

**Universidad de las Ciencias Informáticas**  
**Facultad 6**



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

**Título:** Mercado de Datos Contabilidad para el Sistema de Información de Gobierno.

Autor:

Irela González Piñera

Tutores:

Ing. Idalmys Maza Capote

Ing. Ariel Manresa Sánchez

La Habana, Junio 2013

Año 55 de la Revolución.



*Bueno es ir a la lucha con determinación, abrazar la vida y vivir con pasión.  
Perder con clase y vencer con osadía, porque el mundo pertenece a quien se atreve y  
la vida es mucho más para ser insignificante.*

*Charles Chaplin*

**DECLARACIÓN DE AUTORÍA**

Declaramos ser autores del presente trabajo “Mercado de datos Contabilidad para el Sistema de Información de Gobierno” y reconocemos a la Universidad de las Ciencias Informáticas (UCI) los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmamos la presente a los \_\_\_\_ días del mes de \_\_\_\_\_ del año 2013.

Irela González Piñera

\_\_\_\_\_

Firma de la Autora Firma del Autor

Ing. Ariel Manresa Sánchez

\_\_\_\_\_

Firma del Tutor

Ing. Idalmys Maza Capote

\_\_\_\_\_

Firma de la Tutora

### **DATOS DE CONTACTO:**

#### **Tutores:**

Ing. Ariel Manresa Sánchez

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Categoría docente:

Categoría científica:

Años de experiencia en el tema: 1

Años de graduado: 2

Correo Electrónico: [amanresas@uci.cu](mailto:amanresas@uci.cu)

Ing. Idalmys Maza Capote.

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Categoría docente:

Categoría científica:

Años de experiencia en el tema: 3

Años de graduado: 3

Correo Electrónico: [idalmys@uci.cu](mailto:idalmys@uci.cu)

*Primeramente quisiera agradecer al los seres más importantes en mi vida, a mi madre y padre queridos por todos los años de sacrificio, entrega y de amor para que yo esté hoy aquí cumpliendo sus sueños de ser una profesional.*

*A mi hermanas que las quiero con el alma, por todos sus buenos consejos y por su apoyo en todos los momentos difíciles, por estar siempre a mi lado.*

*A mi querido amor que no por mencionarlo de tercero deja de ser menos importante, al contrario, es junto con mi familia la persona más primordial en mi vida. A ti Arodys, te agradezco todo el apoyo incondicional y todo el amor que me has dado, sinceramente y desde lo más profundo de mi corazón te agradezco, yo mejor que nadie se que sin ti no hubiera logrado llegar hasta aquí y que para ti también es un sueño hecho realidad este momento. Te agradezco por estar siempre para mí y nunca decir que no cuando te he necesitado, gracias por guiarme durante todo este tiempo, por soportarme y sobre todo gracias por amarme.*

*A lo largo de estos cinco años he conocido a personas muy especiales, a las que en un momento como este es imposible olvidar porque han dejado huella en mi corazón, mis queridos amigos, de los cuáles he aprendido siempre, esas personas que han estado para mí cuando más he necesitado y nunca han sabido decir que no, para ustedes mis más entrañables agradecimientos, René, Yurima, Narín, Yanet, Rosy, Haymee, Leidanis, Yanelis, Manolo, Migue, Tomás, Dieter, Luis ángel “mi negro querido”, Chávez, Yeny, Yasma.*

*A mi otra mamá (Milena) como cariñosamente le digo, por todos sus buenos consejos, por ayudarme y quererme tanto.*

*A quien considero mi hermanito porque no tengo, mi querido amigo Adonis, gracias por guiarme siempre, ayudarme estar pendiente de mí, los momentos difíciles hicieron que nuestra amistad se hiciera más fuerte, los buenos nos permitieron disfrutarla.*

*A Ana del Carmen “Anita”, mis más querida amiga que ha estado conmigo durante estos cinco años, una de las personas más intensas que he tenido el placer de conocer, que siempre, siempre me ha brindado su confianza, su cariño.*

*A mis tutores por ayudarme y guiarme especialmente a Manresa por ser más que mi tutor, mi amigo.*

*A todo el colectivo de profesores que he tenido durante estos cinco años, gracias por ayudarme y formarme, a Wendy gracias por ser tan exigente, a Ismael, Leonel, Monchy, Leo, por ayudarme siempre que los ocupé.*

*A todos muchas gracias....*

*A mis padres, mis hermanas y a mis sobrinas queridas por ser mi razón de existir, por creer en mí y por brindarme todo su amor siempre.*

*A mi novio por su apoyo, cariño y por ser una persona tan extraordinaria.*

*A mis amigos por estar siempre pendientes de mí.*

## Resumen

La principal misión de la Oficina Nacional de Estadísticas e Información es dirigir metodológicamente la gestión de la información relevante para el gobierno y responder en particular por la gestión de la información y los documentos de interés para el gobierno central incluyendo la estadística oficial del país. En el área de Contabilidad la información es almacenada mediante tres modelos estadísticos a través del sistema Integrado de Gestión Estadística (SIGE). La solución del presente trabajo de diploma enmarca su desarrollo en los Mercados de Datos siendo estos de gran utilidad para realizar el análisis de la información correspondiente al área anteriormente mencionada. En el proceso de desarrollo se utilizó la metodología realizada por el Departamento de Almacenes de Datos que toma como base la metodología de Kimball y se ajusta a las condiciones y características de producción de la Universidad de la Ciencias Informáticas y del Centro de Tecnología de Gestión de Datos. Apoyándose en las herramientas Visual Paradigm, PostgreSQL, Pentaho Data Integration, DataCleaner, Pentaho Schema Workbench, Pentaho BI Server. Como resultado se obtiene la estructura del modelo de datos dimensional que comprende: las tablas de dimensiones, hechos y las medidas necesarias para la visualización de los datos. Se definen e implementan los mecanismos de extracción, transformación y carga de los datos, además del desarrollo de la capa de Inteligencia de Negocio (BI) para la visualización. Se realizaron pruebas a la aplicación para lograr la aceptación del cliente y verificar su calidad.

## Palabras Claves:

Almacenes de Datos, Mercado de Datos, Inteligencia de Negocio, Contabilidad.

## Índice de Contenido

Introducción .....	1
Capítulo 1: Fundamentación Teórica .....	6
1.1 Contabilidad.....	6
1.2 Almacenes de Datos.....	6
1.3 Modelo Multidimensional .....	9
Figura 1: Esquemas del modelo multidimensional .....	10
1.4 Modo de almacenamiento de datos OLAP.....	10
1.5 Mercado de Datos .....	11
1.6 Metodología para el Desarrollo de Almacenes de Datos.....	12
1.9 Herramienta de Modelado .....	15
1.9.1 Visual Paradigm 8.0.....	15
1.10 Sistema Gestor de BD .....	16
1.10.1 PostgreSQL 9.1 .....	16
1.10.2 pgAdmin III 1.14.1.....	18
1.11 Herramientas para los procesos de Extracción, Transformación y Carga .....	19
1.11.1 DataCleaner 1.5.4.....	19
1.11.2 Pentaho Data Integration 4.2.1 .....	20
1.12 Herramientas para el proceso de Inteligencia de negocios (BI) .....	21
1.12.1 Pentaho Schema Workbench 3.2.0.....	21
1.12.2 Pentaho BI Server 3.10.....	21
1.13 Conclusiones del capítulo.....	22
Capítulo 2: Análisis y Diseño del Mercado de Datos Contabilidad .....	23
2.1 Descripción del Negocio .....	23
2.2 Especificación de Requisitos .....	24



2.2.2	Requisitos de Información .....	24
2.2.2	Requisitos Funcionales.....	24
2.2.3	Requisitos no Funcionales.....	25
2.3	Reglas del Negocio.....	26
2.4	Vistas de Análisis.....	26
2.5	Casos de Uso del Sistema (CUS).....	27
2.5.1	Diagrama de CUS.....	28
Figura 2.	Diagrama de CU del sistema. ....	28
2.5.2	Actores del Sistema.....	28
2.5.3	Especificación de Casos de Uso de Información .....	29
2.6	Arquitectura del Mercado de Datos.....	31
2.7	Diseño del Mercado de Datos.....	32
2.7.1	Diseño del Subsistema de Almacenamiento.....	32
2.7.2	Matriz Dimensional o Matriz Bus.....	34
2.7.3	Modelo de Datos.....	35
2.8	Políticas de Respaldo y Recuperación.....	37
2.9	Diseño del Subsistema de Integración.....	37
2.9.1	Diseño General de las Transformaciones .....	38
2.9.2	Estrategias Generales de Integración de Datos .....	39
2.10	Diseño del Subsistema de Visualización.....	40
2.10.1	Arquitectura de la Información .....	40
2.10.2	Diseño de las Vistas de Análisis .....	41
2.10.3	Diseño de los Cubos OLAP .....	42
2.11	Esquema de Seguridad .....	42
2.12	Conclusiones del Capítulo .....	44

Capítulo 3: Implementación del MD Contabilidad.....	45
3.1  Implementación del Subsistema de Almacenamiento .....	45
3.1.1  Estándares de Codificación .....	47
3.2  Implementación del Subsistema de Integración.....	48
3.2.2  Gestión de los Metadatos del Proceso de Integración .....	51
3.3  Implementación del Subsistema de Visualización.....	51
3.3.1  Implementación de la Capa de Visualización.....	51
3.1  Conclusiones del Capítulo .....	53
Capítulo 4: Pruebas .....	54
4.1  Calidad del Software.....	54
4.2  Pruebas .....	55
4.2.1  Casos de Prueba .....	56
4.2.2  Listas de Chequeo.....	57
4.3  Resultados de las Pruebas .....	59
4.4  Perfilado de los Datos.....	60
4.5  Conclusiones del Capítulo .....	61
Conclusiones .....	62
Recomendaciones .....	63
Referencias Bibliográficas.....	64
Bibliografía.....	67

## **Introducción**

Las tecnologías de la información contribuyen con el desarrollo socio-económico de un país por su vinculación con las diferentes esferas de la sociedad. Los avances alcanzados por estas tecnologías en la actualidad conforman un eslabón fundamental en el ámbito empresarial, específicamente en la contabilidad, disciplina que permite identificar, medir y comunicar la información económica.

Actualmente el análisis estadístico juega un papel decisivo en las organizaciones productivas y sociales, pues desde su surgimiento sus métodos y técnicas han ayudado a la realización de múltiples tareas. Este tipo de análisis es la base para la realización de estudios técnicos e investigaciones y permite la mejora de procesos de producción, de bienes y de servicios o el sustento de la toma de decisiones. Además facilita la supervisión y la evaluación de las actividades que se realizan en cuanto a la marcha de la economía nacional se refiere.

Cuba se encuentra inmersa en un proceso de informatización de todos sus sectores económicos y de mejora en sus sistemas informacionales en aras de alcanzar mayores niveles de productividad. Este proceso se ha implementado con el objetivo de mejorar el análisis de la información que realizan los directivos en estos sectores; y está basado principalmente en conocer el comportamiento de las organizaciones para lograr un avance satisfactorio de la gestión empresarial.

Existen diferentes entidades encargadas de la producción de aplicaciones informáticas para contribuir con la informatización de todo el país. Una de estas instituciones es la Universidad de las Ciencias Informáticas (UCI), la que tiene como misión principal producir aplicaciones y servicios informáticos, a partir de la vinculación estudio-trabajo como modelo de formación y servir de soporte a la industria cubana de la informática. La UCI cuenta con varios centros productivos entre los que se puede citar el Centro de Tecnologías de Gestión de Datos (DATEC). Este centro tiene como objetivo crear bienes y servicios informáticos relacionados con la gestión de datos, área del conocimiento que agrupa tanto a los sistemas de información, como a los denominados sistemas de inteligencia empresarial o de negocios.

El centro DATEC en conjunto con la Oficina Nacional de Estadística e Información (ONEI) está inmerso en la informatización de esta entidad. Dicho centro se encuentra trabajando en la creación de un Almacén de Datos (AD) que permita la realización de las tareas de los especialistas de esta institución en un menor tiempo y con mayor calidad, de forma que apoye el proceso de toma de decisiones por parte del Sistema de Información de Gobierno (SIGOB).

La ONEI es el órgano rector de la estadística en Cuba y se responsabiliza por la gestión e información de los principales indicadores de la actividad socio-económica del país. Su principal misión es recopilar mediante modelos estadísticos, información de todos los sectores de la economía y la sociedad cubana, entre los que se encuentran datos referentes a la esfera de la Contabilidad. Los modelos estadísticos pertenecientes al área de Contabilidad recogen diferentes indicadores entre los que se encuentran: ventas netas de bienes y servicios, depreciación de activos fijos tangibles, subvenciones, inventarios, utilidades y pérdidas, productividad y gastos financieros. Toda esta información es recopilada por los centros informantes con periodicidad mensual, trimestral y semestral, lo que ocasiona que se incremente y acumule en gran medida por lo que resulta prácticamente imposible analizarla en un corto período de tiempo.

Para la reforma de esta área se realizó el Mercado de Datos (MD) “Contabilidad, el Trabajo y los Salarios” que tiene como objetivo mitigar los elementos negativos que entorpecen el trabajo de los especialistas. El MD se nutría de información procedente de varios modelos estadísticos, los cuales se encontraban almacenados en archivos de diferentes formatos, como documentos Excel, documentos Word, archivos DBF, formato duro (papel) y otros. Para lograr una correspondencia en el modo de almacenamiento de la información estadística recogida por las áreas de la ONEI, se crea el Sistema Integrado de Gestión Estadística (SIGE) que capta la información proveniente de los diferentes centros informantes, lo que trae consigo un cambio en la fuente de datos que almacenaba la información de la diferentes áreas de la ONEI.

Debido a que las estadísticas deben ir ajustándose a las nuevas necesidades de la economía, evaluando cada año el comportamiento de los indicadores críticos del sector económico del país y los indicadores contables varían constantemente, los modelos que sirvieron de base para el desarrollo del MD “Contabilidad, el Trabajo y los Salarios” están actualmente en proceso de cambio, pues se le añadieron nuevas aristas de análisis y nuevos aspectos lo que conlleva a completa re-estructuración del modelo de datos que se utilizó para su diseño. Todas estas dificultades unidas al cambio de la fuente de información transfieren que el MD nombrado con anterioridad ha quedado obsoleto.

Por la situación anteriormente descrita surge el siguiente **problema de la investigación**: *¿Cómo contribuir a la toma de decisiones en el área Contabilidad del Sistema de Información de Gobierno?*

Definiendo como **objeto de estudio**: los Almacenes de Datos, enmarcado en el **campo de acción**: Mercado de Datos para el área Contabilidad del Sistema de Información de Gobierno.

Para dar respuesta al problema de la investigación se plantea como **objetivo general**: Desarrollar el Mercado de Datos Contabilidad del Sistema de Información del Gobierno que contribuya a la toma de decisiones, desglosado en los siguientes **objetivos específicos**:

1. Fundamentar la selección de la metodología, herramientas y tecnologías a utilizar en el desarrollo de los AD.
2. Realizar el análisis y diseño del MD del área Contabilidad.
3. Implementar el MD del área Contabilidad.
4. Realizar pruebas al MD del área Contabilidad.

Para darle solución a los objetivos específicos anteriormente mencionados se realizó la siguiente planificación de las **tareas de investigación**:

1. Caracterización de la metodología, herramientas y tecnologías a utilizar en el desarrollo de AD, lo que contribuirá a determinar cuales se utilizaran para el desarrollo del MD Contabilidad.
2. Levantamiento de requisitos para determinar las necesidades de información del cliente.
3. Descripción de los casos de uso del MD para especificar cada una de las funcionalidades del sistema.
4. Definición de los hechos, las medidas y las dimensiones del MD para determinar los elementos que forman parte del modelo lógico de los datos.
5. Diseño del modelo de datos para identificar los elementos que forman parte del modelo físico de los datos.
6. Definición de la arquitectura del MD lo que permitirá identificar los principales subsistemas que componen la solución.
7. Implementación del modelo de datos para definir la estructura de la BD a la hora de realizar la cargar los datos.
8. Diseño del subsistema de integración para guiar la carga de los hechos y las dimensiones.

9. Diseño del subsistema de visualización que permita la definición de la capa de presentación y realizando el diseño de los cubos OLAP (Procesamiento Analítico en Línea).
10. Diseño de los casos de pruebas para identificar los elementos que deben estar disponibles en el MD una vez culminada la implementación.
11. Implementación del subsistema de integración para poblar el MD con los hechos y las medidas correspondientes.
12. Implementación del subsistema de visualización para mostrar las vistas de análisis a los usuarios finales.
13. Realizar pruebas al MD para garantizar la calidad de los procesos de integración y el correcto funcionamiento del sistema.

El documento está estructurado de la siguiente forma:

### **Capítulo 1: Fundamentación teórica.**

En este capítulo se realiza una descripción de los elementos más importantes de los AD y MD. Se exponen conceptos, características, ventajas y desventajas de su utilización; así como las herramientas, metodología y tecnologías existentes para el desarrollo de un MD. Se explican las diferentes etapas del desarrollo de un AD para su mejor entendimiento.

### **Capítulo 2: Análisis y diseño del MD Contabilidad.**

Análisis y diseño del MD Contabilidad. Se realiza un estudio preliminar del negocio, con el objetivo de comprender mejor los elementos que lo componen, se especifican las necesidades de información para identificar las dimensiones, los hechos y las medidas, se definen las vistas de análisis candidatos y el diseño de los procesos de integración.

### **Capítulo 3: Implementación del MD Contabilidad.**

Implementación del MD Contabilidad. Se implementa el proceso de integración de los datos, la capa de inteligencia del negocio, incluyendo la creación de las estructuras necesarias para la navegación y el análisis de los datos.

### **Capítulo 4: Pruebas.**

En este capítulo se hace referencia a las pruebas, las cuales se realizarán mediante la utilización de las listas de chequeo para determinar que los artefactos de los procesos de Extracción, Transformación y Carga tengan la calidad requerida, y mediante los casos de pruebas realizados por Casos de usos, para validar las vistas de análisis del MD.

