

# **Universidad de las Ciencias Informáticas**

## **Facultad 6**



**Título:** Mercado de Datos Agricultura, ganadería y silvicultura para el Sistema de Información de Gobierno

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

**Autoras:** Yelena Islén San Juan

Yudith Recio Milanés

**Tutor:** Ing. José Salvador Bermúdez Rodríguez

Ciudad de La Habana, Junio de 2011



*“Nuestra recompensa se encuentra en el esfuerzo y no en el resultado. Un esfuerzo total es una victoria completa”*

*Mahatma Gandhi*

*1869- 1948 Político y pensador Indio*

**DECLARACIÓN DE AUTORÍA**

Declaramos ser las únicas autoras del presente trabajo y autorizamos a la Facultad 6 de la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

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

Yelena Islen San Juan

\_\_\_\_\_

Firma del Autor

Yudith Recio Milanés

\_\_\_\_\_

Firma del Autor

Ing. José Salvador Bermúdez Rodríguez

\_\_\_\_\_

Firma del Tutor

## **DATOS DE CONTACTO**

### **Tutor:**

Tutor: Ing. José Salvador Bermúdez Rodríguez

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Categoría docente: Trabajador no docente, cargo Recién graduado en adiestramiento

Categoría científica: -

Años de experiencia en el tema: 1

Años de graduado: 1

Correo electrónico: [jsbermudez@uci.cu](mailto:jsbermudez@uci.cu)

Jefe del grupo Análisis del departamento Almacenes de Datos (AD) en el Centro de Tecnologías de Gestión de Datos (DATEC). Su Tesis de Diploma estuvo relacionada con el tema del desarrollo de AD. Presentó dicha investigación en eventos científicos a nivel de Facultad y Universidad obteniendo resultados relevantes en ambos eventos.

- *A Fidel Castro y a la Revolución por brindarnos la oportunidad de estudiar y de convertirnos en lo que hoy somos.*

- *A nuestros padres, por darnos todo el amor, el apoyo, la confianza y la dedicación que un hijo puede desear. Por ser nuestro ejemplo a seguir y demostrarnos día a día que toda meta con esfuerzo se puede alcanzar.*

- *A todos los profesores que contribuyeron a nuestra formación como ingenieros, durante estos cinco años.*

- *A todos los compañeros que a lo largo de estos años han compartido junto a nosotras. Por tantos recuerdos inolvidables.*

- *A todos aquellos amigos que son parte insustituible de nuestros mejores años como estudiantes y que siempre estuvieron presente apoyándonos en las buenas y en las malas: Delmis, Maryin, Yenisleidy, Yurislaine, Jorge Esperanza, Juan José, Dino, Yoendy, Manrresa, Rancel, Yaimé, Yicenia, Dismey, Maibel, Raiza, Edelqui, Lazaro, Dayana, Nidelso, Ismaraysi y Gustavo.*

- *A todos los que de una forma u otra han aportado en nuestra preparación como jóvenes comprometidos con la Revolución para hacer*

*este sueño realidad: Rosayda, Alberto Garnachi, Yanelis Benites, Lesley Yanelis Ramirez, Alien, Pedro.*

- *A nuestro tutor por su disposición y apoyo incondicional con nosotras para ayudarnos a lograr lo que hoy es nuestro trabajo de diploma, y por habernos mostrado el camino correcto.*

- *A todos los trabajadores del departamento de Almacenes de Datos del Centro de Tecnología de Gestión de Datos, en especial a José Salvador, Themis, Marisel, Yosvel por las horas dedicadas en aras de lograr los objetivos propuestos.*

- *Y sobre todas las cosas, agradecemos a Dios, nuestro Señor por darnos fuerzas como las águilas y todo cuanto tenemos y somos.*

*Yelena Islen y Yudith Rocio*

*Dedico el presente Trabajo de Diploma a todas las personas que más amo:*

- *A mis padres por ser dos luceros en mi vida, formarme y educarme con tanto amor y dedicación; por guiar mis pasos con sus consejos para la vida cotidiana, dándome siempre fuerzas para salir adelante.*
- *A mi hermana por ser mi amiga, consejera y darme tanto amor.*
- *A mi futuro esposo por su cariño y comprensión en todo este tiempo.*
- *En especial a mi padre Celestial por darme tan bella familia.*

*Yelena Islen San Juan*

*Dedico mi trabajo de diploma principalmente a las personas que más amo en el mundo:*

