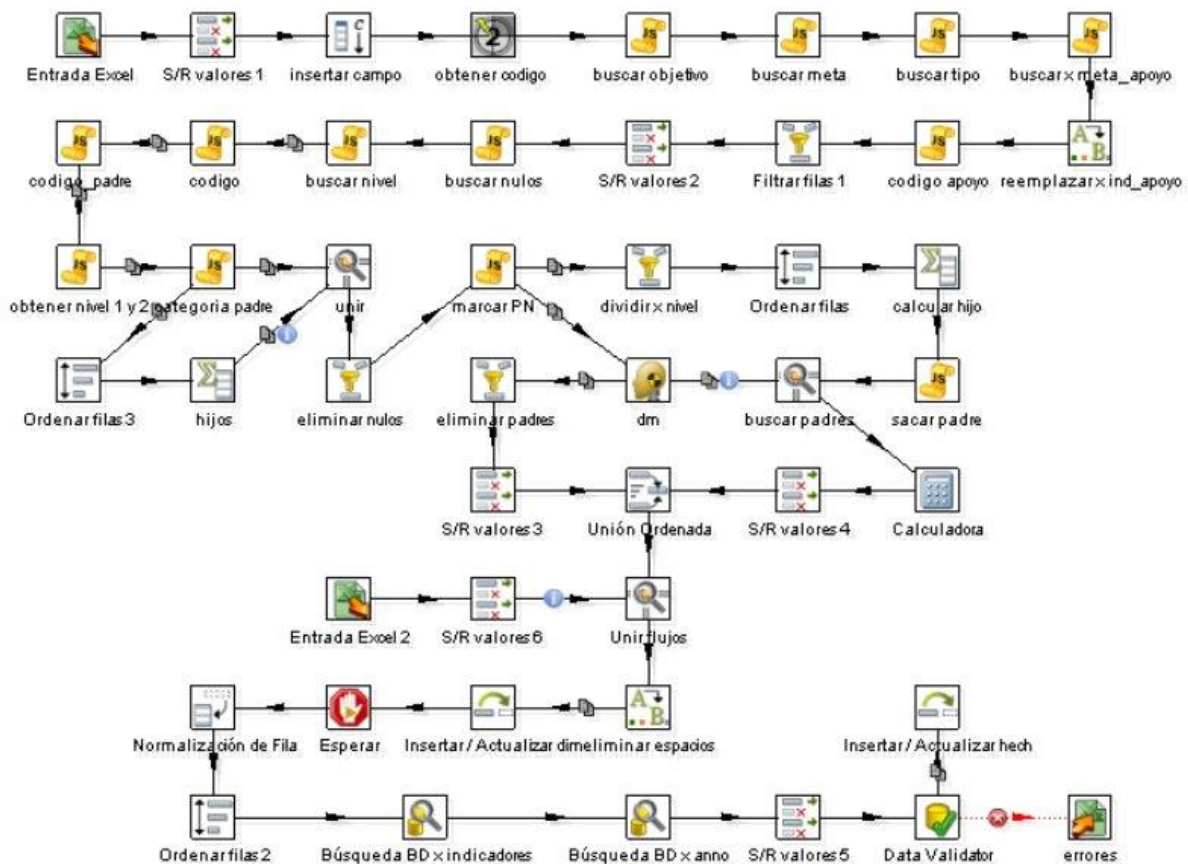
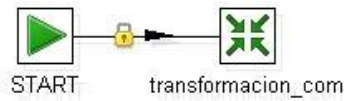


Figura 15. Implementación de la Transformación COM

Después de haber realizado la implementación de la transformación, se procede a desarrollar el trabajo que ejecutará dicha transformación. A continuación se muestra el trabajo implementado para el MD:



**Figura 16. Trabajo para el MD**

### **3.3.3. Control en los procesos de integración de datos**

El éxito de los AD depende en gran medida de la calidad implícita en los procesos de integración de datos. Desde que se encuentran en las fuentes de información hasta que son integrados en el AD, los datos deben ser sometidos a varias técnicas de calidad que van desde el perfilado y limpieza de los datos hasta la auditoría y control de todo el flujo de integración. Una pieza clave y común en estas técnicas, que juega un rol fundamental en los almacenes, es que todas son gestionadas a través de los metadatos. (26)

Los metadatos son definidos usualmente como “datos acerca de los datos”. Estos pueden documentar además: objetos, reglas del negocio, los procesos que manipulan los datos, las tablas que contiene el AD, las columnas que poseen cada una de esas tablas y el tipo de datos que se puede almacenar. Debido a la información que almacenan los metadatos, estos se hacen muy técnicos para el entendimiento de los usuarios finales, no obstante, es imprescindible su documentación ya que ellos demuestran la credibilidad de los datos, su integridad y son de gran interés para los programas que tienen que manejar estos datos.

En los procesos de integración implementados en el MD del presente proyecto, se realizaron dos tablas de metadatos. La primera tuvo como objetivo controlar la ejecución de la transformación implementada, guardando el nombre, fecha de inicio y fin de la ejecución, así como los errores ocurridos en la misma. La segunda se estableció para mantener un historial de los registros y las trazas del comportamiento del trabajo implementado en el MD.

### **3.4. Implementación del subsistema de visualización**

Una vez concluida la fase de integración de los datos, se da inicio a la fase de visualización. En esta etapa es donde se concibe ya la solución final que permitirá sistematizar la información generada de los distintos procesos de la empresa que serán los que ofrezcan la información necesaria para la toma de decisiones. Además, facilitará establecer estrategias, conocer las fortalezas y debilidades del negocio.

### 3.4.1. Implementación de los cubos OLAP

Uno de los elementos más importantes de la fase de visualización son los cubos OLAP, debido a que es en ellos donde se organiza la información estructural, se definen las dimensiones, los niveles de jerarquía de las dimensiones y las medidas. El MD Series históricas objetivos cubanos del milenio en el área de Información presentará un solo cubo OLAP. Para su diseño e implementación se tuvieron en cuenta las tablas de hecho y dimensiones identificadas. Este fue modelado usando la herramienta Pentaho Schema Workbench.



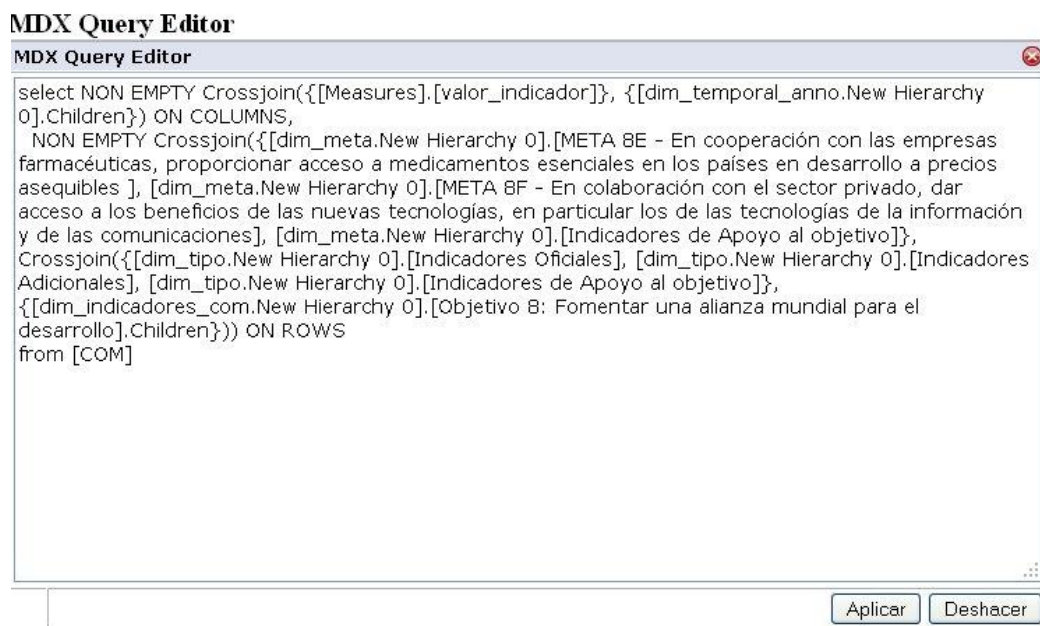
Figura 17. Implementación del cubo OLAP

### 3.4.2. Implementación de los reportes candidatos

Después de haber concluido con la implementación de los cubos OLAP, le corresponde la implementación de los reportes candidatos. Estos representan la forma en la que el cliente desea que

se muestre la información para el análisis de los datos que con anterioridad fueron cargados en el proceso ETL.