## **Capítulo 1: Fundamentación Teórica**

### **Introducción.**

En este capítulo se realiza una descripción de los elementos más importantes de los AD y MD, así como la selección de la metodología, técnicas y herramientas que serán utilizadas para la construcción del MD. También se presentan las tecnologías, arquitectura, metodologías y herramientas que se tendrán en cuenta para desarrollar el MD Contabilidad.

### **1.1 Contabilidad**

La contabilidad es una técnica que se ocupa de registrar, clasificar y resumir operaciones mercantiles de una entidad con el fin de interpretar sus resultados, lo que permite orientar el curso de la economía mediante datos contables y estadísticos. Estos datos permiten conocer la estabilidad, las tendencias de las ventas, costos y gastos generales dentro de una empresa o institución. El propósito de la contabilidad es controlar los recursos monetarios de una empresa para la toma de decisiones en materia de inversión.

Un modelo estadístico es un formulario que recoge indicadores sobre una temática específica dentro de una empresa o entidad con el objetivo de controlar el comportamiento de los indicadores dados. En el caso de la Contabilidad se captarán indicadores a todas las empresas estatales así como empresas mixtas o sociedades mercantiles, independientemente de la sección, división y clase en que clasifique su actividad económica.

### **1.2 Almacenes de Datos**

Las empresas hoy día necesitan herramientas confiables que le permitan actuar de manera correcta en las operaciones que realiza y en las decisiones que deben tomar. Las decisiones tomadas deben ser rápidas y basadas en buenos cimientos, por lo que se necesita de hechos y cifras que deben ser manejados y analizados en el menor tiempo posible debido a que la competencia en los negocios crece rápidamente. El poseer conocimientos correctos significa tener respuestas correctas y trazarse estrategias en beneficio de la empresa.

Existen disímiles tareas como recolectar, procesar, limpiar y transformar la información, las cuales juegan un papel fundamental en el proceso de toma de decisiones y se pueden considerar engorrosas



si se tiene en cuenta que las empresas cuentan con diferentes áreas. Uno de los mecanismos de la inteligencia de negocios que vienen a resolver el problema del análisis y manejo de la información son los AD.

### **Definición de Almacenes de Datos.**

Existen múltiples conceptos de los AD dado por varios autores. En la actualidad los autores más reconocidos en este tema son Ralph Kimball y Bill Inmon, este último considerado como el padre de este término.

Según Ralph Kimball un AD “es una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis”, también determinó que un AD no es más que: “la unión de todos los MD de una entidad. (2)

Según Bill Inmon un AD “es una colección de datos orientada al dominio, integrada, variante en el tiempo y no volátil, diseñada para dar apoyo al proceso de toma de decisiones en una organización. (3)

En la investigación se asume como concepto de AD a la unión de varios MD que contienen información de las principales áreas de una empresa, obteniendo dicha información de una o disímiles fuentes, la que posteriormente será analizada desde diferentes perspectivas permitiendo el soporte al proceso de toma de decisiones.

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

Los AD poseen características únicas que lo hacen muy eficiente para la creación de sistemas que ayuden al proceso de toma de decisiones. Generalmente contienen grandes volúmenes de información.

Entre sus características principales se destacan las siguientes:

- **Temático:** los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales y se centran en entidades de alto nivel, no en los procesos.(5)
- **Histórico:** los datos se organizan y almacenan en jerarquías en el tiempo, lo que permite análisis comparativos de estados actuales y de períodos anteriores. (5)
- **No volátil:** el almacén de datos puede ser leído pero no modificado. Es decir, se incorporan los últimos valores que tomarán las distintas variables contenidas en él, sin ningún tipo de acción sobre los valores que ya existían. (5)

- **Integrado:** como los datos almacenados provienen de fuentes diferentes deben integrarse en una estructura estable que elimine las inconsistencias existentes en los mismos.(5)

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

#### **Ventajas:**

- Facilitan el análisis de los datos en tiempo real.
- Los AD hacen más fácil el acceso a una gran variedad de datos.
- Integración de información procedente de múltiples sistemas externos.(6)

#### **Desventajas.**

- Pueden suponer altos gastos. Además de los consumos de mantenimiento que son muy elevados.
- Riesgo de fracaso en la construcción del sistema por cambios frecuentes en los requisitos de los usuarios.(6)

### **Actualidad de los Almacenes de Datos.**

Desde que las organizaciones comenzaron a guardar los datos de sus operaciones en medios de almacenamiento físico, con el fin de permitirles una mayor administración y control de la información, ha existido una necesidad de utilizarla para atender las necesidades propias de la organización y tomar mejores decisiones. En la actualidad la información juega un papel importante como el principal conocimiento que sostiene a las empresas, pero no basta simplemente tener la información, se debe saber cómo utilizarla, qué hacer con ella. La Inteligencia de Negocios surge a partir de esta información, con el objetivo de que se puedan generar escenarios, pronósticos y vistas de análisis que apoyen a la toma de decisiones, lo que se convierte en una aplicación perfecta para la tecnología de AD. (7)

A escala mundial existen diversos sistemas de AD entre los que se pueden citar:

- AD para el Análisis y Difusión de la Información Estadística del Turismo en España “DATUR. Contiene información estadística procedente fundamentalmente de sus tres fuentes principales:

Frontur, Egatur y Familitur, así como de la reelaboración de fuentes secundarias procedente de otros organismos de esa nación.(8)

- AD para la Oficina Nacional de Estadísticas de la República Dominicana. El Almacén de Datos integra 550 series estadísticas e indicadores con información económica, social, demográfica y de medio ambiente. Contiene anuarios, boletines, estudios especiales y permite el acceso a las principales Base de Datos (BD) disponibles para consulta en línea.(9)
- AD de SIGOB para la ONEI: contiene información estadística de todas las esferas socio-económicas de Cuba, cuenta con varios MD de diferentes áreas departamentales, entre ellas se pueden citar Turismo, Medio Ambiente y Contabilidad.

### 1.3 Modelo Multidimensional

Dadas las características de los AD es ideal la utilización en su diseño de un Modelo Multidimensional (MMD), el cual se sustenta del modelo entidad relación. Este tipo de diseño tiene como ventajas sobre el Modelo Entidad-Relación (MER), que es muy flexible, está desnormalizado y orientado a los intereses de un usuario final, aunque esto no significa que existan inconsistencias en los datos. (28)

El modelo multidimensional se representa a través de la definición de las tablas de hechos y dimensiones.

**Tablas de Hechos:** Las tablas de hechos son las tablas primarias en el modelo dimensional. Generalmente, almacenan medidas numéricas, las que representan valores de las dimensiones. La llave de la tabla de hecho, es una llave compuesta, debido a que se forma de la composición de las llaves primarias de las tablas dimensionales a las que está unida. Existen tablas de hechos que no contienen medidas, a estas tablas se les denomina tablas de hechos sin hechos, es decir, la relación entre las dimensiones que definen la llave de esta tabla de hecho implica por sí sola la ocurrencia de un evento. (10)

**Tablas de Dimensiones:** Contienen, generalmente, una llave simple y atributos que la describen. Las tablas de dimensiones especifican la organización lógica de los datos y proporcionan el medio para analizar el contexto del negocio. Cada tabla de dimensión tiene una clave principal que se corresponde con exactamente uno de los componentes de la clave compuesta de la tabla de hechos. En

dependencia del esquema de diseño que se asuma pueden contener llaves foráneas de otras tablas de dimensión.

Para el modelado de los datos en un AD existen tres variantes de modelación, las cuales se detallan a continuación:

- **Esquema Estrella:** es la técnica más común. Posee una sola tabla de hechos que se relaciona con todas las tablas de dimensiones.(Ver figura 1 a)
- **Esquema Copo de Nieve:** es una ampliación del modelo estrella, las dimensiones son normalizadas en dependencia de la jerarquía existente, pero la estructura de una única tabla de hechos se mantiene.(Ver figura 1 b)
- **Esquema constelación de hechos:** está compuesto por diversos esquemas de estrella, con la particularidad de que varias tablas de hechos comparten algunas tablas de dimensiones. (Ver figura 1c)

En la siguiente figura se muestran los esquemas para el modelado de datos en un AD.

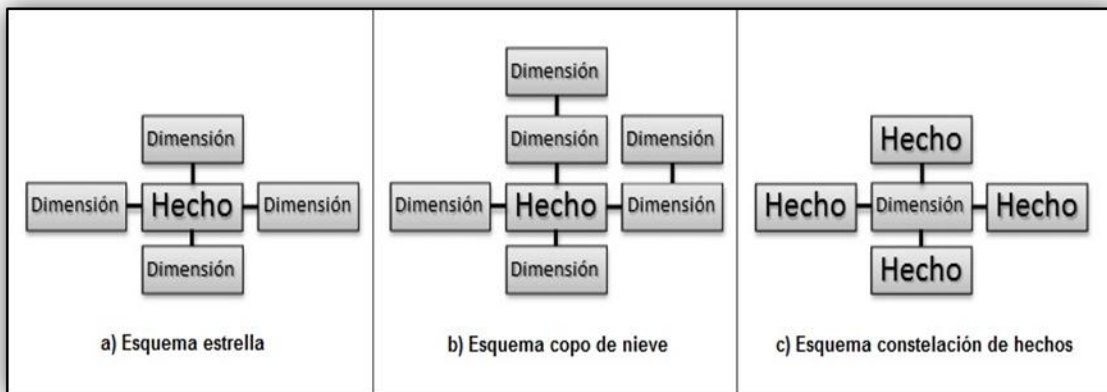


Figura 1: Esquemas del modelo multidimensional

## 1.4 Modo de almacenamiento de datos OLAP

La tecnología de Procesamiento Analítico en Línea (OLAP) permite a los usuarios analizar diferentes dimensiones de datos multidimensionales. Su objetivo es agilizar la consulta de grandes cantidades de

datos, lo que proporciona respuestas rápidas a consultas analíticas complejas e iterativas, utilizadas generalmente para sistemas de ayuda que contribuyan a la toma de decisiones. (11)

Existen tres modelos derivados de OLAP: Proceso analítico relacional en línea (ROLAP por sus siglas en inglés), Proceso analítico multidimensional en línea (MOLAP, por sus siglas en inglés) y Proceso analítico híbrido en línea (HOLAP, por sus siglas en inglés); su principal diferencia está dada por la forma de organizar y almacenar los datos.

**ROLAP:** Los datos son almacenados en filas y columnas de forma relacional. Este modelo presenta la información a los usuarios en forma de dimensiones de negocio. La principal ventaja de esta arquitectura es que permite el análisis de una gran cantidad de datos. (11)

**MOLAP:** Es una herramienta OLAP que accede a datos que no están almacenados en registros de tablas, sino que almacenan los datos en arreglos de varias dimensiones, llamados cubos. Su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente. (11)

**HOLAP:** Permite un análisis híbrido de la información. Lo mejor de las dos tecnologías anteriores. El análisis HOLAP ayuda a reducir costes de hardware ya que se necesita menos espacio en disco que en las BD relacionales. Además, la respuesta de las consultas sobre las BD 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. (11)

En la investigación se decide utilizar como modo de almacenamiento ROLAP, debido a las políticas del Centro DATEC y la Universidad en cuanto a la utilización de software libre, pues se utiliza como sistema gestor de BD PostgreSQL, que además de ser libre y poseer una serie de características, las cuales se detallan posteriormente, permite modelar BD relacionales, no así multidimensionales. ROLAP accede directamente a los datos del almacén, soporta técnicas de optimización de accesos, tales como particionado de los datos a nivel de aplicación, soporte a la desnormalización y uniones múltiples, para acelerar las consultas.

### 1.5 Mercado de Datos

Ralph Kimball define un MD como *“un conjunto flexible de datos, idealmente basado en el dato más atómico posible (granular) para ser extraído de las fuentes operacionales y presentado en un modelo*

*simétrico (dimensional), que es más resistente cuando se enfrentan con las más inesperadas consultas de los usuarios (...) (12)*

Tomando como referencia el concepto de Ralph Kimball se puede caracterizar un MD, como una colección de datos que contiene información de un área de negocio específica, se caracteriza por organizar la información de tal forma que se alcance el más alto nivel de detalle. Permite examinar los datos desde diferentes perspectivas, logrando así un mejor control de la información.

### **Características de los Mercados de Datos.**

- Contiene el grado de granularidad necesaria.
- Se centran en los requisitos de los usuarios asociados a un departamento o área de negocio concretos.
- Mejoran el rendimiento de las consultas al reducir el volumen de los datos que se recorren para responder. (13)

### **1.6 Metodología para el Desarrollo de Almacenes de Datos**

Para guiar el proceso de desarrollo de un software es preciso tener en cuenta la metodología a seguir. La misma se define como la guía que indica qué hacer y cómo actuar cuando se quiere obtener algún tipo de investigación. (14) Las metodologías de desarrollo precisan un conjunto de pasos y procesos a seguir que permiten planificar, estructurar y controlar el proceso de desarrollo de software.

Desde el surgimiento de las soluciones de AD y BI se han venido desarrollando simultáneamente las metodologías para el desarrollo y la implementación de este tipo de soluciones. A nivel internacional existen dos enfoques bien definidos para la creación de un AD: la metodología de Kimball y la metodología de Inmon.

Inmon en su metodología se basa en un enfoque descendente (top-down) proponiendo primero la creación del AD y a partir de este los MD que se nutrirán de él.

Inmon propone la construcción de un repositorio de datos corporativo como fuente de información consistente, consolidada, histórica y de calidad. Como el almacén de datos se construye descendentemente los mercados de datos se nutren del almacén corporativo, convirtiéndose en un complejo empresarial de BD relacionales. (15)

Por otra parte Kimball plantea la creación del almacén siguiendo un enfoque ascendente (bottom-up), creando primero los MD independientes por cada área departamental y el almacén de datos sería la unión de todos los MD.

### **Propuesta de Metodología para el Desarrollo de Almacenes de Datos en DATEC**

Para el desarrollo del MD Contabilidad se define como metodología a utilizar: Propuesta de metodología para el desarrollo de AD en DATEC, que toma como base el enfoque de Kimball y se ajusta a las condiciones y características de la producción del centro y la Universidad.

Se tomó como base la Metodología de Kimball por los siguientes elementos:

- Crea los conceptos de Hechos y Dimensiones, lo que indudablemente es muy eficaz en el proceso de la toma de decisiones y proporciona mayor agilidad en el proceso de desarrollo.
- Propone ir construyendo el AD a través de la construcción de los MD departamentales, lo que constituye una estrategia buena y coincide con la división lógica de las empresas, entidades, organismos, etc.
- Es una metodología madura y reconocida por el resto de la comunidad dedicada al tema. Tiene bien definidas las etapas, actividades, artefactos y roles. (15)

Durante el ciclo de vida de esta metodología se destacan las siguientes fases de desarrollo:

- **Estudio Preliminar o Planeación:** Se realiza el estudio de la entidad cliente, la planeación del proyecto, se definen los objetivos, el alcance preliminar, los costos estimados y otras actividades.
- **Requerimientos:** Se realiza en dos direcciones, una, mediante la identificación de las necesidades de información y reglas del negocio; y la otra con un levantamiento detallado de las fuentes de datos a integrar. Después se procede a la definición de los requerimientos.
- **Arquitectura y Diseño:** Se definen las estructuras de almacenamiento, se diseñan las reglas de extracción, transformación y carga, definiéndose la arquitectura de información que regirá el desarrollo de la solución.
- **Implementación:** Se diseña físicamente el repositorio de datos, se crean las estructuras de almacenamiento, el área temporal de almacenamiento, se ejecutan las reglas de Extracción,

transformación y Carga y se configuran e implementan las herramientas de inteligencia de negocios para la obtención de los elementos que se acordaron con el cliente final.

- **Prueba:** Se realizan las pruebas al sistema desde las Pruebas de Unidad hasta las de Aceptación con el cliente final.
- **Despliegue:** Se realiza un despliegue piloto en el cual se configuran los servidores y se instalan las herramientas y se carga una muestra de los datos para demostrar que el sistema funciona. Posterior a la aceptación del cliente se realiza la carga de los datos así como la Capacitación y Transferencia Tecnológica.
- **Soporte y Mantenimiento:** Tras la implantación de la solución se brindan los servicios de soporte en línea, vía telefónica, web u otras según el contrato firmado y las condiciones de soporte establecidas.
- **Gestión y Administración del Proyecto:** A lo largo del ciclo de vida se realizan actividades de control, gestión y chequeo del desarrollo, los gastos, las utilidades, los recursos y demás actividades por parte del Grupo de Dirección del Proyecto.(15)

El MD realizado en la investigación no transitará por todas las fases del ciclo de vida de la metodología a utilizar, solo llegará hasta la fase de pruebas. Atendiendo a que las dos fases posteriores a esta son ejecutadas por el personal del Departamento, y la Gestión y Administración del proyecto es desarrollada por los especialistas que pertenecen al grupo de dirección. (15)

### 1.7 Proceso de Integración

La integración de datos es usualmente entendido como el proceso que combina datos de diferentes fuentes para proveer una visión simple y comprensible de toda la información combinada.

Se requieren herramientas para la obtención de los datos a partir de uno o más sistemas fuentes (Extracción), cambiar el contenido y/o la forma de la información para hacerla coincidir con la estructura del AD (Transformación) e introducir los datos dentro del AD destino (Carga).

A continuación se explica el proceso de Extracción, Transformación y Carga (ETL).

- **Extracción:** en esta etapa es donde se obtiene la información proveniente de la fuente. En la gran mayoría de los proyectos estas fuentes pueden ser diferentes.



- **Transformación:** una vez extraída la información, como los datos pueden provenir de diferentes fuentes es necesario estandarizarlos para que puedan ser cargados en el AD. En esta etapa del proceso ETL se procesa la calidad de los datos, revisando valores válidos, consistencia, eliminación de valores redundantes.
- **Carga:** este proceso es el encargado de cargar al AD aquellos datos que han sido extraídos de las distintas fuentes y transformados. (16)

## 1.8 Proceso de Inteligencia de Negocio

Uno de los recursos más importantes de cualquier organización es la información, gracias a ella se pueden tomar decisiones que permiten mejorar el rumbo de las empresas y organizaciones. BI es el proceso de integración y tratamiento de los datos para convertirlos en información que permite apoyar a los tomadores de decisiones en la organización y hace referencia a las técnicas de análisis de datos destinados a encontrar información útil para la toma de decisiones. (17)

Actualmente, existe una gran diversidad de software de BI que pueden ser utilizados en las diferentes áreas departamentales de la empresa. Son muchas las empresas que se han favorecido con la implementación de un sistema de BI.