- *A mi madre por ser mi apoyo en todo momento y darme su amor y comprensión siempre. Por ser la persona que más quiero ya que me ha dado la dicha de nacer y crecer bajo todo su amor. Y por demostrarme que en la vida todo llega, sólo hay que esperar.*
- *A mi padre Vicente por ser una de las personas más especiales en mi vida ya que ha sabido aconsejarme, apoyarme y quererme como soy. Y ha sido de la persona de la que más he aprendido a través de mi educación y mi formación como mejor ser humana.*
- *A mis hermanos Yherffry y Aldrick por ser mi hombro de apoyo para reír y llorar en todo momento y por cuidarme y enseñarme todo lo que saben. A mis tías y tíos que tanto los quiero.*
- *Dedico mi trabajo de diploma a mi otra familia Josefa y Zenaide por quererme tanto, a Idelsi y Leonel por enseñarme tantas cosas y toda su dedicación hacia mí, a Yanet y Juan Luis por todo su apoyo y cariño, a Arnaldo (Pipo) por tenerme como una hermana más, a todos muchas gracias por ser parte de mi familia.*
- *En especial a mi novio Jorge Luis por darme todo su amor, su comprensión y su cariño todo este año y por ser mi amigo en los momentos de alegría y de tristeza.*
- *A todos mis amigos que han estado conmigo durante toda mi vida desde mi infancia hasta hoy en especial Mailin, Nelson, Alietis, Saili, Yaime, Yicenia, Dismey, Ismaraisi, Lazaro y muchos más que quiero mucho.*
- *A todos los chicos del grupo 6103 que tanto los quiero y los tendré en mi corazón siempre.*
- *A todos los que de una forma u otra han estado en todo momento brindándome su apoyo y comprensión y que me han ayudado a lograr todos mis sueños.*

*Yudith Reçio Milanés*

La presente investigación surge como parte de la colaboración existente entre la Universidad de las Ciencias Informáticas y la Oficina Nacional de Estadísticas. Esta última es el órgano rector de la estadística en Cuba y la responsable de gestionar los principales indicadores en distintas áreas de nuestro país entre las que se encuentra la Agricultura, ganadería y silvicultura. En la actualidad, las herramientas que se utilizan para la recolección y gestión de la información en la Oficina Nacional de Estadísticas presentan deficiencias, influyendo negativamente en los datos estadísticos. En el presente Trabajo de Diploma se tiene como objetivo principal desarrollar un Mercado de Datos para el área de Agricultura, ganadería y silvicultura del Sistema de Información de Gobierno que contribuya al almacenamiento homogéneo de los datos, posibilitando un adecuado análisis de la información. Para dar cumplimiento a ello, se realizó una caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de los Almacenes de Datos. De igual manera se analizó y diseñó un Mercado de Datos para dicha área. Partiendo de los elementos de análisis y diseño se llevó a cabo la implementación, obteniéndose como resultado un Mercado de Datos que cumple con los requerimientos de información solicitados por los especialistas de la Oficina Nacional de Estadísticas. Por último, se validó el Mercado de Datos implementado aplicando el Modelo V, listas de chequeo y los casos de pruebas diseñados.

**PALABRAS CLAVES:**

Agricultura, Almacén de Datos, ganadería, Mercado de Datos, Oficina Nacional de Estadísticas, silvicultura.

**ÍNDICE**

**INTRODUCCIÓN ..... 1**

**Capítulo 1: Fundamentación teórica para el desarrollo de un Mercado de Datos ..... 6**

**1.1 Introducción ..... 6**

**1.2 Proceso de digitalización de la información que se almacena en el área de Agricultura, ganadería y silvicultura de la ONE ..... 6**