Para la implementación de los reportes con el fin de analizar los indicadores de las series, se realizaron consultas MDX, debido a que estas permiten consultar BD multidimensionales y ofrecen la posibilidad de modificar la estructura o de rotar el cubo. En total ocho consultas MDX fueron implementadas, una por cada reporte. Para la implementación de los reportes utilizados para consultar las notas técnicas de los indicadores de las series, se realizaron ocho consultas SQL, una por cada reporte. Como ejemplo de una de las consultas MDX se presenta la siguiente:



**Figura 18. Consulta MDX del reporte Valor de indicador por indicadores de la alianza mundial**

### 3.4.3. Arquitectura de información

La arquitectura de información del MD objeto de investigación, se encuentra estructurada por un área de análisis que agrupa la información de estas series. Dentro de esta área la información se divide en ocho libros de trabajo, uno para cada objetivo del milenio y dentro de cada libro, el reporte que visualiza los indicadores de ese objetivo y las notas técnicas de dichos indicadores.

Para la visualización de los reportes se utilizó la herramienta Mondrian OLAP Server, como servidor *web* el Apache Tomcat y finalmente, para conformar la estructura del mapa de navegación junto con todos los reportes, el Pentaho BI Server. Los reportes Notas Técnicas fueron desarrollados con la herramienta Pentaho Report Designer, por tener las particularidades de que no son necesarios mostrarlos desde diversas perspectivas de análisis, ellos solo muestran las descripciones y las fuentes



de cada indicador. En la siguiente ilustración se muestra la aplicación final, visualizando el reporte Valor de indicador por indicadores de la alianza mundial y más adelante, una imagen del reporte Notas Técnicas del Objetivo 8:

Indicadores COM	Medidas								
	Valor Indicador								
	Año								
	1990	1995	2000	2005	2006	2007	2008	2009	
8.14 - Líneas de teléfono fijo por cada 100 habitantes	5,6	3,2	4,4	7,6	7,7	8,2	8,5	10	
8.15 - Abonados a teléfonos celulares por cada 100 habitantes				1,2	2,2	3	4,1	5,5	
8.16 - Usuarios de internet por cada 100 habitantes				10	11	12	13	14	
8F.2 - Cantidad de Joven Club de Computación (U)				587	600	602	611	607	
8F.3 - Teléfonos instalados de todo tipo (MU)	599,5	652,6	826,6	1.167,4	1.354,2	1.377,3	1.424,1	1.789,3	
8F.4 - Estaciones públicas instaladas (U)	10.331	6.028	13.996	34.751	40.358	44.126	50.530	53.414	5
8F.5 - Cable de fibra óptica (m/km)				0,9	3,4			1,3	
8F.6 - Fibra instalada (m/km)				14	46,7			29,4	
8F.7 - Proporción de digitalización (%)				51,7	89,8	92,2	94,9	95,6	97,1
8F.8 - Computadoras en el sistema de educación (U)				88.049	117.440	125.171	121.259	118.009	
8a - Estudiantes de otros países en Cuba (U)				15.854			31.459	30.719	2
8b - Graduados de otros países en Cuba (U)				45.352			52.662	55.188	3

Figura 19. Vista del reporte de los indicadores de la Alianza Mundial

Meta	Tipo de Indicador	Indicador	Definición	Fuente
Meta 8E	Indicadores Oficiales	8.13 - Proporción de la población con acceso sostenible a medicamentos esenciales a precios asequibles	Porcentaje de la población con enfermedades crónicas controladas mediante medicamentos (asma bronquial, diabetes mellitus, hipertensión arterial, hipercolesterolemia, trastornos psiquiátricos, epilepsia, entre otras); así como tratamientos eventuales por prescripción facultativa, que disponen de forma mantenida y sistemática de sus medicamentos, a precios asequibles y en dispensarios cercanos a sus domicilios.	Ministerio de Salud Pública
Meta 8F	Indicadores Oficiales	8.14 - Líneas de teléfono fijo por cada 100 habitantes	Es el número de líneas telefónicas fijas que conectan el equipo	Ministerio de la Informática y las Comunicaciones. Oficina Central de Estadística

Figura 20. Vista del reporte Notas Técnicas del Objetivo 8



#### **3.4.4. Configurar la seguridad de los usuarios**

A fin de proporcionar una mayor seguridad al sistema, en la implementación del subsistema de visualización del MD se crearon dos usuarios y roles, los cuales poseen diferentes permisos de acceso a la información:

- Rol de administrador: tiene todos los permisos de la aplicación.
- Rol de analista: tiene permiso de solo lectura, posee única y exclusivamente permisos de ejecución sobre el sistema.

### **3.5. Pruebas**

#### **3.5.1. Pruebas de software**

Una vez ejecutada la etapa del proceso de desarrollo, se debe proceder a la validación de la aplicación final. Las pruebas de software están formadas por los procesos críticos, los procesos necesarios, que permiten verificar y validar la calidad de un producto. Estas son empleadas para identificar posibles fallos de implementación, con el objetivo de comprobar y determinar si el producto cumple con lo esperado por el cliente.

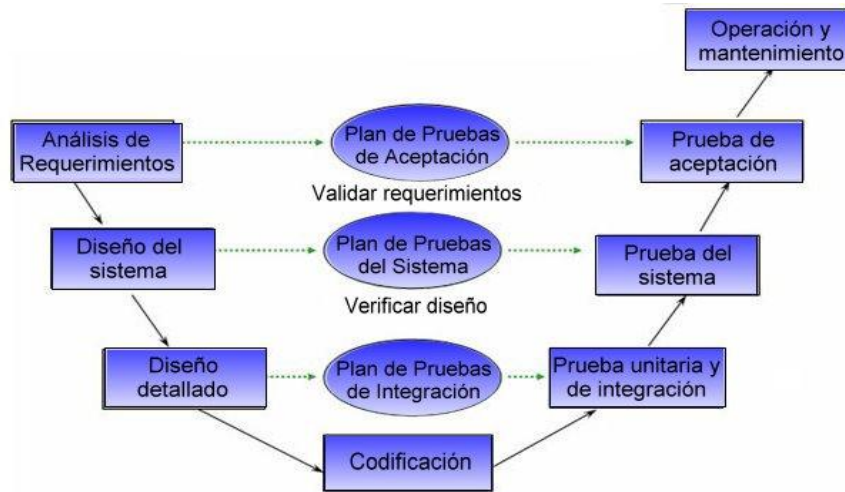
Para determinar el nivel de calidad pueden ejecutarse pruebas dirigidas a componentes del software o al sistema de software completo, de forma tal que se pueda comprobar y medir el grado en que el producto cumple con los requerimientos y especificaciones iniciales del producto. Estas pruebas permiten depurar un software desde sus inicios, a fin de que este llegue a manos del cliente cumpliendo todas sus funcionalidades y libre de defectos.

#### **3.5.2. Modelo V**

Para mejorar el proceso de pruebas es utilizado el Modelo V, el cual ha sido aplicado por DATEC. Este modelo es un procedimiento que posibilita la validación del correcto funcionamiento de los AD. Como principal ventaja presenta la posibilidad de ejecutar las pruebas en todo el ciclo de vida de desarrollo del software.

El Modelo V describe, a un alto nivel de abstracción, las fases del ciclo de desarrollo en las que se involucra la prueba, presentando la interrelación de las actividades de pruebas con las de análisis y diseño. Este procedimiento constituye una representación de dos flujos de actividades relacionadas, donde presenta como punto de intersección la codificación. A la izquierda se presentan las actividades relacionadas al desarrollo y a la derecha, las del aseguramiento de la calidad del software. La vinculación entre ambas partes implica que si se hallan problemas durante la verificación y la validación, entonces el lado izquierdo puede ser ejecutado nuevamente para solucionar el problema y

mejorar los requerimientos, el diseño y el código, antes de retomar las pruebas del lado derecho. El siguiente esquema constituye una representación gráfica de este modelo:



**Figura 21. Modelo V**

Las pruebas que desarrolla este modelo son las siguientes:

- Pruebas unitaria: son diseñadas y ejecutadas por el desarrollador una vez concluido el desarrollo de cada componente. Es la prueba que valida si los componentes individuales de un programa están funcionando adecuadamente, permitiendo descubrir discrepancias entre la especificación de la interfaz de los módulos y su comportamiento real.
- Pruebas de integración: son diseñadas y ejecutadas por el desarrollador de conjunto con los especialistas de DATEC cuando la aplicación está completa. Es la prueba que determina la integración de los componentes dentro de un sistema y evaluar su correcta interfaz, funcionalidad y desempeño.