Las soluciones BI apoyan a los tomadores de decisiones con la información oportuna y de valor, lo que les permite tomar mejores decisiones de negocios. La información adecuada en el momento y en el lugar adecuado incrementa efectividad de cualquier empresa. Son múltiples los beneficios que ofrece a las instituciones, entre ellos se encuentra la generación de una ventaja competitiva, o de índole cognoscitiva en temas no empresariales.

## 1.9 Herramienta de Modelado

### 1.9.1 Visual Paradigm 8.0

Visual Paradigm es una herramienta que emplea Lenguaje Unificado de Modelado (UML por sus siglas en inglés) para modelar el negocio. Es fácil de instalar y actualizar, además de que puede ser utilizado durante todo el ciclo de vida del desarrollo de software. Modela los diagramas para las distintas fases del desarrollo de un producto: análisis, diseño e implementación. Permite la realización de ingeniería directa e ingeniería inversa hacia diversos lenguajes de programación y genera vistas de análisis en diferentes tipos de formatos. (18)

A continuación se muestran algunas de sus características:

- Disponibilidad en múltiples plataformas (Windows, Linux).
- Diseño centrado en casos de uso y enfocado al negocio que generan un software de mayor calidad.
- Uso de un lenguaje estándar común a todo el equipo de desarrollo que facilita la comunicación.
- Capacidades de ingeniería directa e inversa.
- Modelo y código que permanece sincronizado en todo el ciclo de desarrollo.
- Disponibilidad de múltiples versiones, para cada necesidad.
- Licencia: gratuita y comercial.
- Soporta aplicaciones Web.
- Varios idiomas.
- Generación de código para Java y exportación como HTML.
- Fácil de instalar y actualizar.

Se decidió utilizar Visual Paradigm 8.0 para UML porque es una herramienta que soporta todo el ciclo de vida del desarrollo de software. Además, brinda la posibilidad de modelar todo tipo de diagramas de clases, admite la compatibilidad entre ediciones, la documentación asociada a cada etapa del proceso de desarrollo y generar script para diferentes sistemas gestores de BD. Los desarrolladores lo utilizan para facilitar el modelado simultáneo, almacenar los archivos de proyectos y hacer un seguimiento de los cambios. La UCI cuenta con la licencia de uso que presenta esta herramienta

## 1.10 Sistema Gestor de BD

### 1.10.1 PostgreSQL 9.1

El software de BD en los últimos años ha experimentado un auge extraordinario dado por la progresiva informatización de casi la totalidad de las empresas de hoy día. Rapidez y efectividad en los procesos y los grandes flujos de información son las necesidades más indispensables a la hora de optimizar servicios y productos. Con la existencia de múltiples entornos de desarrollo y la notable demanda de soluciones informáticas, han surgido multitud de gestores de BD. (19)

Los SGBD son un tipo de software muy específico, dedicado a servir de interfaz entre la BD, el usuario y las aplicaciones que la utilizan y que prestan servicios para el desarrollo y el manejo de las mismas. Uno de los SGBD de código abierto que mayor auge ha alcanzado en los últimos tiempos y considerado uno de los gestores más completo es **PostgreSQL**, prestando mayor atención a que permite métodos almacenados, restricciones de integridad, vistas, etc.

Algunas Características de PostgreSQL 9.1:

- Atomicidad (Indivisible) es la propiedad que asegura que la operación se ha realizado o no, y por lo tanto ante un fallo del sistema no puede quedar a medias.
- Consistencia es la propiedad que asegura que sólo se empieza aquello que se puede acabar. Por lo tanto se ejecutan aquellas operaciones que no van a romper la reglas y directrices de integridad de la BD.
- Aislamiento es la propiedad que asegura que una operación no puede afectar a otras. Esto asegura que dos transacciones sobre la misma información nunca generará ningún tipo de error.
- Durabilidad es la propiedad que asegura que una vez realizada la operación, ésta persistirá y no se podrá deshacer aunque falle el sistema.
- Compatible con los principales sistemas operativos: Linux, Unix, Mac OS, etc.
- Soporte de todas las características de una BD profesional (triggers, funciones, secuencias, relaciones, reglas, tipos de datos definidos por usuarios, vistas, vistas materializadas, etc.)
- Utilidades para limpieza de la BD.
- Utilidades para análisis y optimización de consultas. (20)

A continuación se muestran las características principales que responden la decisión de tomar PostgreSQL como SGBD:

- Máximo de BD: ilimitado. (De acuerdo a la capacidad de almacenamiento del servidor)
- Máximo de tamaño de tabla: 32 TB.
- Máximo de tamaño de registro: 1.6 TB.

- Máximo de tamaño de campo: 1GB.
- Máximo de registros por tabla: ilimitado.
- Máximo de campos por tabla: 250 a 1600 (depende de los tipos usados).
- Máximo de índices por tabla: ilimitado.
- Compatible con los principales sistemas operativos.
- Adaptable a las necesidades propias de cada usuario.
- Soporta llaves foráneas, tipos de datos definidos por el usuario, secuencias, relaciones, uniones, vistas, reglas y procedimientos almacenados en múltiples lenguajes.(20)

### 1.10.2 pgAdmin III 1.14.1

pgAdmin es el software más utilizado de código abierto para la administración de BD. En pgAdmin se puede ver y trabajar con la mayoría de los objetos de la BD, examinar sus propiedades y realizar tareas administrativas. PgAdmin está diseñado para responder a las necesidades de todos los usuarios, desde escribir simples consultas SQL para crear BD complejas. La interfaz gráfica soporta todas las características de PostgreSQL y facilita la administración. (21)

Seguidamente se mencionan algunas de sus principales características:

- Multiplataforma.
- Diseñado para múltiples versiones de PostgreSQL y derivados.
- Acceso a los datos.
- El acceso a todos los objetos de PostgreSQL. (21)

Para la administración de la BD se decide utilizar pgAdmin III 1.14.0 porque es un motor de BD de código abierto muy avanzado. Es multiplataforma y también funciona con otros motores comerciales basados en PostgreSQL. Se diseña para responder a las necesidades de la mayoría de los usuarios, desde escribir simples consultas SQL hasta desarrollar BD complejas. La interfaz gráfica soporta todas las características de PostgreSQL y facilita la administración. Está disponible en más de una docena de lenguajes y para varios sistemas operativos, incluyendo Microsoft Windows, Linux, FreeBSD, Mac y

OSX.

## 1.11 Herramientas para los procesos de Extracción, Transformación y Carga

### 1.11.1 DataCleaner 1.5.4

Contar con datos de calidad, entre otras cosas, permite: suministrar datos precisos y completos que conlleven a mejores resultados y a un alto rendimiento aumentando la confianza del usuario, al garantizar que pasen más tiempo analizando los datos y menos tiempo comprobando la calidad de las fuentes de información; aumenta el conocimiento de los datos y permite conocer dónde se están produciendo los errores para corregirlos y construir mejores sistemas de información.

Danette McGilvray (2008) define el concepto de calidad de la información como “grado en que la información y los datos pueden ser una fuente confiable para cualquiera que requiera su uso; de forma completa, correcta, consistente, oportuna para las personas que toman decisiones, accesible a los clientes y dirigida a la organización para conseguir sus objetivos”.(22)

DataCleaner es un motor de procesamiento que fue construido para tareas altamente interactivas, de rendimiento y flexibilidad. Es una herramienta de análisis de calidad de la información que le permite realizar perfiles de datos y validación. El monitoreo es un aspecto central del DataCleaner para establecer el punto de partida, los objetivos, y para asegurar un proceso de seguimiento de las cuestiones de calidad de datos. Además encuentra las pautas, valores perdidos, juegos de caracteres y otras características de los valores de la información. (23)

A continuación se presentan las características más importantes de DataCleaner 1.5.4:

- Es posible crear reglas de validación para la entrada de los registros. Esta funcionalidad trabaja como una forma de asegurar la integridad y coherencia de los datos almacenados. Elabora perfiles de datos.
- Realiza la comparación de tablas, columnas y células con el fin de verificar la consistencia y veracidad de los datos.
- La personalización de algunas herramientas de acuerdo con la necesidad de la organización, sin olvidarse de la simplicidad.

- Consigue acceder a las BD más utilizadas en el mercado, incluyendo Oracle, Microsoft SQL Server, MySQL, PostgreSQL, OpenOffice (archivos con extensión ODB). Además, consigue interactuar con archivos en formato XML y planillas de Microsoft Excel. Es una aplicación open source (código abierto y licencia de uso gratuita).
- Mejora el flujo de informaciones y garantiza la calidad de datos esenciales para el funcionamiento de la organización.
- Es compatible con Windows XP, Windows Vista, Windows 7, Windows 2003, Linux y Mac OSX. (23)

### 1.11.2 Pentaho Data Integration 4.2.1

El uso de la herramienta de integración de datos PDI (conocida además como Kettle) proporciona una poderosa capacidad de Extracción, Transformación y Carga de datos usando una solución innovadora basada en metadatos. Esta herramienta provee un ambiente de diseño intuitivo y es extremadamente versátil, ya que se tienen bloques que permiten leer y escribir de cualquier BD, fichero excel, Access y otros que permiten operar con los campos renombrando, normalizando, calculando campos en función de otros, mapeando valores y realizando búsquedas auxiliares en BD (24). La herramienta de integración de datos PDI fue concebida para apoyar el desarrollo de soluciones de BI mediante metodologías ágiles, reduciendo y optimizando el ciclo de vida de aplicaciones BI al permitir avanzar de forma paralela en el diseño de las ETL, modelamiento y visualización de datos. PDI está compuesto por cuatro herramientas:

- **SPOON**: permite diseñar de forma gráfica las transformaciones ETL.
- **PAN**: ejecuta un conjunto de transformaciones diseñadas con SPOON.
- **CHEF**: permite diseñar la carga de datos incluyendo un control de estado de los trabajos.
- **KITCHEN**: permite ejecutar los trabajos *batch* diseñados con CHEF.

Algunas características:

- Librería de transformaciones completa con más de 100 objetos de mapeo.

- 100% Java, amplio soporte multiplataforma y soporte de una amplia cantidad de fuentes de datos, incluyendo aplicaciones integradas, sobre 30 plataformas propietarias y open source, archivos planos, documentos Excel, y más.
- Herramienta gráfica de muy fácil uso (control lógico de flujo).
- Basado en repositorio facilita re-uso de componentes de transformación, colaboración y administración de modelos, conexiones, logs, etc.
- Scheduler (Calendario programador de transformaciones y jobs).(25)

### 1.12 Herramientas para el proceso de Inteligencia de negocios (BI)

#### 1.12.1 Pentaho Schema Workbench 3.2.0

Esta herramienta de la suite Pentaho tiene como objetivo facilitar la tarea de diseño de cubos OLAP. Su sencilla interfaz permite modelar un XML con el diseño del cubo a través de opciones lógicas e intuitivas que no requieren de un manejo avanzado de este formato de archivo.

Dentro de sus características se destacan:

- Diseñador intuitivo de esquemas OLAP
- Permite crear, editar, actualizar y publicar esquemas OLAP para ser desplegados por aplicaciones de visualización Pentaho.
- Acelera de manera considerable la construcción e implementación de este tipo de soluciones. (26)

#### 1.12.2 Pentaho BI Server 3.10

La Plataforma Pentaho BI Server provee el soporte y la infraestructura necesarios para crear soluciones de BI a problemas de negocios. El marco proporciona los servicios básicos, incluidos autenticación, registro, auditoría, servicios web y motor de reglas. La plataforma también incluye un motor de solución que integra vistas de análisis, análisis, tableros de comandos y componentes de minería de datos. El diseño modular y arquitectura basada en plug-in permite a todos o parte de la plataforma para estar embebida en aplicaciones de terceros por los usuarios finales, así como fabricantes de equipos originales. (26)

La Plataforma Pentaho BI Server funciona como un sistema basado en administración web de

informes, 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. (26)

### **Ventajas:**

- Aplicación Java2EE 100% extensible, adaptable y configurable.
- Administra y programa vistas de análisis.
- Administra seguridad de usuarios.
- Se integra con la mayoría de entornos y se puede comunicar con otras aplicaciones vía webservices. (27)

### **1.13 Conclusiones del capítulo**

Una vez analizado el estado del arte de los AD y los MD con las principales características, ventajas e inconvenientes y elementos que conforman a cada uno de estos, se caracterizó la metodología, las técnicas y herramientas a utilizar en el desarrollo de la solución. Para el desarrollo de la investigación se decidió utilizar la metodología del departamento de AD ya que cubre las etapas por las que transita el desarrollo de un AD y brinda diversas ventajas que facilitan su construcción; también se adapta a las tendencias de la universidad teniendo como base la metodología propuesta por Kimball. Se utiliza además como modo de almacenamiento ROLAP debido a que el gestor de BD seleccionado es PostgreSQL y el mismo no soporta los modelos multidimensionales y trabaja solo con los relacionales. Para la realización del modelado se decide utilizar el Visual Paradigm en su versión 8.0. Se seleccionó como herramienta para la administración de los datos pgAdmin en su versión 1.14.1. Para desarrollar el proceso ETL se seleccionaron las herramientas Pentaho Data Integration en su versión 4.2.1 y DataCleaner en su versión 1.5.4 para el perfilado de datos. El uso de Pentaho Schema Workbench en su versión 3.2.0 y Pentaho BI Server en su versión 3.10 permitirán la implementación de la capa de visualización al MD Contabilidad con el objetivo de mostrar las vistas de análisis que servirán de apoyo a la toma de decisiones.



## **Capítulo 2: Análisis y Diseño del Mercado de Datos Contabilidad** **Introducción.**

En este capítulo se abordan una serie de elementos que hacen referencia al levantamiento de requisitos, la definición de las reglas del negocio, los requerimientos que el sistema debe cumplir y las necesidades de información de los usuarios para confeccionar el modelo dimensional, identificando las dimensiones, los hechos y las medidas. Además se describen los casos de uso y se define el diseño de los procesos de almacenamiento, integración y visualización.

### **2.1 Descripción del Negocio**

La ONEI es el órgano del Consejo de Estado encargado de dirigir metodológicamente la gestión de la información relevante para el gobierno y la aplicación de la política estatal en materia de estadística, responder, en particular por la gestión de la información y los documentos de interés nacional para el gobierno central, incluyendo la estadística oficial del país. Esta entidad cuenta con 22 áreas que se encargan del análisis estadístico de los diferentes sectores socioeconómicos del país. Una de estas áreas es Contabilidad, dedicada al análisis de los indicadores de origen contable a todas las uniones, empresas, organizaciones económicas estatales, granjas estatales, sociedades mercantiles cubanas, empresas mixtas y empresas de capital totalmente extranjero, así como todas las unidades presupuestadas.

Mantener un control sobre los indicadores contables es de gran importancia debido al impacto que estos pueden tener para la economía nacional. Generalmente en el área de Contabilidad se analizan los indicadores que se encuentran afectados, los que han venido trayendo pérdidas para la economía y los que mayor interés reporten en cuanto al aumento del capital del país.

La mayor parte de la información relacionada con la contabilidad es recogida por la ONEI a partir de tres modelos estadísticos fundamentales. El primero es el modelo 0005 que se encarga de recopilar datos relacionados con los indicadores generales, donde se informa mensualmente el comportamiento de estos indicadores en cada empresa, organismo o institución del país. El segundo modelo es el 5901 o conocido además como Indicadores seleccionados de la contabilidad donde se recogen datos acerca de la cuentas reales, cuantías nominales y otras informaciones, toda esta información es comunicada semestralmente. El tercer modelo es el 5903 o Cumplimiento del plan económico que capta los indicadores para el control del cumplimiento del plan en todas las entidades que realizan actividades mercantiles dentro de la economía nacional.

### **2.2 Especificación de Requisitos**

Los requisitos son definidos teniendo en cuenta las necesidades de los usuarios y de su área de trabajo, de los cuales depende la implementación de la solución. En el MD Contabilidad se definieron 13 requisitos de información, los que representan toda la información que debe mostrarse en la aplicación. Además se identificaron 9 requisitos funcionales y se especificaron los requisitos no funcionales. Todos estos requisitos se encuentran documentados en el expediente de proyecto en el artefacto Especificación de Requisitos de software. A continuación se muestran detalladamente varios ejemplos de los requisitos. Para mayor información dirigirse al expediente de proyecto el Artefacto especificación de requisitos "DATEC-SIGOB – Especificacion \_de\_ requisitos \_de\_ software\_Contabilidad.

#### **2.2.2 Requisitos de Información**

Los requisitos de información describen qué información debe almacenar el sistema para satisfacer las necesidades de clientes y usuarios. Los mismos deben estar disponibles para el usuario final a la hora de este realizar las consultas necesarias para analizar los datos, con el objetivo de apoyar la toma de decisiones en la institución.

**RI1.** Obtener el plan del año actual según el indicador, tiempo, Nomenclador de Actividades Económicas (NAE), División Político Administrativa (DPA), organismo.

**RI2.** Obtener el valor real del año actual según el indicador, tiempo, NAE, DPA, organismo.

#### **2.2.2 Requisitos Funcionales**

Los requisitos funcionales de un sistema describen la funcionalidad o los servicios que se espera que este provea para dar respuesta a los requerimientos de información. A continuación se muestran los requisitos funcionales identificados por cada Subsistema:

##### **Requisitos funcionales del Subsistema de Almacenamiento.**

RF1. Obtener información a partir del Sistema Integrado de Gestión Estadística (SIGE) a partir del 2011.

##### **Requisitos funcionales del Subsistema de Integración de Datos.**

RF2. Cargar información incremental correspondiente al modelo 0005 con una periodicidad mensual.

##### **Requisitos funcionales del Subsistema de Visualización.**

RF6. Mostrar información mediante vistas de análisis agrupadas en libros de trabajo.