**1.3 Almacenes de Datos ..... 6**

        1.3.1 Mercado de Datos ..... 9

**1.4 Instituciones que utilizan Almacenes de Datos en el mundo ..... 14**

**1.5 Metodologías para el desarrollo de un Almacén de Datos ..... 15**

        1.5.1 Metodología Kimball vs Inmon ..... 15

        1.5.2 Modelo para el desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio en DATEC ..... 17

**1.6 Herramienta de modelado ..... 18**

**1.7 Técnicas de captura de requerimientos ..... 19**

**1.8 Herramientas informáticas para el proceso de Extracción, Transformación y Carga ..... 20**

**1.9 Herramientas informáticas para la inteligencia de negocios, aplicando técnicas OLAP ..... 20**

**1.10 Sistema Gestor de Bases de Datos ..... 22**

**1.11 Conclusiones del capítulo ..... 23**

**Capítulo 2: Análisis y diseño del Mercado de Datos Agricultura, ganadería y silvicultura ..... 24**

**2.1 Introducción ..... 24**

**2.2 Caracterización de las áreas de la organización ..... 24**

**2.3 Necesidades de información ..... 25**

**2.4 Reglas del negocio ..... 25**

**2.5 Especificación de requerimientos ..... 26**

        2.5.1 Requerimientos de información ..... 26

        2.5.2 Requerimientos funcionales ..... 28

        2.5.3 Requerimientos no funcionales ..... 29

**2.6 Modelo de casos de uso del sistema ..... 29**

        2.6.1 Actores del sistema ..... 29

---

2.6.2	Diagrama de Casos de Uso del sistema .....	30
2.6.3	Especificación de Casos de Uso del Sistema .....	30
2.7	Especificación del modelo dimensional .....	32
2.7.1	Matriz dimensional .....	33
2.7.2	Tablas de dimensiones .....	34
2.7.3	Tablas de hechos .....	34
2.7.4	Modelo dimensional .....	35
2.8	Especificación del modelo físico .....	37
2.9	Conclusiones del capítulo .....	37
<b>Capítulo3: Implementación del Mercado de Datos Agricultura, ganadería y silvicultura .....</b>		<b>38</b>
3.1	Introducción .....	38
3.2	Diseño del subsistema de integración.....	38
3.2.1	Registro de sistemas fuentes .....	38
3.3	Implementación del subsistema de integración .....	39
3.3.1	Implementación de los flujos de transformación .....	40
3.3.2	Implementación de los trabajos .....	41
3.4	Diseño del subsistema de visualización .....	41
3.4.1	Arquitectura de información.....	42
3.4.2	Diseño de los reportes candidatos.....	42
3.5	Implementación del subsistema de visualización .....	43
3.5.1	Implementación de los cubos OLAP.....	43
3.5.2	Implementación de los reportes candidatos .....	44
3.6	Configurar la seguridad de los usuarios .....	44
3.7	Conclusiones del capítulo.....	45
<b>Capítulo 4: Validación del Mercado de Datos Agricultura, ganadería y silvicultura .....</b>		<b>46</b>
4.1	Introducción .....	46
4.2	Gestión de la Calidad de Software.....	46
4.3	Casos de prueba .....	48
4.4	Elaboración y evaluación de las listas de chequeo .....	49
4.5	Resultados de las pruebas.....	49
4.6	Conclusiones del capítulo.....	50

<b>Conclusiones</b> .....	<b>51</b>
<b>Recomendaciones</b> .....	<b>52</b>
<b>Bibliografía</b> .....	<b>53</b>
<b>Referencias bibliográficas</b> .....	<b>56</b>
<b>Anexos</b> .....	<b>59</b>
<b>Glosario</b> .....	<b>61</b>

**Tablas de figuras**

Figura 1: Organizado en torno a temas.....	7
Figura 2: Integrado.....	7
Figura 3: No volátil.....	8
Figura 4: Dependiente del tiempo .....	8
Figura 5: Modelo de almacenamiento ROLAP .....	10
Figura 6: Modelo de almacenamiento MOLAP.....	11
Figura 7: Modelo de almacenamiento HOLAP .....	11
Figura 8: Procesamiento de transacciones .....	12
Figura 9: Ciclo de vida Kimball.....	16
Figura 10: Proceso del negocio.....	25
Figura 11: Diagrama de Casos de Uso del Sistema.....	30
Figura 12: Dimensión indicadores ganadería .....	34
Figura 13: Hecho información ganado .....	35
Figura 14: Modelo Dimensional.....	36
Figura 15: Transformación del hecho información ganado.....	41
Figura 16: Trabajo.....	41
Figura 17: Mapa de navegación.....	42
Figura 18: Cubo del hecho información ganado.....	44
Figura 19: Reporte Candidato.....	44
Figura 20: Modelo V.....	48
Figura 21: Casos de prueba Ganadería .....	49

## **INTRODUCCIÓN**

Las ansias de mejora continua han logrado importantes avances socio-económicos para la humanidad, ejemplo de ello es la informática, ciencia definida como el procesamiento automático de la información utilizando dispositivos electrónicos y sistemas computacionales. La informática ha sido una de las áreas del conocimiento que mayor desarrollo ha adquirido en los últimos tiempos. Casi la totalidad de las disciplinas científicas requieren de ella, aplicándose a todos los aspectos de la existencia humana.