Las pruebas unitaria y de integración deben ser realizadas para comprobar el diseño, verificando que todos los aspectos de este y del programa se han implementado correctamente en el código.

- Pruebas de sistema: son ejecutadas por el desarrollador o equipo de pruebas en un ambiente controlado. Es la prueba que debe verificar el diseño del sistema, asegurando que todos los aspectos del diseño estén correctamente implementados. Por consiguiente, este tipo de prueba determina el correcto funcionamiento de un sistema y su cumplimiento contra las especificaciones del producto. Permite detectar discrepancias entre el comportamiento del sistema construido y su especificación.

- Pruebas de aceptación: son pruebas funcionales que se realizan directamente con el cliente para validar los requerimientos y su conformidad con el producto. Son aquellas que demuestran al cliente que la aplicación está terminada y funciona correctamente.

### **3.5.3. Herramientas de prueba**

En la aplicación final, son utilizadas las siguientes herramientas de prueba:

- Casos de pruebas
- Lista de chequeos

#### Casos de prueba:

Constituye el conjunto de entradas de pruebas, condiciones de ejecución y resultados esperados, diseñados con el fin de cumplir un objetivo en particular o una función esperada. Esta herramienta permite obtener una probabilidad muy alta para descubrir un nuevo error, lo cual garantizará que el producto se comporte tal como se describe en las especificaciones funcionales del diseño, verificando así el cumplimiento de los requisitos de información a través de los reportes candidatos. Cada caso de prueba es diseñado de acuerdo a las funcionalidades descritas en los CUI, elaborándose previamente a la ejecución de las pruebas a la aplicación.

#### Lista de chequeo:

La lista de chequeo constituye un documento que contiene un conjunto de parámetros a medir sobre un aspecto determinado con el objetivo de verificar el grado de cumplimiento de determinadas reglas establecidas previamente. A través de un formulario de preguntas referidas a los artefactos y el software en general, esta importante herramienta metodológica permite medir y evaluar la confiabilidad, así como la seguridad del producto, creando un mecanismo fiable y reproducible que garantice mejorar las normas de calidad.

Cada pregunta se encuentra asociada a una evaluación en una escala que da una medida del grado de cumplimiento y disponibilidad de la propiedad evaluada, de esta forma, se determina la evaluación del elemento probado. Las respuestas deben determinar de modo satisfactorio el cumplimiento o no de los procedimientos que se estén evaluando.

### **3.5.4. Pruebas realizadas al MD**

#### Pruebas unitaria y de integración:

En el trabajo de diploma que se expone, las pruebas unitarias y de integración fueron aplicadas a medida que se implementaron los subsistemas de integración, almacenamiento y visualización.

Durante los procesos de ETL se verificó que la carga de los datos se realizara correctamente. El subsistema de visualización se validó a través de las consultas realizadas, teniendo en cuenta que estas cumplieran con lo definido en los requisitos de información, lo cual corresponde con la necesidad del cliente. Durante la ejecución de estas pruebas fueron detectadas un total de tres no conformidades distribuidas en los tres subsistemas.

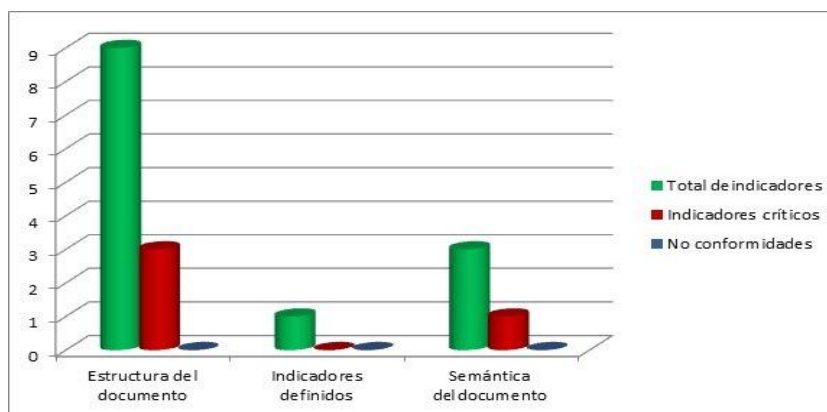
- Subsistema de integración
  - Poner el componente “Esperar” detrás de “Insertar/Actualizar dim” y no del componente “eliminar espacios”.
- Subsistema de almacenamiento
  - Arreglar el tamaño de los textos.
- Subsistema de visualización
  - Poner las metas, los tipos de indicadores y los indicadores en el mismo orden que aparecen en la fuente de datos.

La herramienta de prueba utilizada en estas pruebas fue la lista de chequeo. Para su elaboración se tomaron en consideración los elementos de evaluación, que no deben ser excluidos una vez que se comiencen a redactar los artefactos de ETL, permitiendo obtener los puntos satisfactorios o no que posean dichos artefactos.

La lista de chequeo se encuentra dividida en tres secciones:

- Estructura del documento: abarca todos los aspectos definidos por el expediente de proyecto o el formato establecido por el proyecto.
- Indicadores definidos: abarca todos los indicadores a evaluar de los artefactos generados en los procesos de ETL.
- Semántica del documento: contempla todos los indicadores a evaluar respecto a la ortografía, redacción y demás.

Después de haberse aplicado la lista de chequeo en el artefacto “Diccionario de datos”, de un total de 13 indicadores identificados y cinco definidos de críticos, no fueron detectadas no conformidades. En la figura 22, se muestra gráficamente cómo fue evaluado el comportamiento de los indicadores en este artefacto. La lista de chequeo también fue aplicada a los artefactos “Mapa lógico de los datos”, “Perfilado de datos” y “Registro del sistema fuente”, los cuales fueron evaluados de forma satisfactoria al no presentar no conformidades.



**Figura 22. Evaluación del comportamiento de los indicadores del artefacto Diccionario de datos**

Pruebas de sistema:

Las pruebas de sistema fueron ejecutadas por los especialistas. En ellas fueron encontradas cuatro no conformidades: las variables de entrada y salida no se correspondían con la aplicación, así como un error ortográfico y dos escrituras en bloques pertenecientes a los indicadores que fueron cargados al almacén extraídos desde las fuentes de datos.

La herramienta de prueba que se utilizó en las pruebas de sistema fue casos de prueba. En el presente trabajo de diploma fue diseñado un caso de prueba debido a que solo fue identificado un CUI en la etapa de análisis. Para este caso de prueba se procedió a elaborar una planilla en la que fueron detalladas las funcionalidades descritas en el CUI, así como todas sus variables. En la tabla no. 5 se muestra el caso de prueba diseñado.

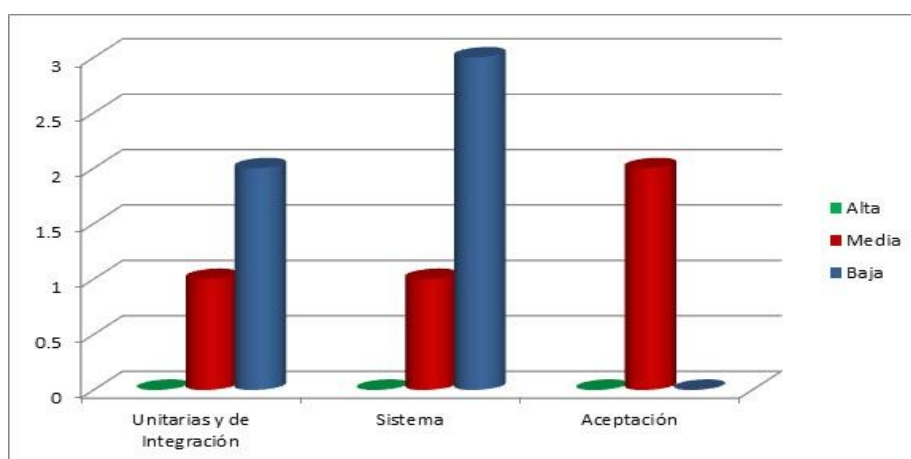
Escenario	Descripción	Variabes de Entrada	Variabes de Salida	Respuesta del sistema	Flujo central
EC 1.1 reporte prueba	Muestra los valores de los indicadores de los objetivos cubanos del milenio desde el año 1990 hasta el 2010.	Meta Tipo de indicador Indicadores COM Año	Valor de indicador	El sistema muestra todas las variables disponibles para el análisis, ubicadas en las filas y las columnas que pueden ser visualizadas para cada reporte.	En la parte superior izquierda se selecciona el A.A.G SIGOB. Se selecciona el A.A Series históricas objetivos cubanos del milenio. Se selecciona el L.T Prueba. En la parte inferior izquierda se selecciona el reporte prueba. En el área de trabajo se visualiza el reporte seleccionado.