### **2.2.3 Requisitos no Funcionales**

Los requerimientos no funcionales son propiedades o cualidades que el sistema debe cumplir para lograr el correcto funcionamiento de la aplicación. A continuación se muestran los requerimientos no funcionales identificados en el desarrollo del MD Contabilidad. Todos estos requerimientos se encuentran detallados en el Artefacto especificación de requisitos “DATEC-SIGOB – Especificacion \_de\_ requisitos \_de\_ software\_Contabilidad.

#### **Usabilidad:**

**RNF1** Agilizar el acceso a las vistas de análisis del almacén de datos mediante la distribución de la información por áreas de análisis.

El usuario podrá acceder de manera rápida a la información que solicita en el área correspondiente de acuerdo al objetivo de su solicitud.

#### **Confiabilidad:**

**RNF2** Garantizar la persistencia de la información.

Para garantizar la persistencia de la información se realizará un respaldo total de los datos del Almacén de Datos con una frecuencia semestral.

#### **Soporte:**

**RNF3** Lograr la homogeneidad de la estructura de los elementos definidos en el almacén.

Las estructuras del almacén de datos deben tener un nombre estándar teniendo en cuenta el tipo de estructura que sea.

#### **Restricciones de diseño:**

**RNF4** Utilizar el Sistema Gestor de Base de Datos definido durante la investigación.

El gestor de base de datos que se utilizará es PostgreSQL y como interfaz de administración de dicho gestor pgAdmin.

#### **Requerimientos para la documentación de usuarios en línea y ayuda del sistema:**

**RnF5.** Confección de un manual de usuario.

#### **Interfaz:**

### **Requerimientos legales y de derecho de autor.**

**RnF6.** Entregar el sistema a la ONEI.

El sistema debe ser transferido a la ONEI mediante un proceso de transferencia una vez que esté en explotación, incluyendo el código fuente y la documentación correspondiente

### **2.3 Reglas del Negocio**

Las reglas de negocios (o las directivas empresariales) definen y controlan la estructura, el funcionamiento y la estrategia de una organización, son la acción, práctica o procedimiento dentro de una determinada actividad o esfera. (29). Son además las condiciones que regulan un proceso de negocio, las cuales se deben seguir para cumplir con las necesidades de información de los clientes. Las reglas de negocio para el desarrollo del MD Contabilidad se agrupan en cuatro grupos: las reglas de variables, reglas de almacenamiento, reglas de transformación y reglas de visualización. Para mayor información dirigirse al expediente de proyecto el Artefacto especificación de requisitos "DATEC-SIGOB – Reglas\_de negocio \_y \_transformacion\_Contabilidad.

Para el desarrollo del MD Contabilidad se definieron las reglas de negocio que se muestran a continuación:

#### **Reglas de Variables:**

**RN1.** Los identificadores de los indicadores no pueden estar repetidos ni tomar valores nulos.

#### **Reglas de almacenamiento:**

**RN2.** Los valores de los indicadores del Modelo 5903 deben ser números con cifras decimales.

#### **Reglas de Visualización:**

**RN3.** En el modelo 5901 las columnas Saldo inicial año anterior y Saldo inicial año actual estarán bloqueadas desde la fila 9 a la 61.

### **2.4 Vistas de Análisis**

Mediante los requisitos de información anteriormente mencionados y por petición del cliente se identificaron nueve vistas de análisis los que son detallados a continuación.

1. Indicadores generales por organismo y tiempo.

2. Indicadores generales por NAE y tiempo.
3. Indicadores generales por División Político Administrativa (DPA) y tiempo.
4. Cumplimiento del plan económico por organismo y tiempo.
5. Cumplimiento del plan económico por NAE y tiempo.
6. Cumplimiento del plan económico por DPA y tiempo.
7. Indicadores seleccionados de la contabilidad por organismo y tiempo.
8. Indicadores seleccionados de la contabilidad por NAE y tiempo.
9. Indicadores seleccionados de la contabilidad por DPA y tiempo.

### **2.5 Casos de Uso del Sistema (CUS)**

Un caso de uso (CU) es una secuencia de acciones que realiza un sistema que produce un resultado observable de valor (un trabajo de salida) a un agente particular, que es alguien o algo fuera del sistema que interactúa con el mismo. Los CU demuestran la interacción del sistema con los agentes, incluidos los usuarios y otros sistemas. (30) Los CUS se dividen en dos categorías: los CU funcionales que agrupan los requisitos funcionales y los CU de información que agrupan los requisitos de información. (Figura 2)

Para lograr una mejor comprensión del funcionamiento del sistema, se realiza una representación gráfica que contiene las relaciones entre los casos de uso y sus actores. A continuación se muestra la figura del diagrama de CUS.

2.5.1 Diagrama de CUS

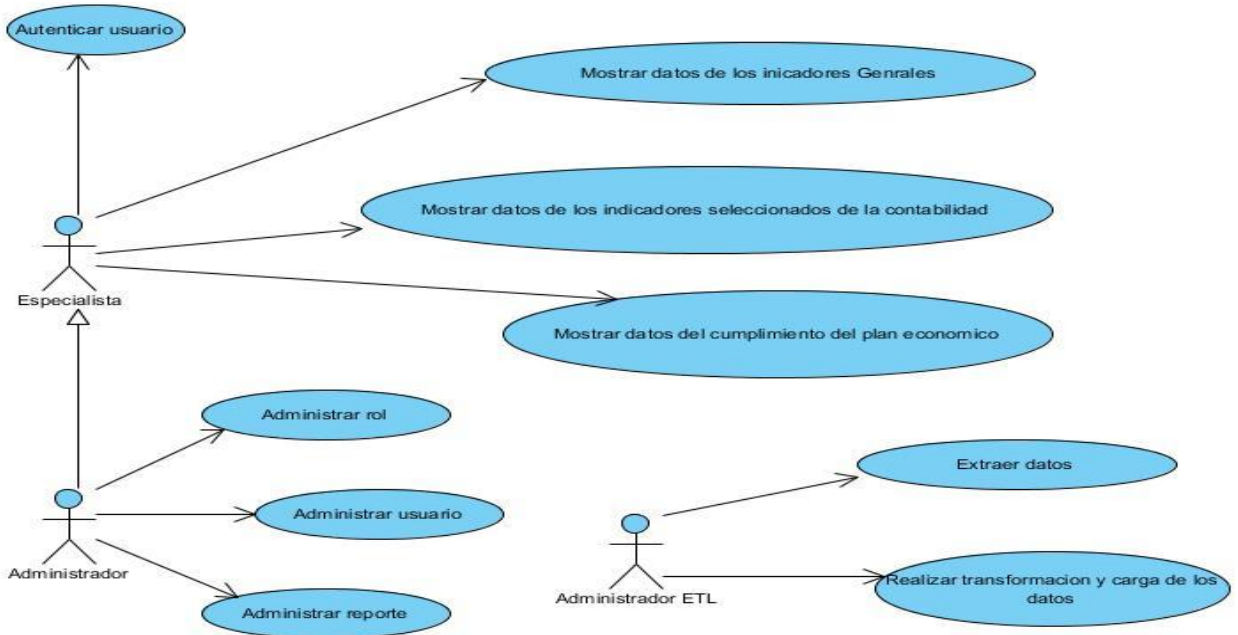


Figura 2. Diagrama de CU del sistema.

2.5.2 Actores del Sistema

Actor	Responsabilidad en el sistema
Administrador	Inserta y elimina los usuarios en el sistema. Inserta roles que van a ser asignados a los usuarios y elimina los innecesarios en el sistema. Inserta y modifica vistas de análisis que estarán disponibles en el sistema y elimina aquellos innecesarios. Consulta información de las vistas de análisis, los usuarios que han accedido al sistema y la información consultada por estos.

## *Capítulo 2: Análisis y Diseño del MD Contabilidad*

Administrador de ETL	Realiza la extracción de los datos de los ficheros fuentes, así como la transformación y la carga de los datos hacia el MD.
Especialista	Consulta la información de las vistas de análisis disponibles en el sistema.

Tabla 1: Descripción de los actores.

### 2.5.3 Especificación de Casos de Uso de Información

A continuación se describe el Caso de Uso de información (CUI) mostrar datos del cumplimiento del plan económico, los restantes se encuentran en el artefacto Especificación de casos de uso del expediente de proyecto del MD Contabilidad. (Ver Tabla 2: Descripción del CU Mostrar datos de los Indicadores Generales)

<b>Objetivo</b>	Obtener los datos de los Indicadores Generales.	
<b>Actores</b>	Analista.	
<b>Resumen</b>	El CU inicia cuando el analista selecciona visualizar información de los indicadores generales, luego se realiza la selección de la información a mostrar. Finaliza cuando se muestran los resultados.	
<b>Complejidad</b>	Alta	
<b>Prioridad</b>	Alta	
<b>Precondiciones</b>	Carga de los datos.	
<b>Postcondiciones</b>	Disponibilidad de opciones (cruce de variables) de reportes relacionados con el CU Mostrar datos de los Indicadores Generales.	
<b>Flujo de eventos</b>		
<b>Flujo básico &lt; Mostrar datos de los Indicadores Generales &gt;</b>		
	<b>Actor</b>	<b>Sistema</b>

## *Capítulo 2: Análisis y Diseño del MD Contabilidad*

1.	El especialista se autentica en el sistema.	
2.		Muestra la interfaz principal con las áreas de análisis existentes.
3.	El actor selecciona el A.A.G SIGOB.	
4.		Muestra las A.A que están contenidas dentro del A.A.G SIGOB.
5.	El actor selecciona el A. A Contabilidad.	
6.		Muestra los libros de trabajo que están contenidos dentro del A. A Contabilidad.
7.	El actor selecciona el L.T Indicadores Generales.	
8.		Muestra los reportes contenidos dentro del L.T Indicadores Generales.
9.	El actor selecciona el reporte deseado	
10.		Muestra la información contenida en el reporte seleccionado y brinda la posibilidad al actor de hacerle cambios al reporte para su análisis.  Ir al C.U <b>Visualizar reporte.</b>  Finaliza el caso de uso.
<b>Flujos alternos</b>		
	<b>Actor</b>	<b>Sistema</b>



1.			
2.			
	<b>Perspectivas de análisis</b>	<b>Posibles resultados</b>	
		<b>Mediadas.</b>	<b>Periodicidad.</b>
	Variables de entrada relacionadas con el <b>CU</b> Mostrar datos de los Indicadores Generales. <ul style="list-style-type: none"> <li>• Indicador</li> <li>• Entidad</li> <li>• Tiempo</li> <li>• NAE</li> <li>• DPA</li> <li>• Organismo</li> <li>• Formas de financiamiento</li> <li>• Variante.</li> </ul>	Variables de salida disponibles en el hecho <b>Indicadores Generales.</b> <ul style="list-style-type: none"> <li>• Plan real año actual.</li> <li>• Plan real año anterior.</li> <li>• Año anterior.</li> </ul>	Rango de tiempo en que se solicitan las variables de salida: <ul style="list-style-type: none"> <li>• Mensual.</li> </ul>

Tabla 2: Descripción del CU Mostrar datos del cumplimiento del Plan Económico.

### 2.6 Arquitectura del Mercado de Datos.

El MD está ordenado mediante una arquitectura compuesta por la fuente de datos, el Subsistema de integración, Subsistema de almacenamiento y Subsistema de visualización. A continuación una descripción de cada uno de estos Subsistemas.

- En el subsistema de integración de datos es donde se realizan todos los procesos ETL en los cuales se extrae, se limpia e integra toda la información almacenada en los sistemas fuentes a través de

transformaciones y trabajos.

- En el subsistema de almacenamiento es donde se guarda toda la información que ha sido transformada en el subsistema de integración.
- En el subsistema de visualización de la información es donde se muestra toda la información almacenada al cliente, a través de vistas de análisis. Los mismos permiten al cliente realizar un análisis de toda la información procesada.
- La arquitectura del MD Contabilidad queda diseñada de la siguiente manera. (Ver figura 3: Arquitectura del MD).

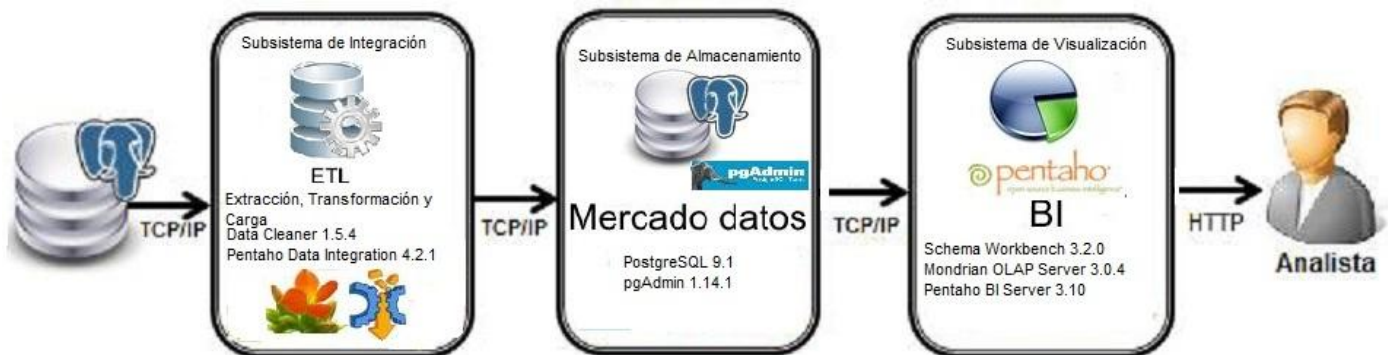


Figura 3: Arquitectura del MD Contabilidad.

Primeramente se encuentra el Subsistema de integración que obtiene la información de la BD de SIGE, en el cual se realiza el proceso de integración de los datos para su almacenamiento posterior. Luego el subsistema de almacenamiento recibe la información integrada y la almacena en la BD destino, la cual es soportada por el SGBD PostgreSQL 9.1 y es administrada por los usuarios que tienen acceso a la información mediante pgAdminIII 1.14.1. Finalmente en el Subsistema de visualización se muestra la información a través de vistas de análisis mediante la herramienta Pentaho BI Server 3.10, la cual es accedida por los usuarios autorizados a la visualización de las vistas de análisis.

## 2.7 Diseño del Mercado de Datos

### 2.7.1 Diseño del Subsistema de Almacenamiento

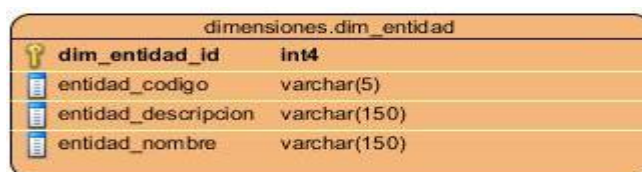
Para la realización del MD se realiza el modelo dimensional de los datos, el que contiene las tablas de hechos y las dimensiones, así como las relaciones entre estas. Como modo de almacenamiento se elige

ROLAP debido a que el SGBD utilizado PostgreSQL, no soporta el almacenamiento multidimensional de la información.

- **Dimensiones.**

Como parte del diseño del Subsistema de almacenamiento se identifican las dimensiones que constituyen las perspectivas de análisis de la información y las características de los hechos con los que está relacionada. Las dimensiones están constituidas por jerarquías y niveles, los que detallan la información de forma dinámica.

En el MD Contabilidad se identificaron 9 dimensiones, a continuación se detalla la dimensión entidad “dim\_entidad”, las demás dimensiones se encuentran en el artefacto DATEC\_SIGOB\_MD\_Cont\_Especificaciones\_del\_Modelo\_de\_Datos\_Dimensional. (Ver Figura 4: Dimensión Entidad)



dimensiones.dim_entidad	
dim_entidad_id	int4
entidad_codigo	varchar(5)
entidad_descripcion	varchar(150)
entidad_nombre	varchar(150)

Figura 4: Dimensión Entidad

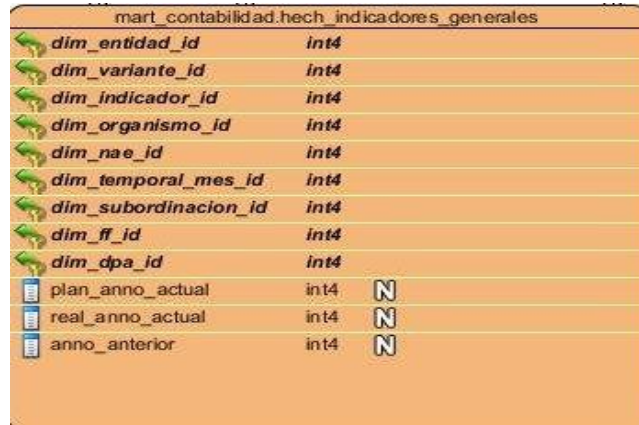
La dimensión entidad contiene la información de las diferentes de entidades que brindan información a la ONEI.

- **Hechos.**

Las tablas de hechos son las tablas principales del esquema dimensional, contienen los valores de las medidas del negocio. En el MD Contabilidad se definen como medida numérica el valor que se obtiene en la fuente de datos de esta área. Las tablas de hechos solo almacenan las mediadas numéricas y las llaves primarias de cada una de las dimensiones que se relacionan con la misma. En el MD Contabilidad se identificaron tres tablas de hechos, a continuación se detalla un ejemplo, las demás las pueden consultar en el artefacto DATEC\_SIGOB\_MD\_Cont\_Especificaciones\_del\_Modelo\_de\_Datos\_Dimensional.

Hecho “indicadores generales” (hech\_indicadores\_generales): este hecho recopila información referente a los indicadores de producción e ingresos y los indicadores del sector externo de las Entidades Estatales,

las Sociedades Mercantiles, las Empresas Mixtas, las Empresas de capital totalmente extranjero. (Ver Figura 5: Hecho Indicadores Generales)



mart_contabilidad.hech_indicadores_generales		
dim_entidad_id	int4	
dim_variante_id	int4	
dim_indicador_id	int4	
dim_organismo_id	int4	
dim_nae_id	int4	
dim_temporal_mes_id	int4	
dim_subordinacion_id	int4	
dim_ff_id	int4	
dim_dpa_id	int4	
plan_anno_actual	int4	N
real_anno_actual	int4	N
anno_anterior	int4	N

Figura 5: Hecho Indicadores Generales

- **Medidas.**

Las medidas son los campos de la tabla que se desean analizar.