**Tabla 5. Caso de prueba del CUI Presentar información de los indicadores de los objetivos cubanos del milenio**

### Pruebas de aceptación:

Finalmente, se realizaron las pruebas de aceptación por la representante de la ONEI en la universidad: Elena Leonila Fernández García, quien al verificar que los reportes candidatos se correspondieran con las fuentes de información, encontró dos no conformidades: los totales de las mujeres en el Objetivo 7 no coincidían con los totales de la fuente. Estas fueron resueltas satisfactoriamente y en una segunda iteración, el cliente confirmó que el MD satisfacía todas las necesidades de información. Además, pudo constatar el correcto funcionamiento y cumplimiento de los requisitos del producto, quedando oficializada su aprobación en la carta de aceptación.

En la siguiente figura se muestran las no conformidades encontradas en las pruebas aplicadas al MD, las cuales fueron corregidas satisfactoriamente:



**Figura 23. No conformidades encontradas en las pruebas aplicadas al MD**

### **3.6. Conclusiones**

En este último capítulo se realizó la implementación y pruebas para el MD, llegando a las conclusiones siguientes:

- En la Implementación del modelo de datos, se definió la estructura física con tres esquemas y cinco tablas que brindarán una correcta organización de los datos en la BD. Esta estructura facilitará la manipulación de los datos, determinando a su vez si el proyecto cumplirá o no con su propósito.
- Durante la implementación del subsistema de integración fue ejecutada una transformación para cargar los datos al MD que permitió poblarlo satisfactoriamente. Además, fueron cargados al MD las tablas de metadatos, que mantendrán un historial de los registros y trazas del comportamiento de la transformación y el trabajo que fueron implementados, a fin de controlar los procesos de integración.

- En la implementación del subsistema de visualización fue definida un área de análisis, que quedó dividida en ocho libros de trabajos, uno para cada objetivo del milenio y dentro de cada libro, el reporte que visualizará los indicadores de ese objetivo y las notas técnicas de dichos indicadores, dando como resultado 16 reportes destinados a los usuarios finales.
- Finalmente, fueron aplicadas las listas de chequeo que determinaron que los artefactos generados en la etapa de ETL, tuvieran la calidad necesaria. Además, se aplicaron los casos de pruebas que permitieron validar los reportes implementados. Todo lo anterior, permitirá garantizar que la solución final cumpla con lo establecido por el cliente.

## CONCLUSIONES

Como resultado de la investigación realizada, se desarrolló el Mercado de Datos Series históricas objetivos cubanos del milenio en el área de Información para el Sistema de Información de Gobierno, el cual permitió resolver satisfactoriamente el problema de investigación planteado, cumpliendo de esta forma con los objetivos propuestos y arribando por tanto, a las siguientes conclusiones:

- La revisión bibliográfica y documental del tema permitió fundamentar la selección de las metodologías, herramientas y tecnologías a utilizar en el desarrollo del MD objeto de investigación, definiéndose como metodología a emplear el Modelo para el Desarrollo de Soluciones de AD y BI aplicada en DATEC. Se establecieron los términos básicos necesarios relacionados con las herramientas que permitieron garantizar el desarrollo íntegro del MD, las cuales son: Visual Paradigm 8.4, el SGBD relacional PostgreSQL 9.1, el PgAdmin III 1.14.1, el DataCleaner 1.5.4, Pentaho Data Integration 4.2.1, Pentaho Schema Workbench 3.2.1, Pentaho Report Designer 3.6.1, Apache Tomcat 6.6, Pentaho BI Server 3.8.0 y Mondrian OLAP Server 3.0.4.
- En el análisis y diseño del MD se realizó el levantamiento de requisitos, definiéndose los CUS que permitieron la interacción del usuario con el sistema. La arquitectura del MD elaborada permitió identificar y diseñar los principales subsistemas que la componen: integración, almacenamiento y visualización, lo que constituyó una guía para la implementación de la aplicación final.
- Como resultado de la implementación, se logró obtener un MD poblado y funcional, con un formato estándar, organizado y centralizado de los datos, que posibilitó el análisis y la integración de toda la información disponible para ser consultada por parte de los usuarios, apoyando al proceso de toma de decisiones. En la validación de la calidad del MD se consideraron las necesidades de información del cliente, alcanzándose resultados satisfactorios a través del conjunto de casos de pruebas realizados y las listas de chequeos, lo que permitió demostrar la utilidad del proyecto presentado y cumplir por tanto, con las necesidades del cliente.



## **RECOMENDACIONES**

A partir del diseño realizado aplicando las metodologías, herramientas y tecnologías a utilizar en el desarrollo del MD, así como una vez presentado el proyecto de trabajo, se recomienda lo siguiente:

- Desarrollar nuevas formas de visualización de la información y aplicar otras técnicas de Inteligencia de Negocios.
- Elaborar una estrategia para cargar los datos que quedaron registrados en los ficheros de errores, una vez que los especialistas los hayan revisado.

## REFERENCIAS BIBLIOGRÁFICAS

1. Evaluación en cifras Objetivos de Desarrollo del Milenio a nivel local. Una mirada desde la región oriental de Cuba. ONEI. Imprenta Alejo Carpentier. ISBN: 978-959-7119-46-3.
2. Inmon, William H. (1996). - Building the Data Warehouse, Wiley Computer, New York.
3. Almacén de Datos. *EcuRed*. [En línea] [Citado el: 20 de Noviembre de 2010.]  
[http://www.ecured.cu/index.php/Almac%C3%A9n\\_de\\_Datos#Definici.C3.B3n\\_de\\_Ralph\\_Kimball](http://www.ecured.cu/index.php/Almac%C3%A9n_de_Datos#Definici.C3.B3n_de_Ralph_Kimball)
4. Sinergia e Inteligencia de Negocio. [En línea] 18 de enero de 2007. [Citado el: 2010 de octubre de 25.] <http://www.sinnexus.com/empresa/creditos.aspx>.
5. 8. R, Elmasri y S. B, Navathe. *Fundamentos de sistemas de bases de datos*. s.l.: University of Texas at Arlington, 2002.
6. Sánchez, Jean. “*Descubrimiento de conocimiento en Base de Datos (KDD) Fase de integración y recopilación*”. [En línea] [Citado el: 26 de Octubre de 2011.]  
<http://issuu.com/jculacio/docs/kddfase1>.
7. KIMBALL, Ralph y ROSS, Margy. *The Data Warehouse Toolkit*. EUA: Wiley Publishing Inc, 2002.
8. Ing. Javier Nader, Tesis de Magister en Ingenieria del Software “Sistema de Apoyo Gerencial Universitario
9. Manual para la construcción de un Data Warehouse, publicado por el Instituto Nacional de Estadística e Informatica INEI Lima, año 1997
10. Rodríguez Yunta, Luis. BASES DE DATOS DOCUMENTALES: ESTRUCTURA Y PRINCIPIOS DE USO. *Universidad de Navarra*. [En línea] 2001. [Citado el: 22 de Octubre de 2010.]  
[www.unav.es/dpp/documentacion/proteger/lryunta.pdf](http://www.unav.es/dpp/documentacion/proteger/lryunta.pdf).
11. Colectivo de autores: Trujillo, Juan Carlos. Mazón, José Norberto y Pardillo, Jesús. *Diseño y Explotación de Almacenes de Datos. Conceptos Básicos de Modelado Multidimensional*. Editorial Club Universitario. ISBN: 978-84-8454-902-4.
12. ¿Por qué Business Intelligence? [En línea]  
[http://www.sinnexus.com/business\\_intelligence/sistemas\\_informacion.aspx](http://www.sinnexus.com/business_intelligence/sistemas_informacion.aspx)
13. ¿Qué es Business Intelligence? [En línea] [http://www.sinnexus.com/business\\_intelligence/](http://www.sinnexus.com/business_intelligence/)
14. La metodología. [Online] [Cited: 01 10, 2011.]  
<http://www.ecured.cu/index.php/Metodolog%C3%ADa>.
15. Bitam. “Business Intelligence”. [en línea]. <http://www.bitam.com/spanish/AcercaDeBI.htm>  
[Consulta: Febrero de 2002].

16. Mundy, Joy. Kimball University: Building a Foundation for Smart Applications. *InformationWeek*. [En línea] 1 de Diciembre de 2006. [Citado el: 20 de Octubre de 2010.] [http://www.kimballgroup.com/html/articles\\_search/articles2006/0612IE.html](http://www.kimballgroup.com/html/articles_search/articles2006/0612IE.html).
17. Apuntes. Ingeniería del software. Sistemas Informaticos. Nivel de madurez software. Informatica Aplicada a la Gestión Pública. Página profesor Rafael Barzanallana. Universidad de Murcia (España). [En línea] 16 de Diciembre de 2006. [Citado el: 1 de Diciembre de 2010.] <http://www.um.es/docencia/barzana/IAGP/lagp2.html>.
18. Rodríguez Sanz, Miguel "ANÁLISIS Y DISEÑO DE UN DATA MART PARA EL SEGUIMIENTO ACADÉMICO DE ALUMNOS EN UN ENTORNO UNIVERSITARIO". Proyecto Fin de Carrera. Universidad Carlos III de Madrid Escuela Politécnica Superior Ingeniería en Informática. 22 Julio, 2010. <http://es.scribd.com/doc/52203545/16/Metodologia-propuesta-por-Bill-Inmon>.
19. Kimball vs Inmon. Ampliación de conceptos del Modelado Dimensional. [En línea] [Citado el: 13 de noviembre del 2011] <http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimencional/>
20. González Hernández, Yanisbel "Propuesta de Metodología para el desarrollo de Almacenes de Datos en DATEC". Centro de Tecnologías de Gestión de Datos (DATEC). La Habana UCI, septiembre de 2011.
21. Zepeda Sánchez, Leopoldo Zenaido. Metodología para el Diseño Conceptual de Almacenes de Datos. Tesis de Doctorado. Universidad Politécnica de Valencia. Valencia: s.n., 2008. [En línea] junio de 2008. [Citado el: 23 de noviembre de 2010.] <http://tesis.com.es/autores/zepeda-sanchez-leopoldo-zenaido>.
22. SILBERSCHATZ, ABRAHAM, KORTH, HENRY F. y S., SUDARSHAN. *Fundamentos de bases de datos*. s.l. : MCGRAW-HILL, 1995
23. En que se clasifican las bases de datos? [Online] [Cited: 01 10, 2011.] [http://www.netronycs.com/clasificacion\\_de\\_base\\_datos.html](http://www.netronycs.com/clasificacion_de_base_datos.html).
24. Conesa Caralt, Jordi y Curto Díaz, Josep. Introducción al Bussines Intelligence. s.l. : UOC - EDIUOC, 2010.
25. Geiger, Claudia Imhoff. Nicholas Galemno. Jonathan G. *Mastering Data Warehouse Design, Relational and Dimentional Techniques*. 2003. 0-471-32421-3
26. Medina, Doris. *Modelo de Auditoría y Metadatos para la Gestión de Proceso de Integración de Datos*. La Habana, Cuba, UCIENCIA 2012

## BIBLIOGRAFÍA

-  Almacén de Datos. EcuRed. [En línea] [Citado el: 20 de Noviembre de 2010.] [http://www.ecured.cu/index.php/Almac%C3%A9n\\_de\\_Datos#Definici.C3.B3n\\_de\\_Ralph\\_Kimball](http://www.ecured.cu/index.php/Almac%C3%A9n_de_Datos#Definici.C3.B3n_de_Ralph_Kimball)
-  Álvarez, Miguel Angel. Desarrollo Web. [En línea] 2007. [Citado el: 03 de 03 de 2010.] <http://www.desarrolloweb.com/articulos/sistemas-gestores-bases-datos>.
-  Barzanallana, Rafael. Apuntes Ingeniería del software. Sistemas Informáticos. Nivel de madurez software. Informática Aplicada a la Gestión Pública Universidad de Murcia (España). [En línea] 16 de Diciembre de 2006. [Citado el: 1 de Diciembre de 2010.] <http://www.um.es/docencia/barzana/IAGP/lagp2.html>.
-  Bitam. "Business Intelligence". [en línea]. <http://www.bitam.com/spanish/AcercaDeBI.htm> [Consulta: Febrero de 2002].
-  Bouman, Roland y Van Dongen, Jos. "Pentaho Solutions. Business Intelligence and Data Warehousing with Pentaho and MySQL ". <http://www.wiley.com/compbooks>.
-  Bustos, Jorge. Business Intelligence y Data Warehousing en Win. Equipo Técnico Danysoft. [Citado el: 17 de Julio de 2005.] [j.bustos@danysoft.com](mailto:j.bustos@danysoft.com). <http://www.danysoft>.
-  Cerezal Tamargo, Lourdes. Integración de Procesos y Datos para el Diseño de un Data Warehouse. Tesis para optar por el título de Máster en Ciencia de la Computación. Universidad de La Habana, año 2003.
-  Colectivo de autores: Juan Ángel Enríquez de Salamanca de la Fuente, Tomás Martínez Ruiz, Almudena Quevedo Navarro, M<sup>a</sup> Carmen Rafael Bastante, Pedro Reales Mateo y Laura Sánchez González. "Funcionalidad 4: Almacenes de Datos y Bases de Datos XML". Universidad de Castilla La Mancha Escuela Superior de Informática, 2007.
-  Colectivo de autores: Hurtado Torres, M. Visitación, Abad Grau, M. Mar, Hornos Barranco, Miguel J. y Montes Soldado, Rosana. Bases de Datos y Data Warehouse: Herramientas estratégicas para la eficacia comercial. Departamento de Lenguajes y Sistemas Informáticos. Facultad de Ciencias Económicas y Empresariales. Universidad de Granada.
-  Colectivo de autores: Trujillo, Juan Carlos. Mazón, José Norberto y Pardillo, Jesús. Diseño y Explotación de Almacenes de Datos. Conceptos Básicos de Modelado Multidimensional. Editorial Club Universitario. ISBN: 978-84-8454-902-4.
-  Conesa Caralt, Jordi y Curto Díaz, Josep. Introducción al Bussines Intelligence. s.l.: UOC - EDIUOC, 2010.
-  Cuba en cifras Objetivos de Desarrollo del Milenio. ONEI. Edición 2009. ISBN: 978-959-7119-46-9.

- 📖 ¿En qué se clasifican las bases de datos? [Online] [Cited: 01 10, 2011.] [http://www.netronycs.com/clasificacion\\_de\\_base\\_datos.html](http://www.netronycs.com/clasificacion_de_base_datos.html).
- 📖 Evaluación en cifras Objetivos de Desarrollo del Milenio a nivel local. Una mirada desde la región oriental de Cuba. ONEI. Imprenta Alejo Carpentier. ISBN: 978-959-7119-46-3.
- 📖 Geiger, Claudia Imhoff. Nicholas Galembo. Jonathan G. Mastering Data Warehouse Design, Relational and Dimensional Techniques. 2003.
- 📖 González Hernández, Yanisbel "Propuesta de Metodología para el desarrollo de Almacenes de Datos en DATEC". Centro de Tecnologías de Gestión de Datos (DATEC). La Habana UCI, septiembre de 2011.
- 📖 Inmon, William H. (1996). Building the Data Warehouse, Wiley Computer, New York.
- 📖 Jorge Oduardo, Yanisleidis. Sistema de Información de Gobierno. Mercado de Datos para las series históricas del Sector Externo. Trabajo de Diploma para optar de por el título Ingeniero en Ciencias Informáticas. La Habana.UCI, 2011.
- 📖 Kimball vs Inmon. Ampliación de conceptos del Modelado Dimensional. [En línea] [Citado el: 13 de noviembre del 2011] <http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimencional/>
- 📖 La metodología. [Online] [Cited: 01 10, 2011.] <http://www.ecured.cu/index.php/Metodolog%C3%ADa>.
- 📖 Lluís Cano, Joseph. "Business Intelligence: Competir con Información". ESADE.
- 📖 Manual para la construcción de un Data Warehouse. Instituto Nacional de Estadística e Informática INEI Lima, año 1997.
- 📖 Manual para la adquisición de un sistema de Data Warehouse. *Dataprix*. [En línea] [Citado el: 18 de Noviembre de 2010.] <http://www.dataprix.com/olap-rolap-molap>
- 📖 Marqués Andrés, María Mercedes. Modelos de datos. Universitat Jaume-I. [Citado el: 5 de Abril de 2011] <http://www3.uji.es/~mmarques/f47/apun/node32>.
- 📖 Medina, Doris. Modelo de Auditoría y Metadatos para la Gestión de Proceso de Integración de Datos. La Habana, Cuba, UCIENCIA 2012.
- 📖 Mundy, Joy. Kimball University: Building a Foundation for Smart Applications. InformationWeek. [En línea] 1 de Diciembre de 2006. [Citado el: 20 de Octubre de 2010.] [http://www.kimballgroup.com/html/articles\\_search/articles2006/0612IE.html](http://www.kimballgroup.com/html/articles_search/articles2006/0612IE.html).
- 📖 Pérez Leyva, Delmis Arianna. García Devesa, Maryin. Sistema de Información de Gobierno. Mercado de Datos Comercio Interior. Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas. La Habana.UCI, 2011.