Las medidas se clasifican en:

- Aditivas: pueden ser agregada a lo largo de cualquier dimensión.
- Semiaditivas: pueden no ser agregada a lo largo de una o más dimensiones.
- No aditivas: no pueden ser agregada a lo largo de ninguna dimensión.

En el MD contabilidad se identificaron 13 medidas, las cuales son físicas además de ser todas aditivas. A continuación se muestra un ejemplo de una de las medidas identificadas en el MD Contabilidad y perteneciente al hecho “indicadores generales” (hech\_indicadores\_generales).

**Medida** Plan Año Actual (plan\_anno\_actual): valor contable que refleja las cifras oficiales (de los indicadores que se analizan) del plan acumulado hasta el período que se informa.

### 2.7.2 Matriz Dimensional o Matriz Bus

La matriz bus refleja las relaciones existentes entre los hechos y las dimensiones, la misma permite verificar que no haya más de un hecho que se relacione con las mismas dimensiones, o sea que no existan ambigüedades en las relaciones. (Ver Tabla 3: Matriz bus)

## Capítulo 2: Análisis y Diseño del MD Contabilidad

Lista de Hechos:

H1: Indicadores Generales.

H2: Indicadores Seleccionados de la Contabilidad.

H3: Cumplimiento del plan económico.

Dimensiones	Hechos		
	H1	H2	H3
Indicador General	X	X	X
Entidad	X	X	X
DPA	X	X	X
NAE	X	X	X
Organismo	X	X	X
Formas de Financiamiento	X		
Subordinación	X	X	X
Variante	X		X
Tiempo	X	X	X

Tabla 3. Matriz Bus

En el MD Contabilidad la matriz bus permitió reconocer que los 3 hechos identificados comparten dimensiones, sin embargo no existen hechos solapados porque no existen dos o más hechos que compartan exactamente las mismas dimensiones. Se identificó además que de las 9 dimensiones 7 son compartidas por los 3 hechos entre las que se encuentran DPA, NAE y organismo.

### 2.7.3 Modelo de Datos

Una vez definidas en el negocio las dimensiones y los hechos se conforma el modelo de datos del MD Contabilidad cuyo objetivo es representar la estructura lógica de la información, evidenciando además las relaciones existentes entre las tablas de hechos y las tablas dimensiones. A continuación se muestra en la Figura 6 el modelo de datos dimensional propuesto para el desarrollo del MD Contabilidad.

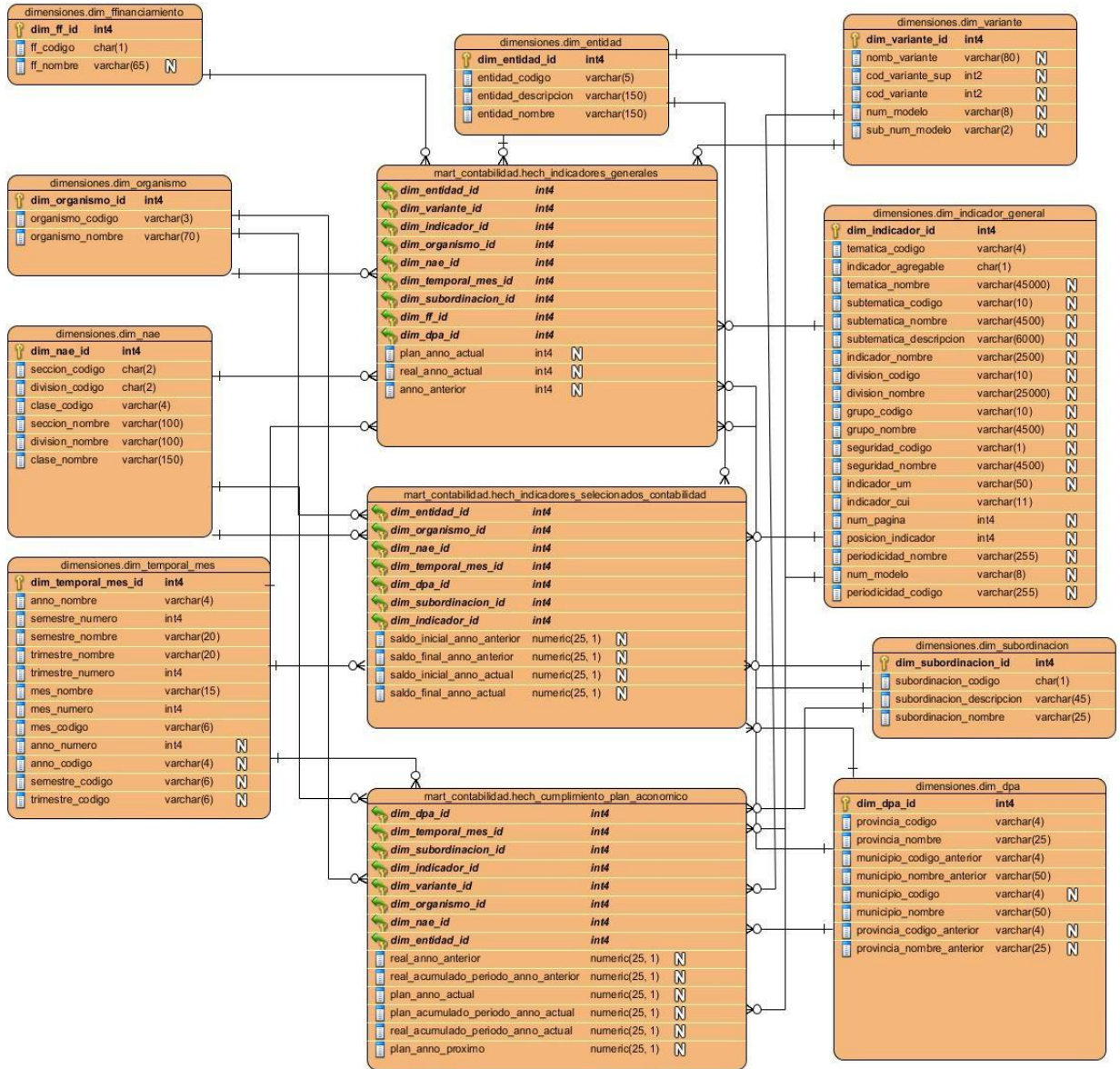


Figura 6: Modelo de datos MD contabilidad.

## Topología de esquema

La topología que se empleó en el diseño del modelo de datos físico y lógico del MD Contabilidad fue el esquema Constelación de Hechos, porque existen tablas de hechos que comparten algunas tablas de dimensiones. En el desarrollo del MD Contabilidad se evidencia la utilización de este esquema pues las tablas de hechos Indicadores Generales y Cumplimiento del Plan Económico comparten las dimensiones

DPA, variante, organismo, entre otras.

### 2.8 Políticas de Respaldo y Recuperación

En el MD Contabilidad la política de respaldo y recuperación se lleva a cabo teniendo en cuenta tres aspectos principales: tablas involucradas, periodicidad de las salvas y backups existentes.

- **Tablas Involucradas:** en la realización de las salvas se involucran las siguientes tablas, hech\_cumplimiento\_plan\_economico, dim\_nae, hech\_indicadores\_generales\_, hech\_indicadores\_seleccionados\_contabilidad, dim\_dpa, dim\_variante, dim\_organismo, dim\_subordinacion, dim\_entidad, dim\_temporal\_mes, dim\_ffinanciamiento y dim\_indicador\_general.
- **Periodicidad de las Salvas:** las salvas de la información contenida en el MD se hará en correspondencia con la periodicidad de cada modelo. La salva del modelo “0005” o Indicadores Generales se realizará mensualmente y la salva de los modelos “5901” y “5903” se realizará semestral y trimestral respectivamente.
- **Backups existentes:** cada año se realizarán dos salvas, una finalizado el mes junio y otra finalizado el mes de diciembre. Esta salva contiene la información de cada una de las entidades nacionales e internacionales al cierre de cada semestre.

### 2.9 Diseño del Subsistema de Integración

Uno de los elementos más importantes del MD es el diseño del Subsistema de integración del cual forma parte el **perfilado de datos**, que permite realizar un análisis profundo de los datos provenientes de la fuente para conocer el estado en que se encuentra la información así como su calidad y estructura.

Una vez realizado el perfilado de datos a la fuente del MD Contabilidad se pudo observar que la información que se encontraba almacenada corresponde a los años 2011 y 2012, los que utilizan como máscara de fecha yyyy/mm/dd. Los datos encontrados son: textos, números enteros y fechas además no existen valores negativos ni se encontraron valores vacíos.

Permite además verificar los distintos tipos de datos que contiene la fuente de Información del MD Contabilidad, como se muestra en la Figura 7: Distribución del tipo de dato.



### Distribución del Tipo de dato.

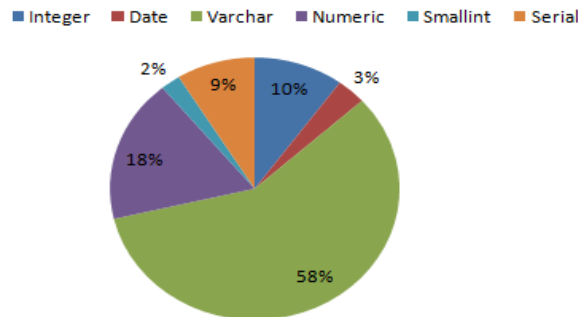


Figura 7: Distribución del tipo de dato.

#### 2.9.1 Diseño General de las Transformaciones

Las transformaciones son el elemento fundamental del proceso de ETL, en el diseño de las transformaciones se detalla cada uno de los pasos a seguir para efectuar la carga de las dimensiones y los hechos en el MD. A continuación se muestra detallado de forma general el diseño de las transformaciones del MD Contabilidad. (Ver Figura 8: Diseño de Transformaciones.)



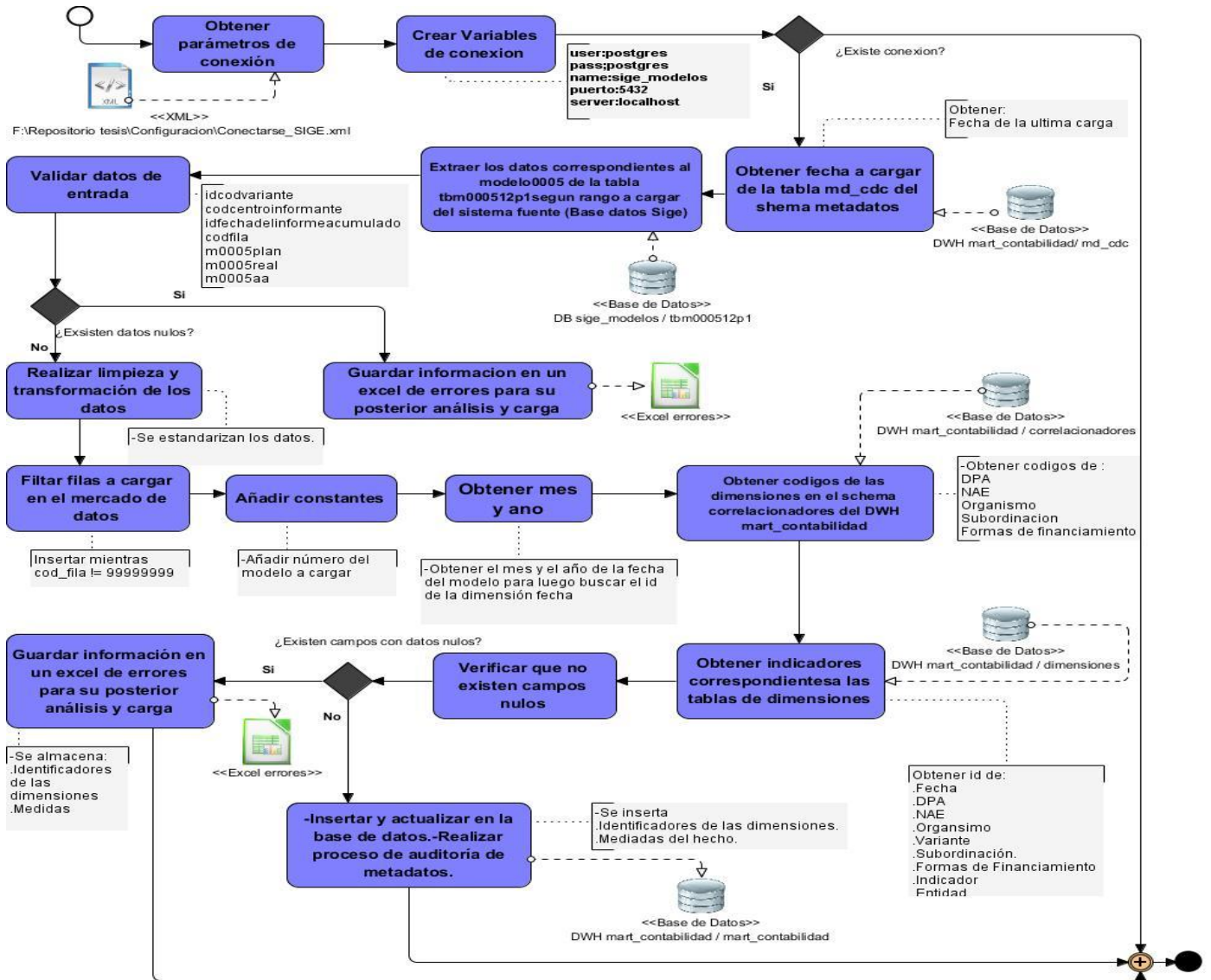


Figura 8: Diseño de Transformaciones.

## 2.9.2 Estrategias Generales de Integración de Datos

Para el desarrollo del presente trabajo se utiliza como estrategia de integración de datos: ETL. Se decide la utilización de esta estrategia pues da soporte a distintos orígenes de datos desde BD relacionales y no relacionales hasta archivos XML, archivos en formatos Excel y DBF, entre otro gran conjunto de ficheros. ETL brinda la posibilidad de enviar datos a otros software en tiempo real, tiene una amplia capacidad de transformación de datos, desde la más simple conversión de un tipo de datos, cálculos simples, hasta

transformaciones algo más complejas como agregaciones y sumalizaciones. Posee una amplia visualización de flujos, ya que se puede observar gráficamente cual es el recorrido del flujo de información desde que los mismos son leídos desde los sistemas fuentes hasta que son escritos en las distintas tablas de hechos y dimensiones y posee funcionalidades avanzadas que permiten limpiar los datos de origen evitando cargar información errónea en el AD.

La primera parte del proceso de ETL consiste en extraer la información desde los sistemas fuentes, en este caso sería la BD SIGE, cuyo formato es una BD relacional. Una parte importante de este subproceso consiste en analizar los datos extraídos para verificar si los mismos cumplen con la estructura requerida. Una vez extraídos estos datos quedan preparados para la fase de transformación.

En la fase de transformación del proceso de ETL, a los datos extraídos se le aplican una serie de reglas de negocio o una manipulación específica en dependencia del negocio y de la fuente de información, para convertirlos en los datos que realmente serán cargados. Además en esta fase se define la información de la fuente que se desea cargar. Una vez terminado este subproceso los datos quedan listos para su posterior almacenamiento.

En el subproceso de carga se toman los datos de la fase de transformación para cargarlos en el sistema destino, el MD Contabilidad. Esta fase puede contener varias acciones pues en los AD generalmente se mantiene un historial de la carga de la información, cuyo objetivo es no actualizar los datos almacenados con anterioridad, para disponer de un historial de uno o más valores a lo largo del tiempo.

### 2.10 Diseño del Subsistema de Visualización

El diseño del Subsistema de visualización se realiza con el objetivo de organizar las vistas de análisis por áreas de análisis, facilitando al usuario una búsqueda rápida de la información.

#### 2.10.1 Arquitectura de la Información

Con la creación del MD se podrá almacenar la información que pertenece al área de Contabilidad, para ser analizada por las personas capacitadas y sirviendo de apoyo al proceso de toma de decisiones. En la investigación se identificó un área de análisis que se divide en tres libros de trabajo los que incluyen las tablas de salida definidas por el cliente.

- **Área de análisis (AA):** Son las secciones en las que se divide el AD, las que contienen los diferentes libros de trabajo y las vistas de análisis.

- **Libro de Trabajo (LT):** Son las diferentes categorías en las que se pueden agrupar cada una de las vistas de análisis.
- **Vistas de análisis:** Informe que expresa mediante tablas los valores de los indicadores analizados.

En la siguiente imagen se muestra el mapa de navegación del MD Contabilidad. (Ver Figura 9: Mapa de Navegación)

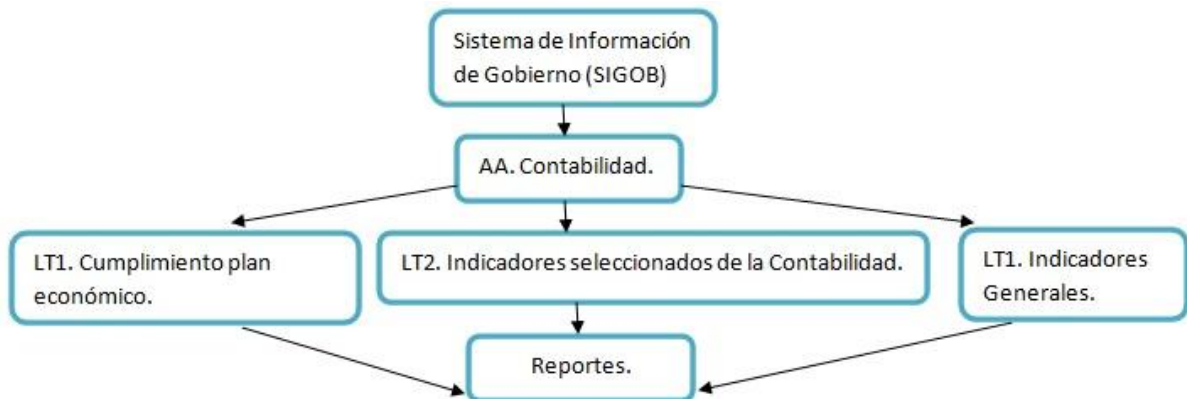


Figura 9: Mapa de Navegación MD Contabilidad.