- 📖 Pérez Nieblas, Leonel. Sistema de Información de Gobierno. Mercado de Datos para el área de Construcción. Trabajo de Diploma para optar de por el título Ingeniero en Ciencias Informáticas. La Habana.UCI, 2011.
- 📖 ¿Por qué Business Intelligence? [En línea]  
[http://www.sinnexus.com/business\\_intelligence/sistemas\\_informacion.aspx](http://www.sinnexus.com/business_intelligence/sistemas_informacion.aspx)
- 📖 ¿Qué es Business Intelligence? [En línea] [http://www.sinnexus.com/business\\_intelligence/](http://www.sinnexus.com/business_intelligence/)
- 📖 R, Elmasri y S. B, Navathe. Fundamentos de sistemas de bases de datos. s.l.: University of Texas at Arlington, 2002.
- 📖 Reportes OLAP: Opciones tecnológicas para quienes toman decisiones: [Online]  
<http://www.spss.com>
- 📖 Rodríguez Sanz, Miguel “ANÁLISIS Y DISEÑO DE UN DATA MART PARA EL SEGUIMIENTO ACADÉMICO DE ALUMNOS EN UN ENTORNO UNIVERSITARIO”. Proyecto Fin de Carrera. Universidad Carlos III de Madrid Escuela Politécnica Superior Ingeniería en Informática. 22 Julio, 2010. <http://es.scribd.com/doc/52203545/16/Metodologia-propuesta-por-Bill-Inmon>.
- 📖 Rodríguez Yunta, Luis. Bases de datos documentales: estructura y principios de uso. Universidad de Navarra. [En línea] 2001. [Citado el: 22 de Octubre de 2010.] [www.unav.es/dpp/documentacion/proteger/lryunta.pdf](http://www.unav.es/dpp/documentacion/proteger/lryunta.pdf).
- 📖 Rolap, Molap, Holap. [En línea] [Citado el: 2 de noviembre de 2010.] <http://www.csae.map.es/csi/silice/DW2251.html>
- 📖 Sinergia e Inteligencia de Negocio. [En línea] 18 de enero de 2007. [Citado el: 2010 de octubre de 25.] <http://www.sinnexus.com/empresa/creditos.aspx>.
- 📖 Silberschatz, Abraham, Korth, Henry F. y S., Sudarshan. Fundamentos de bases de datos. s.l.: MCGRAW-HILL, 1995.
- 📖 Solaz, Katrina. “OLAP es una necesidad del mercado y consecuencia de la madurez del mismo“.Directora General de Hyperion Solutions España. [Citado el: 4 de Junio de 1999].
- 📖 Wolf, Carmen. Implementado un DataWarehouse. [Citado el: 28 de febrero de 2002]. <http://www.inf.udec.cl/revista/edicion5/cwolff.htm>
- 📖 Zepeda Sánchez, Leopoldo Zenaido. Metodología para el Diseño Conceptual de Almacenes de Datos. Tesis de Doctorado. Universidad Politécnica de Valencia. Valencia: s.n., 2008. [En línea] junio de 2008. [Citado el: 23 de noviembre de 2010.] <http://tesis.com.es/autores/zepeda-sanchez-leopoldo-zenaido>.

## ANEXOS

Anexo1: CUI Presentar información de los indicadores de los objetivos cubanos del milenio

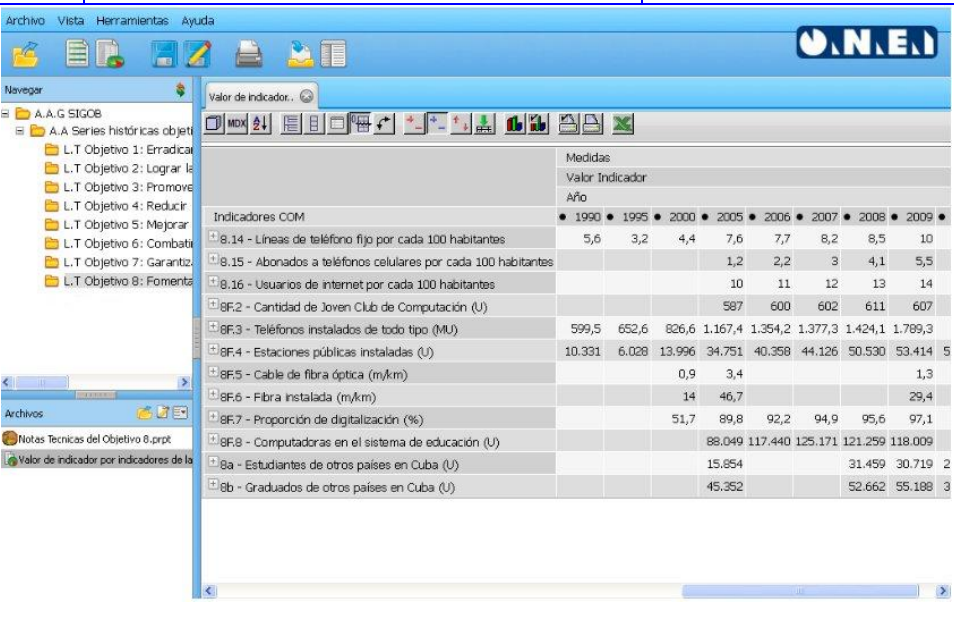
<b>Objetivo</b>	Analizar la información de los indicadores de los objetivos cubanos del milenio.	
<b>Actores</b>	Analista, Administrador del Sistema (Inicia) Presentar información de los indicadores de los objetivos cubanos del milenio	
<b>Resumen</b>	El CU inicia cuando el actor entra al sistema, selecciona uno de los reportes de los objetivos cubanos del milenio. El CU termina cuando se muestra el reporte seleccionado.	
<b>Complejidad</b>	Alta	
<b>Prioridad</b>	Media	
<b>Precondiciones</b>	El actor ha sido autenticado correctamente. Los datos correspondientes han sido cargados en el MD. Los reportes relacionados con los objetivos cubanos del milenio han sido creados.	
<b>Postcondiciones</b>	Los reportes correspondientes a los libros de trabajo del área de análisis A.A Series históricas objetivos cubanos del milenio fueron consultados por el actor.	
<b>Flujo de eventos</b>		
<b>Flujo básico &lt; Presentar información de los indicadores de la pobreza y el hambre &gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Selecciona el área de análisis general A.A.G SIGOB.	
2.		Muestra las áreas de análisis que están contenidas dentro del A.A.G SIGOB.
3.	Selecciona el área de análisis A.A Series históricas objetivos cubanos del milenio.	
4.		Muestra los libros de trabajo que están contenidos dentro del A.A Series históricas objetivos cubanos del milenio.