### 2.10.2 Diseño de las Vistas de Análisis

Para el desarrollo del MD Contabilidad se identificaron nueve vistas de análisis, todos por petición del cliente. A continuación se muestran las vistas de análisis agrupadas por AA Contabilidad y los LT.

#### AA. Contabilidad.

**LT Indicadores Generales:** El libro de trabajo indicadores generales contiene tres vistas de análisis que se realizan en función de la variante, DPA, NAE, entidad e indicador general.

**LT Indicadores seleccionados de la contabilidad:** El L.T indicadores seleccionados de la contabilidad cuenta con tres vistas de análisis, las mismas se realizaron en función de la entidad, DPA, NAE e indicador general.

**LT Cumplimiento del plan económico:** El L.T cumplimiento del plan económico contiene tres vistas de análisis, las que se realizan en función de organismo, DPA y NAE.

### 2.10.3 Diseño de los Cubos OLAP

En el MD Contabilidad se definieron tres cubos multidimensionales que fueron realizados con el Pentaho Schema Workbench, los cuales corresponden con cada uno de los hechos identificados. En la solución propuesta se identificaron los siguientes cubos: indicadores generales, cumplimiento del plan económico e indicadores seleccionados de la contabilidad. La Figura 10 muestra el diseño del cubo correspondiente al hecho Indicadores generales.



Figura 10: Diseño del cubo multidimensional.

### 2.11 Esquema de Seguridad

En el MD Contabilidad es de gran importancia la seguridad de la información, pues los datos que maneja son de vital importancia para la economía del país. Con este objetivo se definieron roles para darle permisos a cada uno de los usuarios que interactúan directamente con el sistema.

#### Seguridad en el Subsistema de Almacenamiento

Para la seguridad de la BD se creó el rol Administrador de BD el que posee acceso total a la BD y el Administrador de ETL que se encarga de los procesos de extracción transformación y carga. En la Tabla 4

se muestra el esquema de seguridad definido para el Subsistema de Visualización.

Roles	Permisos
Administrador BD.	Total acceso a la BD. Realiza la administración de la BD, que contiene todas las tablas de hechos y dimensiones del AD SIGOB. Autoriza permiso a cada uno de los usuarios.
Administrador ETL.	Realiza todos los procesos de extracción, transformación y carga de la información. Y tiene todos los permisos sobre las tablas de hechos y dimensiones del MD Contabilidad.

Tabla 4: Esquema de Seguridad Subsistema de almacenamiento.

### Seguridad en el Subsistema de Integración

Como el MD Contabilidad se desarrolla sobre el sistema operativo Linux, se garantiza asignar permisos sobre archivos a un grupo determinado de usuarios. Esto permitirá restringir el acceso sobre los archivos que contienen las transformaciones y trabajos encargados de los procesos de integración del MD Contabilidad solo al personal autorizado, en este caso al administrador de ETL.

### Seguridad en el Subsistema de Visualización

Para la seguridad de la aplicación se definió el rol Administrador, que posee total acceso al área de análisis SIGOB, además de ser el encargado de la creación de nuevos usuarios, así como asignarles los roles y permisos a los mismos y el rol especialista que tiene acceso al área de análisis Contabilidad para de este modo consultar las vistas de análisis correspondientes a cada uno de los libros de trabajo de esta área. En la Tabla 5 se muestra el esquema de seguridad definido para el Subsistema de Visualización.

Roles	Permisos
Administrador	Acceso total al Área Análisis General de SIGOB, creación de roles, usuarios y gestión de permisos y vistas de análisis candidatos.
Especialista.	Acceso solo para visualizar las vistas de análisis del Área Análisis Contabilidad.

Tabla 5: Esquema de Seguridad subsistema de visualización.

### **2.12 Conclusiones del Capítulo**

En el capítulo fueron especificados los procesos que se llevan a cabo en el desarrollo del análisis y diseño del MD Contabilidad. Las necesidades de información identificadas permitieron conocer las necesidades de información del área Contabilidad de la ONEI pues se identificaron trece requisitos de información y nueve requisitos funcionales agrupados en tres casos de uso de información y seis casos de uso funcionales respectivamente. Se definieron las reglas del negocio que servirán de base para realizar las transformaciones. Con el diseño del Subsistema de almacenamiento se determinaron los elementos que formaron parte del modelo físico identificando tres hechos, nueve dimensiones y trece medidas. El diseño del Subsistema de integración realizado servirá de apoyo a la implementación de los procesos de ETL. Una vez diseñado el Subsistema de visualización se definió la estructura de la capa de presentación compuesta por el área de análisis Contabilidad, tres libros de trabajo y once vistas de análisis. También se realizó el diseño de los tres cubos OLAP correspondientes a los tres casos de uso de información identificados.

### Capítulo 3: Implementación del MD Contabilidad

#### Introducción:

El presente capítulo está encaminado a la implementación de los diferentes aspectos relacionados con los procesos de integración de datos, con el propósito de brindar una mayor comprensión de las estrategias y procedimientos utilizados. Además, se abordan elementos relacionados con la implementación de la capa de visualización, incluyendo la creación de las estructuras necesarias para la navegación y el análisis de los datos.

#### 3.1 Implementación del Subsistema de Almacenamiento

Posterior al diseño del modelo de datos dimensional se realizó la transformación al modelo físico, el que facilita la descripción del almacenamiento de los datos y la relación existente entre las tablas que lo componen.

#### Estructura de los datos

Una estructura de datos es una vía factible para organizar una colección de datos. En la BD la información se encuentra organizada en estructuras lógicas que facilitan una correcta manipulación de la misma. Estas estructuras son denominadas esquemas que están conformados por una o varias tablas.

**Esquemas:** los esquemas en una DB representan una manera eficiente de tener organizada la información. Estos pueden contener funciones, operadores y tipos de datos. Esta estructura le permite al usuario tener acceso a ellos siempre y cuando posea los permisos adecuados.

Para el desarrollo de la solución se cuenta con cuatro esquemas:

- Esquema dimensiones: este esquema contiene las dimensiones comunes del AD SIGOB.
- Esquema correlacionadores: este esquema contiene los correlacionadores de los modelos “0005”, “5901” y “5903”.
- Esquema metadatos: este esquema contiene información relacionada con el control de cambios, cuenta con cinco tablas, las que permiten capturar los cambios del registro histórico de las cargas realizadas, información del flujo de ejecución de las transformaciones y control del fecha de ejecución de cada una de las transformaciones correspondientes a cada uno de los hechos identificados.

## Capítulo 3: Implementación del MD Contabilidad

- Esquema mart\_contabilidad: en este esquema se encuentran las tablas de hechos del MD contabilidad.

**Tablas:** Las tablas representan el conjunto de registros que te van a permitir describir un elemento individual de la información. La solución propuesta contiene un total de nueve tablas contenidas en el esquema dimensiones, una tabla contenida en el esquema correlacionadores, cinco tablas contenidas en el esquema metadatos y tres tablas contenidas en el esquema mart\_contabilidad, para un total de catorce tablas que se representan en la siguiente tabla que describe dicha estructura.

No.	Esquema	Tablas
1.	dimensiones	dim_dpa
		dim_nae
		dim_ffinanciamiento
		dim_entidad
		dim_organismo
		dim_variante
		dim_temporal_mes
		dim_subordinacion
		dim_indicador_general
2.	correlacionadores	correlacionadores
3.	metadatos	md_cdc
		md_carga_historcia
		md_mercado
		md_transformacion
		md_temporal



## *Capítulo 3: Implementación del MD Contabilidad*

<b>4.</b>	mart_contabilidad	hech_cumplimiento_plan_economico
		hech_indicadores_generales
		hech_indicadores_seleccionados_contabilidad

Tabla 6: Esquemas y tablas.

### 3.1.1 Estándares de Codificación

Los estándares de codificación se utilizan para lograr un entendimiento entre las partes implicadas en un proyecto. Para el desarrollo del MD Contabilidad fueron utilizados varios estándares de codificación que forman parte del proyecto SIGOB, a continuación se exponen los mismos en la siguiente tabla:

Tipo de objeto	Función	Nomenclatura	Descripción
Esquema	Dimensiones compartidas	dimensiones	Esquema donde se organizan las dimensiones compartidas por varios MD (tablas dimensionales y secuencias).
	Esquemas de datos	mart_[temática]	Esquema donde se almacenan las tablas de hechos definidas para gestionar los datos asociados a cada área temática, en el caso del MD Contabilidad sería: mart_contabilidad.
Tablas	Dimensiones	dim_[nombre]	Tablas de dimensiones utilizadas. Ejemplos: dim_dpa, dim_nae, entre otras.
	Hechos	hech_[nombre]	Tablas de hechos que definen las principales medidas requeridas para calcular indicadores y otras medidas derivadas. Ejemplo: hech_indicadores_generales.

	Metadatos	md_[nombre]	Tablas de metadatos donde se guarda la información correspondiente a los datos del mercado. Ejemplo md_cdc.
--	-----------	-------------	---

Tabla 7: Estándares de codificación del MD Contabilidad.

### 3.2 Implementación del Subsistema de Integración

En el proceso de integración de datos no es recomendable iniciar los procesos de ETL sin haber analizado previamente la fuente de datos. La información histórica del área de Contabilidad de la ONEI se encuentra almacenada en una BD relacional soportada por el SGBD PostgreSQL9.1. Una de las etapas más importantes de este proceso es la de transformación y limpieza, una vez realizada estas, ya los datos se encuentran listos para ser almacenados en la BD. Con la limpieza de la información se detectan los datos que contienen errores, las entradas duplicadas, y las transformaciones nos permiten filtrar información, combinar y ordenar los datos.

#### Subsistema de Extracción

En el subsistema de extracción se obtienen los datos de la fuente proporcionada por el cliente que poblarán el MD Contabilidad. Estos datos fueron extraídos del SIGE, donde se encuentran las tablas “tbn000512p1”, “tbn590108p1” y “tbn590307p1” correspondientes a los modelos “0005” Indicadores Generales, “5901” Indicadores Seleccionados de la Contabilidad, “5903” Cumplimiento del Plan Económico, que contienen la información del área Contabilidad de la ONEI desde el año 2011 al 2013.

#### Subsistema de Transformación

Una vez finalizada la extracción de los datos se prosigue a la transformación de los mismos. En este proceso se detectan los posibles errores que contenga la información y se les da un tratamiento para corregirlos aplicando las reglas de negocio identificadas garantizando que los datos sean cargados con la calidad requerida por el usuario final. En el desarrollo del presente trabajo se implementaron un total de 4 transformaciones con el objetivo de poblar el MD Contabilidad, tres correspondientes a la carga de cada uno de los hechos y una a la carga de las variables de conexión, tanto a la fuente de información como al MD Contabilidad.

## *Capítulo 3: Implementación del MD Contabilidad*

---

En la Figura 11 se muestra el flujo de la transformación realizado para la carga del hecho Indicadores Generales (hech\_indicadores\_generales), en el mismo se hace uso del repositorio de metadatos, específicamente del metadato que controla el cambio en la fuente, con el objetivo de capturar la última fecha de carga de información del hecho Indicadores Generales. Luego se realizó la extracción de los datos a partir de la tabla “tbm000512p1” la cual hace alusión a la información recogida por el modelo “0005” o Indicadores Generales y utilizando el componente “Data validator” se verificó que ninguno de los campos tenga valores nulos. A través del componente “Filtrar Fila” se descartaron las filas que no se deseaban cargar en el MD. Además se realizan cálculos utilizando el componente “Calculadora” para obtener dos nuevos campos necesarios para buscar el identificador de la dimensión temporal mes. Se realizó la búsqueda en las tablas de dimensiones dim\_nae, dim\_temporal\_mes, dim\_dpa, dim\_organismo, dim\_variante, dim\_ffinanciamiento, dim\_indicador, dim\_entidad y dim\_organismo para obtener los identificadores de cada una de ellas. Una vez que el flujo estuvo listo con todos los datos correspondientes a cargar en la tabla de hechos se validó que no existiera ningún campo nulo para ser cargado, y de haber algún tipo de error se capturaron estos en un fichero Excel con la finalidad de ser revisados por el cliente para luego realizar una transformación con una entrada Excel y cargar todos esos datos una vez arreglado el error. Una vez comprobado que los datos estén correctos se procede a insertarlos en la tabla de hechos “hech\_indicadores\_generales”. Para finalizar se actualizó la tabla de metadatos “md\_cdc” con la fecha de carga, con el objetivo de garantizar la gestión histórica de los datos.

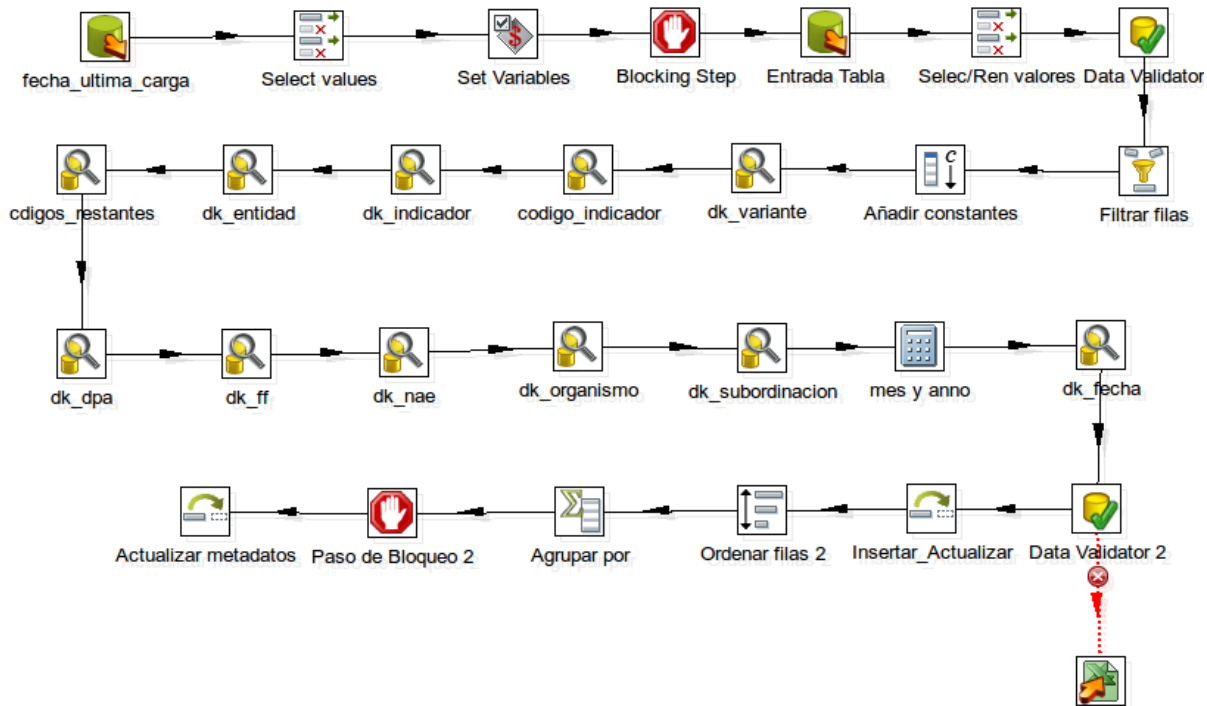


Figura 11: Transformación del hecho Indicadores Generales.

### Nomenclatura

Con el objetivo de lograr una estandarización en el desarrollo de las transformaciones y los trabajos se utilizó la siguiente nomenclatura:

Cuando se agregó una nueva transformación en el proceso de integración de datos, para cargar los hechos se nombró de la siguiente manera:

- transf\_hech\_<nombre>

Ejemplo: transf\_hech\_indicadores\_generales.ktr

Cuando se agregó un nuevo flujo de transformaciones (trabajo), en el proceso de integración de datos, se nombró de la siguiente manera:

- trabajo\_<nombre>

Ejemplo: trabajo\_cargar\_hechos.kjb

### Carga de Datos

El proceso de carga de datos constituye el paso final del desarrollo del AD, donde la información que fue transformada en los procesos de extracción y transformación se cargó hacia las tablas de hechos correspondientes. Estas tablas conforman el modelo dimensional del MD Contabilidad, posibilitando su consulta por los usuarios finales.

### **3.2.2 Gestión de los Metadatos del Proceso de Integración**

Con el objetivo de definir la estrategia de gestión de cambios en el MD Contabilidad se tuvo en cuenta la periodicidad con que se carga la información de cada uno de los modelos. Teniendo en cuenta que cada modelo tiene una periodicidad distinta, se incluye en la tabla de metadatos un campo con el nombre de la transformación, nombre del hecho y la última fecha de carga para llevar constancia de la fecha en la que se ejecutan las transformaciones para la carga de los hechos correspondientes a los modelos. Como se mencionaba anteriormente la tabla de metadatos “md\_cdc” contiene el nombre de la transformación, además del identificador, la fecha en la que se ejecutó la transformación, nombre del modelo fuente y nombre del MD. Una vez realizada la carga de la información, es insertada la nueva fecha en la que se realizó la última carga para la ejecución de la siguiente. Los metadatos utilizados en el desarrollo de este trabajo son metadatos de procesos y metadatos técnicos, los que nos posibilitan obtener información de los procesos de las transformaciones y los trabajos referentes a los subprocessos de ETL. Además de este tipo de metadatos existen los metadatos de negocio que permiten obtener los datos y la información referente a los aspectos del negocio, como son los datos provenientes de la fuente.

### **3.3 Implementación del Subsistema de Visualización**

#### **3.3.1 Implementación de la Capa de Visualización**

La capa de visualización se estructuró según la arquitectura de información propuesta en el mapa de visualización, donde se evidencia la información que contiene la misma. El MD Contabilidad está compuesto por un Área de Análisis General (AAG), y otra específica, que hace referencia al MD anteriormente mencionado, además de tres Libros de Trabajo (L.T), en los cuales se encuentran las vistas de análisis implementadas para cada uno de los modelos. A continuación se muestra la Figura 12 donde se muestra la Arquitectura de la Información del MD Contabilidad.

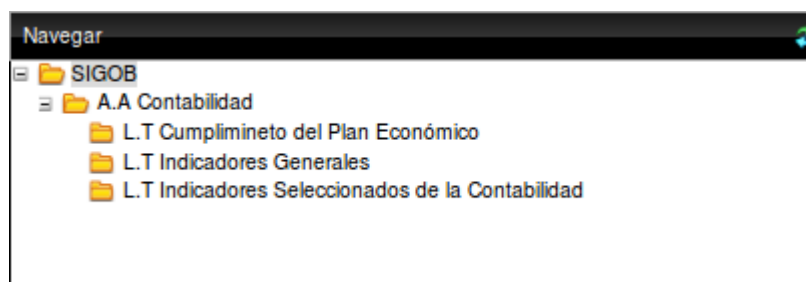


Figura 12: Arquitectura de la Información del MD Contabilidad.

A continuación se detalla la estructura que posee esta capa:

### Descripción del A.A.G:

**A.A.G SIGOB:** agrupa la información referente a los mercados de datos de las distintas áreas de la ONEI, los cuales conforman en su conjunto el almacén de datos SIGOB. Descripción del Área de Análisis:

**A.A Contabilidad:** contiene la información referente a los indicadores estadísticos del área de Contabilidad de la ONEI.

### Descripción de los Libros de Trabajo:

**L.T Indicadores Generales:** presenta las tres vistas de análisis referentes a los indicadores generales del área contabilidad de las disímiles entidades nacionales e internacionales existentes en el país.

**L.T Indicadores Seleccionados de la Contabilidad:** presenta 4 vistas de análisis relacionados con los indicadores seleccionados de la contabilidad de las diferentes entidades nacionales e internacionales existentes en el país.

**L.T Cumplimiento del Plan Económico:** presenta 5 vistas de análisis relacionados con los indicadores del cumplimiento del plan económico de las diferentes entidades nacionales e internacionales existentes en el país.

A continuación se presenta la Tabla de Salida (T.S) o reporte “Indicadores generales según División Político Administrativa” (ver Fig. 13) que se encuentra en el L.T Indicadores generales, así como la estructura en la capa de visualización:

Indicadores	Division Político Administrativa	Medidas		
		Año anterior	Plan año actual	Real año actual
		Fecha	Fecha	Fecha
Ventas netas de bienes y servicios				
	Pinar del Río	5.686.176	5.349.165	5.607.844
	Artemisa	3.472.186	3.611.860	36.791.733
	La Habana	80.604.928	86.679.167	93.787.356
	Mayabeque	3.439.385	3.692.777	3.377.559
	Matanzas	4.759.673	5.301.157	5.149.063
	Villa Clara	11.907.344	10.914.921	11.177.757
	Cienfuegos	7.674.143	7.943.162	7.933.977
	Sancti Spiritus	2.776.714	2.624.243	2.858.276
	Ciego de Ávila	11.836.141	13.091.460	12.526.321
	Camagüey	3.525.374	4.326.175	4.272.935
	Las Tunas	5.668.159	5.395.111	5.459.628
	Holguín	21.104.508	20.805.089	20.073.654
	Granma	7.272.131	7.584.505	7.199.411
	Santiago de Cuba	20.228.687	20.409.543	22.062.416

Figura 13: Reporte Indicadores Generales por División Política Administrativa.

### 3.1 Conclusiones del Capítulo

En el capítulo fueron tratados temas que dieron paso a la implementación del MD. Se realizó la transformación al modelo físico de la BD, facilitando una implementación de la BD consecuente con el diseño multidimensional. La estructura de los datos quedó organizada mediante 4 esquemas, logrando así una mejor y mayor organización de los datos. Se obtuvo una buena sincronización entre las fuentes de datos y el destino de los mismos, al ejecutar la extracción y transformación de la información para cargar la BD. En el proceso de BI se logra que los datos aporten información valiosa para el análisis y la toma de decisiones.

### Capítulo 4: Pruebas

#### Introducción

En el presente capítulo se realizan pruebas a la solución propuesta a través de las distintas vías, para de esta forma verificar su correcto funcionamiento y que cuente con la calidad requerida por el cliente. Estas pruebas se realizan a partir de la aplicación de listas de chequeo y los casos de prueba diseñados, aplicados a la aplicación.

#### 4.1 Calidad del Software

La calidad del software es el grado con el que un sistema, componente o proceso cumple los requerimientos especificados y las necesidades o expectativas del cliente o usuario. Concordancia del software producido con los requerimientos explícitamente establecidos, con los estándares de desarrollo prefijados y con los requerimientos implícitos no establecidos formalmente, que desea el usuario. (29)

#### Características de un Software de Calidad

El modelo de calidad ISO/IEC 9216 (Internacional Standard “Informationtechnology– Software Product Quality”) define 6 características de calidad que debe de cumplir toda aplicación Web para que sea un producto de calidad.

##### 1. Funcionalidad:

Es la capacidad de un producto software de satisfacer los requisitos funcionales prescriptos y las necesidades implícitas de los usuarios.

##### 2. Fiabilidad:

Es la capacidad de un producto software de mantener su nivel de desempeño, bajo condiciones establecidas, por un periodo de tiempo.

##### 3. Usabilidad:

Es la capacidad de un producto software de ser comprendido, usado y conforme con las reglamentaciones y guías de usabilidad.

##### 4 Eficiencia:

Es la capacidad de un producto software de proporcionar un rendimiento apropiado, de acuerdo a la cantidad de recursos usados bajo condiciones establecidas.



## 5 Mantenibilidad:

Es la capacidad de un producto software para ser modificado. Las modificaciones pueden incluir correcciones, mejoras o adaptación del software a cambios en el entorno, en los requisitos o en las especificaciones funcionales.

## 6 Portabilidad:

Es la capacidad de un producto software de ser transferido de un ambiente a otro. Nota: El ambiente puede ser organizacional, de software o de hardware. (30)

### 4.2 Pruebas

Las pruebas constituyen una de las fases más importantes en el desarrollo de cualquier producto de software y permiten determinar su calidad. Las mismas son aplicadas desde el inicio del desarrollo del software hasta que este llegue a manos del cliente, esperando que cumpla con sus expectativas.

Disímiles son los tipos de pruebas que pueden ser aplicadas a los MD con el fin de obtener un producto con la calidad requerida, para lo que existen tanto en el desarrollo como en las pruebas modelos de mejora. En la investigación se evaluó la Calidad del MD Contabilidad mediante el Modelo V, el cual es utilizado por el centro DATEC para garantizar la calidad de sus productos de software. La Figura 14 muestra cómo el Modelo V relaciona las actividades de prueba con el análisis y el diseño:

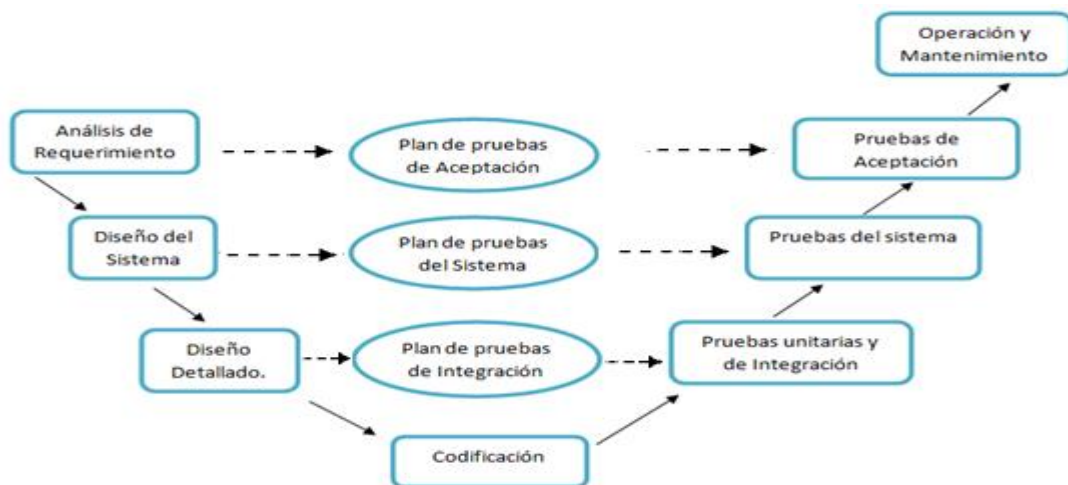


Figura 14: Modelo V

El modelo en V muestra la relación entre las actividades de prueba con el análisis y el diseño como se puede observar en la Figura 14. La unión mediante líneas discontinuas entre las fases de la parte izquierda y las pruebas de la derecha representa una doble información. Por un lado sirve para indicar en qué fase de desarrollo se deben definir las pruebas correspondientes. Por otro lado sirve para saber a qué fase de desarrollo hay que volver si se encuentran fallos en las pruebas correspondientes. (31)

A continuación se muestran las pruebas aplicadas al MD Contabilidad haciendo uso del modelo V.

- **Prueba unitaria:** Es el proceso de probar los componentes individuales (subprogramas o procedimientos) de un programa. El propósito es descubrir discrepancias entre la especificación de la interfaz de los módulos y su comportamiento real. Estas pruebas son diseñadas y ejecutadas por el desarrollador una vez terminado el desarrollo de cada componente.
- **Prueba de integración:** Son las pruebas que se realizan para determinar la integración de los componentes dentro de un sistema y evaluar su correcta interfaz, funcionalidad y desempeño. Estas pruebas son diseñadas y ejecutadas por el desarrollador cuando la solución está completa junto a los especialistas del centro.
- **Prueba de sistema:** Son las pruebas que se realizan para determinar el correcto funcionamiento de un sistema y su cumplimiento contra las especificaciones del producto.
- **Pruebas de aceptación:** Pruebas que se realizan directamente con el cliente para validar su conformidad con el producto.

### 4.2.1 Casos de Prueba

Los casos de prueba son utilizados por el analista para determinar si los requisitos de una aplicación son parciales o completamente satisfactorios y para identificar los posibles fallos de implementación. En el MD Contabilidad fueron diseñados tres casos de prueba asociados a los casos de uso de información identificados en la etapa de análisis, los cuales están regidos por las vistas de análisis candidatos asociados a cada tema de análisis, con el fin de verificar que sean visualizados con las variables correspondientes. A continuación en la Figura 15 se muestra un ejemplo de uno de los casos de prueba diseñados.

Escenario	Descripción	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5	Variable 6	Variable 7	Variable 8	Variable 9	Variable 10	Variable 11	Respuesta del sistema	Flujo central
EC 1.1 01-Contabilidad	Realizar entrevistas a usuarios del negocio y revisar las principales publicaciones de la entidad para identificar las necesidades informativas en el/los temas de análisis definidos, en función de la solución a desarrollar.	NAE	DPA	Indicador	Organismo	Subordinación	Tiempo	Entidad	Obtener el saldo inicial año	Obtener el saldo final año	Obtener el saldo inicial año actual	Obtener el saldo final de año	Se muestra la información correspondiente al escenario.	1. En la parte superior izquierda se selecciona el Área de Análisis General A.A.G SIGOB. 2. Se selecciona el Área de Análisis A.A Contabilidad 3. Se selecciona el Libro de Trabajo L.T Indicadores Seleccionados de la Contabilidad. 4. En la parte inferior izquierda, se selecciona el reporte que corresponde al escenario especificado. 5. Se visualiza el reporte en el área de trabajo que se indicó.

Figura 15: Caso de Prueba del CU “Mostrar Indicadores Seleccionados de la Contabilidad”

#### 4.2.2 Listas de Chequeo

Las listas de chequeo se crean con el fin de concretar y propiciar un buen desarrollo en el trabajo. Son un conjunto de preguntas que sirven para verificar el cumplimiento de los objetivos.

En esta investigación se aplicaron las siguientes listas de chequeo a los artefactos de los procesos de ETL con el fin de evaluar y verificar el potencial de cada uno de ellos, midiendo la confiabilidad y seguridad de los datos cargados:

- Lista de chequeo del Mapa Lógico de Datos.
- Lista de chequeo del Diccionario de Datos.
- Lista de chequeo de Registro de Sistemas Fuentes.
- Listas de chequeo del Perfilado de Datos

#### Estructuras de las Listas de Chequeo

Contienen diferentes indicadores a evaluar, los cuales se encuentran distribuidos en tres secciones:

- **Estructura del documento:** Contiene todos los aspectos definidos por el expediente del proyecto.
- **Indicadores definidos por la etapa:** Contiene todos los indicadores a evaluar durante la etapa de análisis de datos.
- **Semántica del documento:** Contiene todos los indicadores a evaluar respecto a la redacción y ortografía.

A cada uno de los indicadores anteriores los describen los siguientes elementos:

- **Peso:** Define si el indicador a evaluar es crítico o no.
- **Indicadores a evaluar:** Son los indicadores a evaluar en las secciones Estructura del documento, semántica del documento e indicadores definidos por las diferentes etapas.
- **Evaluación:** Es la forma de evaluar el indicador en cuestión. El mismo se evalúa de 1 en caso de que exista alguna dificultad sobre el indicador y 0 en caso de que el indicador revisado no presente problemas.
- **No procede:** Se usa para especificar que el indicador no es necesario evaluarlo en ese caso.
- **Cantidad de elementos afectados:** Especifica la cantidad de errores encontrados sobre el mismo indicador.
- **Comentario:** Especifica los señalamientos o sugerencias que quiera incluir la persona que aplica la lista de chequeo.

Luego de quedar definida la estructura de la lista de chequeo (Consultar Expediente de proyecto), la misma es aplicada al MD Contabilidad. Los resultados arrojados se muestran en forma de gráfico de barras (Ver Figura 16), donde se visualiza el comportamiento de los trece indicadores, de los cuales cinco son críticos y no fueron detectadas No Conformidades (NC).

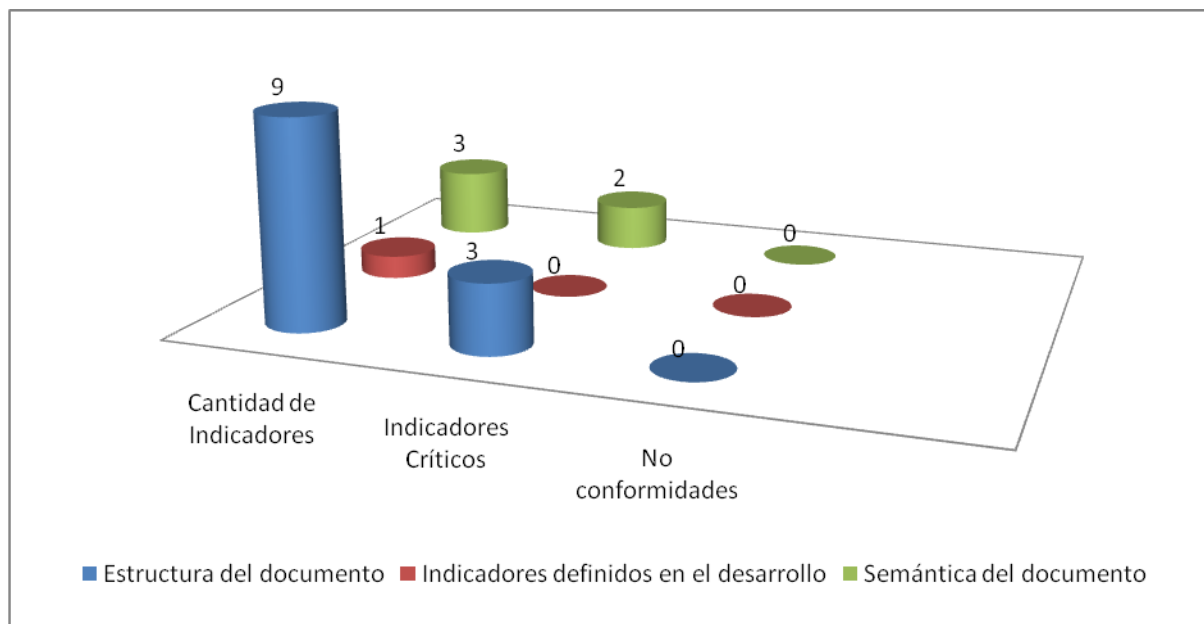


Figura 16: Comportamiento de los Indicadores de la Lista de Chequeo.

### 4.3 Resultados de las Pruebas

Las pruebas realizadas al MD Contabilidad arrojaron los siguientes resultados:

- Pruebas unitarias y de Integración:** Las pruebas unitarias y de integración realizadas al subsistema de integración de datos y a los diferentes componentes relacionados con la capa de visualización por los especialistas del departamento y los miembros del equipo de desarrollo, detectaron cinco NC las cuales fueron resueltas en su totalidad.
- Pruebas del sistema:** Las pruebas del sistema fueron realizadas por el grupo de calidad interno del departamento de AD, los tres casos de pruebas diseñados por cada uno de los casos de uso identificados permitieron identificar treinta y una NC, seis de complejidad baja, veinticinco de complejidad media y ninguna de complejidad alta, las cuales fueron resueltas en su totalidad. Los resultados obtenidos se muestran a través de un gráfico de barras (Ver Figura 17) que contiene las NC y su distribución según la complejidad:

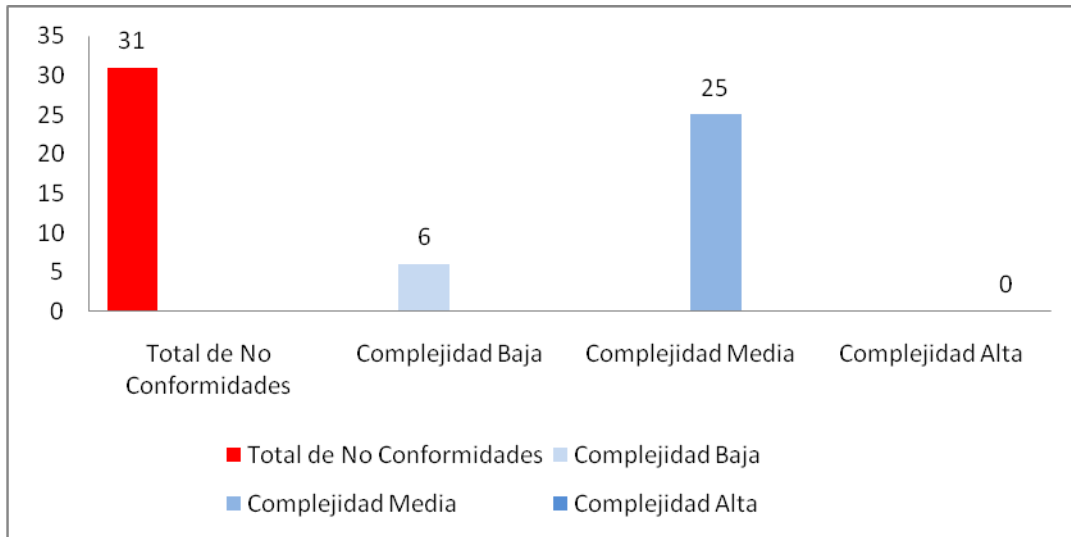


Figura 17: Resultados de los Casos de Pruebas basados en los CUS.