5.	<p>Selecciona el libro de trabajo que desee modificar:</p> <ul style="list-style-type: none"> <li>• L.T Objetivo 1: Erradicar la pobreza extrema y hambre. Ir a la Sección1: L.T Objetivo 1: Erradicar la pobreza extrema y hambre.</li> <li>• L.T Objetivo 2: Lograr la enseñanza primaria universal. Ir a la Sección2: LT Objetivo 2: Lograr la enseñanza primaria universal.</li> <li>• L.T Objetivo 3: Promover la igualdad de género y empoderamiento de la mujer. Ir a la Sección3: L.T Objetivo 3: Promover la igualdad de género y empoderamiento de la mujer.</li> <li>• L.T Objetivo 4: Reducir la mortalidad de los niños menores de 5 años. Ir a la Sección4: L.T Objetivo 4: Reducir la mortalidad de los niños menores de 5 años.</li> <li>• L.T Objetivo 5: Mejorar la salud materna. Ir a la Sección5: L.T Objetivo 5: Mejorar la salud materna.</li> <li>• L.T Objetivo 6: Combatir el VIH/SIDA, paludismo y otras enfermedades. Ir a la Sección6: L.T Objetivo 6: Combatir el VIH/SIDA, paludismo y otras enfermedades.</li> <li>• L.T Objetivo 7: Garantizar la sostenibilidad del medio ambiente. Ir a la Sección7: L.T Objetivo 7: Garantizar la sostenibilidad del medio ambiente.</li> <li>• L.T Objetivo 8: Fomentar una alianza mundial para el desarrollo. Ir a la Sección8: L.T Objetivo 8: Fomentar una alianza mundial para el desarrollo.</li> </ul>	
<b>Sección 1: “L.T Objetivo 1: Erradicar la pobreza extrema y hambre”</b>		
	<b>Actor</b>	<b>Sistema</b>
1.		Muestra los reportes contenidos en el libro de trabajo L.T Objetivo 1: Erradicar la pobreza extrema y hambre.
2.	Selecciona el reporte que desea analizar.	
3.		Muestra la información contenida en el reporte seleccionado y brinda opciones



		al actor para hacer cambios al reporte durante su análisis. Ir al CU Modificar reporte. Termina el CU.
<b>Sección 2: “L.T Objetivo 2: Lograr la enseñanza primaria universal”</b>		
	<b>Actor</b>	<b>Sistema</b>
1.		Muestra los reportes contenidos en el libro de trabajo L.T Objetivo 2: Lograr la enseñanza primaria universal.
2.	Selecciona el reporte que desea analizar.	
3.		Muestra la información contenida en el reporte seleccionado y brinda opciones al actor para hacer cambios al reporte durante su análisis. Ir al CU Modificar reporte. Termina el CU.
<b>Sección 3: “L.T Objetivo 3: Promover la igualdad de género y empoderamiento de la mujer”</b>		
	<b>Actor</b>	<b>Sistema</b>
1.		Muestra los reportes contenidos en el libro de trabajo L.T Objetivo 3: Promover la igualdad de género y empoderamiento de la mujer.
2.	Selecciona el reporte que desea analizar.	
3.		Muestra la información contenida en el reporte seleccionado y brinda opciones al actor para hacer cambios al reporte durante su análisis. Ir al CU Modificar reporte. Termina el CU.
<b>Sección 4: “L.T Objetivo 4: Reducir la mortalidad de los niños menores de 5 años”</b>		
	<b>Actor</b>	<b>Sistema</b>
1.		Muestra los reportes contenidos en el libro de trabajo L.T Objetivo 4: Reducir la mortalidad de los niños menores de 5 años.
2.	Selecciona el reporte que desea analizar.	
3.		Muestra la información contenida en el reporte seleccionado y brinda opciones al actor para hacer cambios al reporte durante su análisis. Ir al CU Modificar reporte. Termina el CU.
<b>Sección 5: “L.T Objetivo 5: Mejorar la salud materna”</b>		
	<b>Actor</b>	<b>Sistema</b>
1.		Muestra los reportes contenidos en el libro de trabajo L.T Objetivo 5: Mejorar la salud materna.

2.	Selecciona el reporte que desea analizar.	
3.		Muestra la información contenida en el reporte seleccionado y brinda opciones al actor para hacer cambios al reporte durante su análisis. Ir al CU Modificar reporte. Termina el CU.
<b>Sección 6: “L.T Objetivo 6: Combatir el VIH/SIDA, paludismo y otras enfermedades”</b>		
	<b>Actor</b>	<b>Sistema</b>
1.		Muestra los reportes contenidos en el libro de trabajo L.T Objetivo 6: Combatir el VIH/SIDA, paludismo y otras enfermedades.
2.	Selecciona el reporte que desea analizar.	
3.		Muestra la información contenida en el reporte seleccionado y brinda opciones al actor para hacer cambios al reporte durante su análisis. Ir al CU Modificar reporte. Termina el CU.
<b>Sección 7: “L.T Objetivo 7: Garantizar la sostenibilidad del medio ambiente”</b>		
	<b>Actor</b>	<b>Sistema</b>
1.		Muestra los reportes contenidos en el libro de trabajo L.T Objetivo 7: Garantizar la sostenibilidad del medio ambiente.
2.	Selecciona el reporte que desea analizar.	
3.		Muestra la información contenida en el reporte seleccionado y brinda opciones al actor para hacer cambios al reporte durante su análisis. Ir al CU Modificar reporte. Termina el CU.
<b>Sección 8: “L.T Objetivo 8: Fomentar una alianza mundial para el desarrollo”</b>		
	<b>Actor</b>	<b>Sistema</b>
1.		Muestra los reportes contenidos en el libro de trabajo L.T Objetivo 8: Fomentar una alianza mundial para el desarrollo.
2.	Selecciona el reporte que desea analizar.	
3.		Muestra la información contenida en el reporte seleccionado y brinda opciones al actor para hacer cambios al reporte durante su análisis. Ir al CU Modificar reporte. Termina el CU.
<b>Flujos alternos</b>		

2 <Error en los datos>		
	Actor	Sistema
2ª.	Se autentifica Se entra al sistema Se selecciona el A.A Series históricas objetivos cubanos del milenio Se selecciona el L.T Objetivo 1: Pobreza extrema y hambre En la parte inferior izquierda se selecciona el reporte deseado En el área de trabajo se visualiza el reporte	Muestra un mensaje informando que los datos son incorrectos y vuelve al punto 1 del flujo normal de eventos.
Opciones de reportes de < Presentar información de los indicadores de la pobreza y el hambre >		
Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad
Variables de entrada relacionadas con el CU <Presentar información de los indicadores de la pobreza extrema y el hambre>: <ul style="list-style-type: none"> <li>• Año</li> <li>• Indicadores_COM</li> </ul>	Variables de salida disponibles en el hecho < Presentar información de los indicadores de la pobreza extrema y el hambre >: <ul style="list-style-type: none"> <li>• valor_indicador</li> </ul>	Rango de tiempo en que se solicitan las variables de salida: <ul style="list-style-type: none"> <li>• Anual</li> </ul>
Prototipo de interfaz de usuario	 <p>The screenshot shows the O.N.E.I. application window. On the left is a navigation tree with folders for 'A.A.G SIGOB' and 'A.A Series históricas objeti'. The main area displays a table titled 'Medidas' with columns for 'Valor Indicador' and 'Año' (1990-2009). The table lists various indicators such as 'Líneas de teléfono fijo por cada 100 habitantes', 'Abonados a teléfonos celulares por cada 100 habitantes', and 'Estaciones públicas instaladas (U)'. A 'Valor de indicador...' dialog box is open over the table.</p>	
Relaciones	CU Incluidos	
	CU Extendidos	Modificar reporte

<b>Requisitos no funcionales</b>	<p>Los siguientes RNF:</p> <ul style="list-style-type: none"> <li>• RNF1</li> <li>• RNF2</li> <li>• RNF4</li> <li>• RNF7</li> <li>• RNF9</li> <li>• RNF10</li> <li>• RNF11</li> <li>• RNF12</li> <li>• RNF14</li> <li>• RNF15</li> <li>• RNF18</li> <li>• RNF19</li> </ul> <p>Del epígrafe # 3.2 del documento Especificación de Requisitos.</p>
<b>Asuntos pendientes</b>	