**Pruebas de aceptación:** En conjunto con el cliente se realizaron las pruebas de aceptación de la solución, las cuales arrojaron resultados satisfactorios, quedando comprobado que el sistema cumple con sus necesidades y que están satisfechos con el producto elaborado.

#### 4.4 Perfilado de los Datos

Una vez finalizada la implementación del MD Contabilidad se realizó un perfilado de datos que permitió examinar los Información almacenada del área de Contabilidad de la ONEI. Se utilizó la herramienta DataCleaner 1.5.4 para realizar el perfilado de los datos a la BD relacional que contiene dicha información. El resultado de este proceso indica que no se almacenaron valores nulos y que las tablas de hechos contienen únicamente valores numéricos. La Figura 18 que se observa a continuación muestra los resultados del perfilado de datos realizado a la tabla hech\_indicadores\_generales.

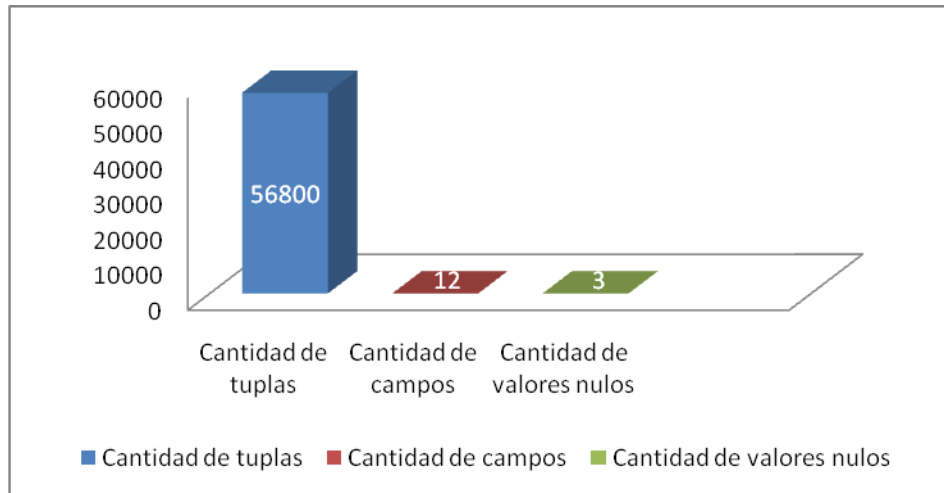


Figura 18: Resultados del perfilado de datos realizado a la Tabla hech\_indicadores\_generales.

#### 4.5 Conclusiones del Capítulo

Las pruebas realizadas al MD Contabilidad detectaron NC, las cuales fueron resueltas en el tiempo fijado con el objetivo de entregarle al cliente un producto de calidad. Con la aplicación de las listas de chequeo a los artefactos de ETL y los casos de pruebas basados en casos de uso se logró probar que el MD Contabilidad garantiza la calidad de la solución y que los artefactos cuentan con la estructura correspondiente.

### Conclusiones

Al concluir la investigación, se puede plantear que fueron cumplidos los objetivos trazados y las tareas de la investigación propuestas.

- En el marco teórico de la investigación se seleccionó la metodología propuesta por el Departamento de AD, la cual guió adecuadamente todo el proceso de desarrollo. Además se definieron las herramientas necesarias a utilizar en el MD. Tanto la metodología como las herramientas ya han sido utilizadas en el departamento y cumplen con las características pertinentes para el buen desarrollo del MD.
- En el análisis y diseño efectuado al MD se identificaron los requisitos funcionales, no funcionales y los informativos, logrando cumplir con las necesidades del cliente.
- La implementación de las estructuras modeladas permitió la integración de los datos, su almacenamiento y la visualización de la información en forma de vistas de análisis.
- Las pruebas realizadas al mercado haciendo uso del modelo V, mediante la utilización de las herramientas de prueba listas de chequeo y casos de prueba, arrojaron un número pequeño de no conformidades las cuales ya han sido resueltas.



### **Recomendaciones**

- Desplegar el Mercado de Datos desarrollado en la ONEI.
- Proporcionar cursos de capacitación a los especialistas de la ONEI, para que puedan trabajar con la aplicación satisfactoriamente.
- Incluir nuevas formas de visualización de la información y aplicar otras técnicas de inteligencia de negocios como cuadro de mando integral.

### Referencias Bibliográficas

1. **Publicación del Instituto de Contabilidad, dependiente de la Facultad de Ciencias Económicas y de Administración, de la Universidad Católica de Córdoba.** [En línea] 2010. [http://www.ucc.edu.ar/portalucc/archivos/File/FACEA/Instituto\\_de\\_Contabilidad/revista-n2-inst-contabilidad.pdf](http://www.ucc.edu.ar/portalucc/archivos/File/FACEA/Instituto_de_Contabilidad/revista-n2-inst-contabilidad.pdf).
2. Design of DW. [En línea] [Citado el: 15 de 11 de 2012.] [https://dspace.ist.utl.pt/bitstream/2295/630366/1/licao\\_6.pdf](https://dspace.ist.utl.pt/bitstream/2295/630366/1/licao_6.pdf).
3. **Empresas, BIPARA.** [En línea] 3 de 10 de 2008. [Citado el: 15 de 11 de 2012.] <http://www.bi-argentina.com.ar/que-es-un-data-warehouse>.
4. **Velasco, Roberto Hernando.** Almacén de Datos (Data Warehouse). [En línea] 2008. [Citado el: 16 de 11 de 2012.] <http://www.rhernando.net/modules/tutorials/doc/bd/dw.html>.
5. **Inmon, Bill.** Building the Data Warehouse.
6. **Jiménez, Margarita.** [En línea] 2012. [Citado el: 18 de 11 de 2012.] <http://www.google.com/cu/url?sa=t&rct=j&q=MINISTERIO+DE+ECONOM%C3%8DA%2C+PLANIFICACI%C3%93N+Y+DESARROLLO+OFICINA+NACIONAL+DE+ESTAD%C3%8DSTICA+Santo+Domingo%2C+Rep%C3%BAblica+Dominicana&source=web&cd=2&ved=0CDQQFjAB&url=http%3A%2F%2Fwww.one.gob.do%2Findex>.
7. **Calzada, Leticia.** [En línea] 2009. [Citado el: 16 de 11 de 2012.] [www.spentamexico.org/v4-n2/4\(2\)%2016-52.pdf](http://www.spentamexico.org/v4-n2/4(2)%2016-52.pdf).
8. **Navarro, Jorge Rubio y Salinas, José Manuel.** [En línea] 2003. [Citado el: 16 de 11 de 2012.] [www.turismo.uma.es/turitec/turitec/actas/2002/3.pdf](http://www.turismo.uma.es/turitec/turitec/actas/2002/3.pdf).
9. **Jiménez, Margarita.** [En línea] 2012. [Citado el: 18 de 11 de 2012.] <http://www.google.com/cu/url?sa=t&rct=j&q=MINISTERIO+DE+ECONOM%C3%8DA%2C+PLANIFICACI%C3%93N+Y+DESARROLLO+OFICINA+NACIONAL+DE+ESTAD%C3%8DSTICA+Santo+Domingo%2C+Rep%C3%BAblica+Dominicana&source=web&cd=2&ved=0CDQQFjAB&url=http%3A%2F%2Fwww.one.gob.do%2Findex>.
10. **Tabla de Hechos y Dimensiones.** [En línea] 2009. <http://msdn.microsoft.com/es-es/library/ms244679%28v=vs.80%29.aspx>.
11. **BD OLTP y OLAP.** [En línea] 2009. [http://www.sinnexus.com/business\\_intelligence/olap\\_vs\\_oltp.aspx](http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx).

12. **Arley, Ricardo Chinchilla.** MD: conceptos y metodologías de desarrollo. [En línea] Julio-septiembre de 2011. [Citado el: 18 de 11 de 2012.] <http://www.kerwa.ucr.ac.cr/bitstream/handle/10669/619/TM%2024-3%20art%206.pdf?sequence=1>.
13. **Ramos, Ing. Jonathan D. Nima.** Scribd. *Elaboración de soluciones de Inteligencia de Negocios.* [En línea] [Citado el: 19 de 11 de 2012.] <http://es.scribd.com/doc/61857452/8/g-CARACTERISTICAS-DE-UN-DATA-MART>.
14. **Rodolfo Quispe-Otazu** “Crear y usar reglas de Negocio” *Diciembre 2008.* Disponible en: <http://msdn.microsoft.com/es-es/library/aa577691%28v=bts.10%29.aspx>
15. **Hernández, Ing. Yanisbel González.** *PROPUESTA DE METODOLOGIA PARA EL DASARROLLO DE AD EN DATEC.* La Habana : s.n., 2010.
16. **Espinosa, Roberto.** El Rincon del Bl. [En línea] 1 de 5 de 2010. [Citado el: 22 de 11 de 2012.] <http://churriwifi.wordpress.com/2010/05/01/16-procesos-etl-escenarios-para-el-diseno-de-los-procesos/>.
17. **Guzmán, Elizabeth León.** Scribd. [En línea] [Citado el: 25 de 11 de 2012.] <http://es.scribd.com/doc/62869760/Intro-Bi>.
18. **Yuliana, Giraldo Luis. Zapata.** *Herramientas de desarrollo de Ingeniería de SW para Linux.* [En línea] 2008. [Citado el: 25 de 11 de 2012.] [http://hugolopez.phi.com.co/docs/download/file=Giraldo-Zapata-Herramientas%20de%20ISW.pdf\\_id=17](http://hugolopez.phi.com.co/docs/download/file=Giraldo-Zapata-Herramientas%20de%20ISW.pdf_id=17).
19. **Garzón, Raquel Boullón.** Intitec. [En línea] 2009. [www.intitec.com/varios/ComparativaSQL.pdf](http://www.intitec.com/varios/ComparativaSQL.pdf).
20. **Quiñones, Ernesto.** apesaol.org. [En línea] [Citado el: 27 de 11 de 2012.] [http://www.postgresql.org.pe/articulos/introduccion\\_a\\_postgresql.pdf](http://www.postgresql.org.pe/articulos/introduccion_a_postgresql.pdf).
21. pgAdmin PostgreSQL Tools. [En línea] 2012. [Citado el: 27 de 11 de 2012.] <http://www.pgadmin.org/>; [http://rpm.pbone.net/index.php3/stat/4/idpl/17347092/dir/fedora\\_16/com/pgadmin3-1.14.0-1.fc16.x86\\_64.rpm.html](http://rpm.pbone.net/index.php3/stat/4/idpl/17347092/dir/fedora_16/com/pgadmin3-1.14.0-1.fc16.x86_64.rpm.html).
22. **McGilvray, D.** Executing Data Quality Projects, from Science Direct. [En línea] 2008. [Citado el: 30 de 11 de 2012.] <http://ds.freecode.com/projects/datacleaner>.
23. DataCleaner. [En línea] 2012. [Citado el: 3 de 12 de 2012.] <http://datacleaner.org/>.
24. DataCleaner, una sencilla opción para mantener las BD ordenadas . [En línea] 2012. [Citado el: 4 de 12 de 2012.] <http://www.genbeta.com/herramientas/datacleaner-una-sencilla-opcion-para-mantener-las-bases-de-datos-ordenadas>.
25. **Pillasagua, Flor Maricela Pibaque.** [En línea] 2011.

- <http://repositorio.cisc.ug.edu.ec/bitstream/123/67/1/TOMO%201.pdf>.
26. Pentaho Analysis Services. [En línea] 2009. [Citado el: 3 de 12 de 2012.]  
<http://www.dataprix.com/empresa/productos/pentaho-bi-suite>.
27. **Open source Business Intelligence.** DESCRIPCION DE LA SUITE ENTERPRISE,Pentaho BI Suite. [En línea] 2009. [Citado el: 3 de 12 de 2012.]  
[http://www.cognus.cl/media/users/1/92208/files/10961/Pentaho\\_BI\\_Suite.pdf](http://www.cognus.cl/media/users/1/92208/files/10961/Pentaho_BI_Suite.pdf).
28. **Comparativa B.I. Open Source.**Open Busines Intelligence. [En línea] 6 de 2010.  
[http://www.stratebi.es/todobi/jun10/Comparativa\\_OSBI.pdf](http://www.stratebi.es/todobi/jun10/Comparativa_OSBI.pdf).
29. AUQUI, D. O. (Enero de 2013). *Portal de Tecnologías Aplicadas a Negocios*. Recuperado el Enero de 2013, de [pis2.wikispaces.com/file/view/TESIS+II+-+UGEL+04.doc](http://pis2.wikispaces.com/file/view/TESIS+II+-+UGEL+04.doc)
30. **Enrique J. Abril Pérez.** *Disponible en:* <http://zarza.usal.es/~fgarcia/doctorado/iweb/05-07/Trabajos/CalidadAplicsWeb.pdf>
31. **Martha Rojas Vera.** “Ciclo de vida Modelo en V” Julio 2010. *Disponible en:* <http://spanishpmo.com/index.php/ciclos-de-vida-modelo-en-v/>

## Bibliografía

1. **Berzal, Fernando.** El modelo multidimensional. [En línea] <http://elvex.ugr.es/idbis/db/docs/intro/F%20Modelo%20multidimensional.pdf>.
2. **Castro, Hugo M.** [En línea] <http://www.google.com/cu/url?sa=t&rct=j&q=variables+aditivas,+semiaditivas&source=web&cd=5&ved=0CEEQFjAE&url=http%3A%2F%2Fmateria-caece.googlecode.com%2Fsvn%2Ftrunk%2FAdministracion%2520del%2520Conocimiento%2FDocumentos%2520Foro%2FPresentaciones%2FDWh.Mod>.
3. **community, Pentaho.** Pentaho community. [En línea] <http://wiki.pentaho.com/display/EALes/.07+Variables>.
4. **Isabel Dapena Bosquet, Antonio Muñoz San Roque , Álvaro Sánchez Miralles.** Sistemas de Información Orientada a la toma de decisiones: el enfoque multidimensional. [En línea] 2005. [www.icaei.es/publicaciones/anales\\_get.php?id=470](http://www.icaei.es/publicaciones/anales_get.php?id=470).
5. **Kimball, Ralph.** *The Data Warehouse ETL Toolkit*. Canadá : Wiley Publishing, 2004.
6. **Pentaho Data Integration.** [En línea] 2005-2013. <http://www.pentaho.com/explore/pentaho-data-integration/>.
7. **Pressman, Roger S.** *Ingeniería de Software. Un Enfoque práctico*. La Habana : Felix Varela, 2005.
8. **Velázquez, Eugenio.** ¿Que es la inteligencia de negocios (Business Intelligence)? [En línea] <http://www.tecnologiapyme.com/administracion-electronica/que-es-la-inteligencia-de-negocios-business-intelligence>.
9. **Zepeda, Leopoldo Zenaido.** Universidad Politécnica de Valencia. departamento de sistemas Informáticos y Computación. [En línea] Junio de 2008. <http://riunet.upv.es/bitstream/handle/10251/2506/tesisUPV2841.pdf>.
10. **Suárez, José Carlos Puente.** Oficina Nacional de Estadística e Información República de Cuba. [En línea] <http://http://www.one.cu/>.

**HTML:** HTML, siglas de HyperText Markup Language (Lenguaje de Marcado de Hipertexto), es el lenguaje de marcado predominante para la elaboración de páginas web. Es usado para describir la estructura y el contenido en forma de texto, así como para complementar el texto con objetos tales como imágenes

**UML:** Lenguaje Unificado de Modelado (LUM) o (UML, por sus siglas en inglés, Unified Modeling Language) es el lenguaje de modelado de sistemas de software más conocido y utilizado en la actualidad; está respaldado por el OMG (Object Management Group)

**TCP/IP:** son las siglas de Protocolo de Control de Transmisión/Protocolo de Internet (por el inglés Transmission Control Protocol/Internet Protocol), un sistema de protocolos que hacen posibles servicios Telnet, FTP, E-mail, y otros, entre ordenadores que no pertenecen a la misma red. TCP garantiza la entrega de datos y que los paquetes sean entregados en el mismo orden en el cual fueron enviados. IP utiliza direcciones que son series de cuatro números octetos (byte) con un formato de punto decimal.

**Área de Análisis:** agrupación de información según su propósito, aunque el criterio depende de las necesidades de la institución o empresa donde se aplica el sistema. Permite restringir el número de usuarios que acceden a los datos.

**No conformidades:** defecto, error o sugerencia que se le hace al equipo de desarrollo cuando se encuentra alguna dificultad en lo que se está evaluando.

**OLAP:** es el acrónimo en inglés de procesamiento analítico en línea (On-Line Analytical Processing). Es una solución utilizada en el campo de inteligencia de negocio, cuyo objetivo es agilizar la consulta de grandes cantidades de datos.

**Almacén de datos:** gran repositorio lógico de datos que permite el acceso y la manipulación flexible de grandes volúmenes de información procedente de diferentes fuentes.

**BI (Inteligencia de negocio):** habilidad corporativa para tomar decisiones mediante el uso de metodologías, herramientas y tecnologías que permiten aplicar en ellos técnicas analíticas de extracción de información.

**ETL (Extracción, Transformación y Carga):** extrae los datos desde los sistemas fuentes, los transforma y luego los carga al almacén o mercado de datos.