**Tabla 6. CUI Presentar información de los indicadores de los objetivos cubanos del milenio**

Anexo2: Lista de chequeo

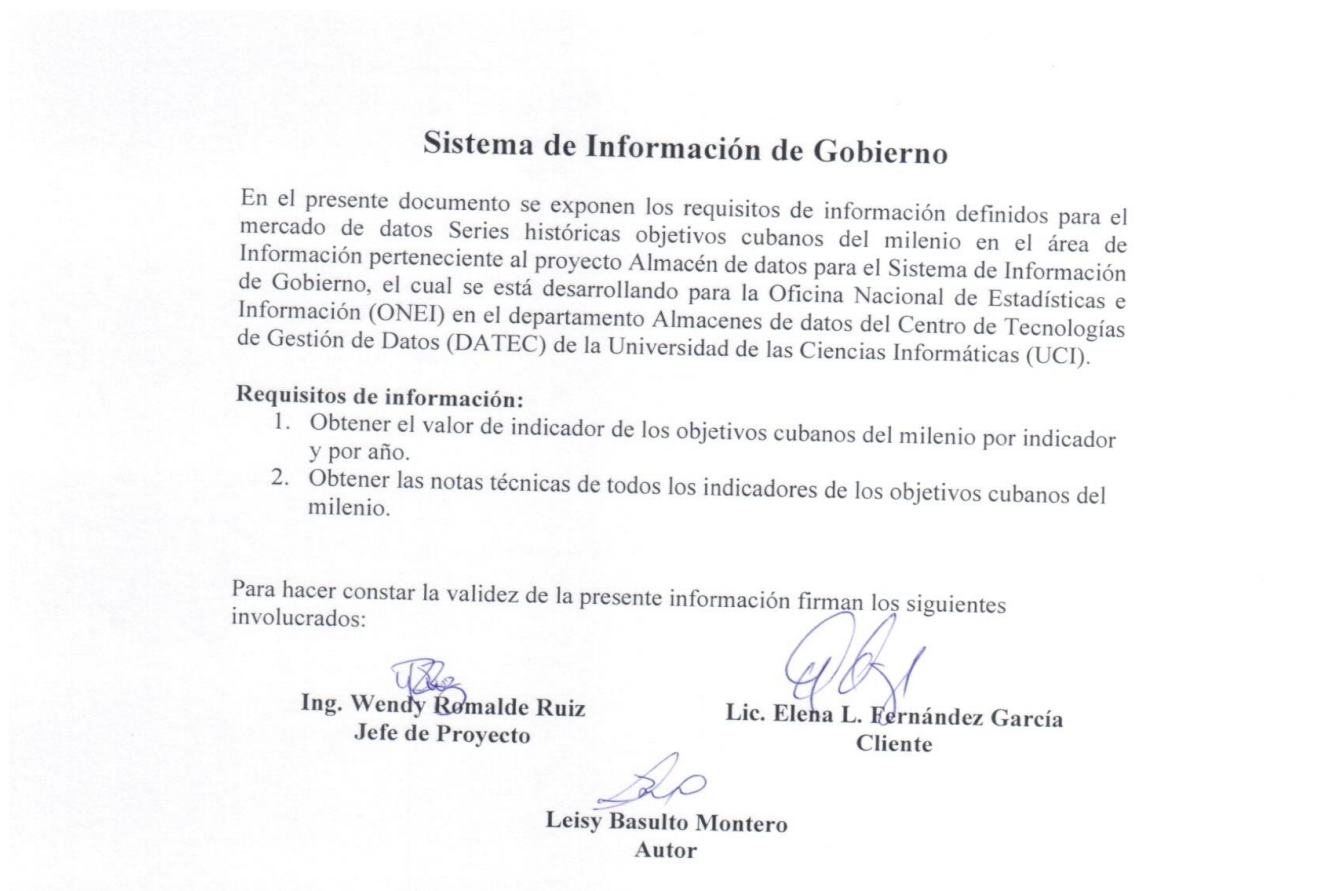
<b>Estructura del documento</b>					
<b>Peso</b>	<b>Indicadores a evaluar</b>	<b>Eval</b>	<b>(NP)</b>	<b>Cantidad de elementos afectados</b>	<b>Comentarios</b>
crítico	1. ¿El entregable contiene las secciones obligatorias de la plantilla estándar definida para el expediente de proyecto?				
crítico	2. ¿El alcance del proyecto describe correctamente los datos de las dimensiones y hechos del mercado de datos?				
crítico	3. ¿El objetivo expresa correctamente el propósito del documento?				
	4. ¿Se hace un uso adecuado del control del documento?				

	5. ¿En la sección de acrónimos se definen todos los acrónimos utilizados en el documento?				
	6. ¿En el entregable, la definición de las variables se hace correctamente?				
	7. ¿Existe una adecuada correspondencia entre las variables definidas y las descripciones que tienen estas variables?				
	8. ¿En el entregable se crea una hoja por cada variable definida?				
	9. ¿Queda registrado en el entregable todos los posibles valores que van a tener las variables definidas?				
<b>Indicadores definidos en el desarrollo</b>					
<b>Peso</b>	<b>Indicadores a evaluar</b>	<b>Eval</b>	<b>(NP)</b>	<b>Cantidad de elementos afectados</b>	<b>Comentarios</b>
	1. ¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?				
<b>Semántica del documento</b>					
<b>Peso</b>	<b>Indicadores a evaluar</b>	<b>Eval</b>	<b>(NP)</b>	<b>Cantidad de elementos afectados</b>	<b>Comentarios</b>
crítico	1. ¿Se ha identificado errores ortográficos en el entregable?				
crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?				

	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?				
--	--	--	--	--	--

**Tabla 7. Lista de chequeo**

Anexo3: Requisitos de información



**Figura 24. Requisitos de información**

## Anexo4: Acta de aceptación del cliente

### ACTA DE ACEPTACIÓN

En La Habana, a los 4 días del mes de junio del 2012

De una parte, la Oficina Nacional de Estadísticas e Información (ONEI) de Cuba, representado en este acto por Elena Leonila Fernández García, quien a los fines y efectos derivados del presente documento se denominará como "El cliente", y de otra Parte, el centro de Tecnologías de Gestión de Datos, conocido de forma abreviada como DATEC de la Universidad de las Ciencias Informáticas (UCI), representada en este acto por José Salvador Bermúdez Rodríguez, que a los fines y efectos derivados del presente documento se denominará DATEC.

**Primero:** Que en cumplimiento de los acuerdos, han sido efectuadas las actividades que se describen, **Las partes** DECLARAN:

**CONSIDERANDO:** Que se han efectuado las actividades siguientes:

1. – Diseño del mercado de datos: Series históricas objetivos cubanos del milenio en el área de Información.
2. – Proceso de extracción, transformación y carga de los datos.
3. – Implementación de las vistas de análisis OLAP.

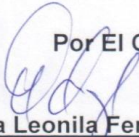
**CONSIDERANDO:** Que las actividades realizadas han sido desarrolladas con la calidad requerida y bajo las condiciones pactadas y aprobadas por **Las Partes**.

**CONSIDERANDO:** Que las actividades que se han ejecutado cumplen con los requerimientos de **El Cliente**.

**CONSIDERANDO:** Que DATEC ha entregado la documentación que avala la ejecución de este acto al **El Cliente**.

**POR TANTO:** **Las Partes** acuerdan formalizar mediante la presente Acta, Aceptadas las actividades que han sido ejecutadas en esta fecha.

Y para que así conste, se extiende la presente Acta en dos (2) ejemplares, rubricados por **Las Partes**.

  
Por El Cliente  
Elena Leonila Fernández García  
Nombre y Apellidos


  
Por DATEC  
José Salvador Bermúdez Rodríguez  
Nombre y Apellidos

Figura 25. Acta de aceptación



## GLOSARIO

Ad hoc: se utiliza en informática para referirse a consultas en bases de datos *ad hoc querying* o *ad hoc reporting*. Esto implica que el sistema permite al usuario personalizar una consulta en tiempo real, en vez de estar atado a las consultas prediseñadas para informes. Generalmente las consultas *ad hoc* permiten a los usuarios con poca experiencia en SQL tener el mismo acceso a la información de la BD, para esto los sistemas que soportan *ad hoc* poseen GUIs para generarlas.

Minería de datos: es el proceso de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente, estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiados datos.

No conformidad: defecto, error o sugerencia que se le hace al equipo de desarrollo una vez encontrada alguna dificultad en lo que se está evaluando.

Repositorio: es un término utilizado en el dominio de las herramientas CASE. El repositorio podría definirse como la BD fundamental para el diseño; no solo guarda datos, sino también algoritmos de diseño y, en general, elementos necesarios para llevar a cabo nuestro trabajo.

ERP: Son sistemas de información gerenciales que integran y controlan negocios asociados con las operaciones de producción y de los aspectos de distribución de una compañía en la producción de bienes o servicios. Estos sistemas básicamente manejan de forma modular muchas actividades de negocios como: la producción, logística, distribución, ventas, inventarios, envíos, facturas, contabilidad y administración de recursos humanos de la compañía. Un ejemplo de su aplicación en Cuba lo constituye el módulo recursos humanos y contabilidad implementada en la UCI.