Ante el cúmulo de información y la necesidad de utilizar nuevos métodos y herramientas que permitiesen la consulta de información relevante y aparejada al incremento insostenible en el poder de cálculo, dio lugar al desarrollo de la industria del software, convirtiéndose en un eslabón esencial para potenciar la economía y otras esferas de la vida social como la estadística. En la actualidad, existe una estrecha relación entre la estadística y la industria del software. Dicha correspondencia se evidencia mediante el uso de herramientas enfocadas al análisis estadísticos de los datos; de esta forma, se evitan los problemas que presentan las organizaciones que hacen uso de gran información para la toma de decisiones. Hoy día las estadísticas abarcan todas las esferas de un país anteponiéndose a los problemas, es por esta razón que se ha convertido en un eslabón importante de la economía.

Con el fin de garantizar la estadística en Cuba y responder a las necesidades de información, se creó la Oficina Nacional de Estadísticas (ONE) en 1994. Su principal objetivo es captar, analizar y difundir los datos recopilados en todo el país. Para ello cuenta con una serie de modelos estadísticos (Excel, "dbf") los cuales recogen información de todos los sectores de la economía y la sociedad.

Este centro cuenta con 15 delegaciones territoriales y 168 oficinas municipales en todo el país subordinadas a las oficinas provinciales. Todas las oficinas son atendidas administrativamente y metodológicamente por la ONE. La información estadística de Cuba se integra en el Sistema Estadístico Nacional (SEN) y para ello se agrupa en tres subsistemas fundamentales:

Sistema de Información Estadístico Nacional (SIEN): comprende la elaboración de estadísticas y análisis destinados a satisfacer las necesidades informativas del estado y del gobierno.

Sistema de Información Estadística Territorial (SIET): comprende la elaboración de las estadísticas destinadas a satisfacer los requerimientos informativos de los órganos del estado y gobierno a nivel territorial, por lo que este sistema consta de quince subsistemas independientes.

Sistema de Información Estadística Complementaria (SIEC): comprende la elaboración de las estadísticas y análisis destinados a satisfacer las necesidades informativas de los órganos, organismos e instituciones estatales a los efectos del control administrativo de sus entidades.

Tras varios años de esfuerzo y sacrificio como parte del programa de informatización de la sociedad surge la Universidad de las Ciencias Informáticas (UCI), enfocada a fomentar la industria de software en Cuba. La vinculación docencia-producción es uno de los objetivos fundamentales de dicho centro, desarrollando en diversas líneas de software productos para clientes nacionales y extranjeros. En la Facultad 6 de la universidad antes mencionada se encuentra el Centro de Tecnologías de Gestión de Datos (DATEC), el cual mantiene una colaboración con distintas instituciones de la capital, ejemplo de ello la ONE. Este centro presenta deficiencias en la gestión de los datos en el área de Agricultura, ganadería y silvicultura como:

- Utilización de herramientas de recolección de información basada en aplicaciones informáticas de oficina (Excel, Word, dbf).
- La información no pueden ser consultada a no ser por un especialista de la informática y de la información con alto conocimiento del negocio: algunas de las fuentes están codificadas de forma que si no se es conocedor de la terminología usada (por especialistas de la información) no puede comprenderse el contenido de estas. Por otra parte, los datos de las fuentes pueden extraerse de forma manual o por un informático.
- Se generan ficheros anuales con los cuales se hace muy difícil la obtención de información: la información va acumulándose año tras año y cada vez resulta más complicado analizarla.
- Información no integrada e inconsistente: la integración se refiere en este caso al hecho de existir referencias a la misma información que usan diferente codificación, diferente cantidad de caracteres, entre otros. Ocurre porque existen múltiples versiones de los mismos datos.
- Proceso de recuperación y elaboración de informes costosos en esfuerzo y tiempo.

Todas estas deficiencias provocan un engorroso análisis de los datos influyendo negativamente en el análisis estadístico de distintas variables en el área de Agricultura, ganadería y silvicultura, aumentando las probabilidades de errores estadísticos.

La búsqueda de mejoras en las formas de almacenar, recuperar y presentar la información proveniente de los organismos, tales como principales reportes, cruces de variables, indicadores, porcentajes y demás aspectos de interés es una necesidad urgente para aumentar la disponibilidad de información y el proceso de toma de decisiones en el área de Agricultura, ganadería y silvicultura.

Por toda la situación anteriormente descrita, se plantea como **problema de la investigación**: el almacenamiento de la información en diferentes formatos del área de Agricultura, ganadería y silvicultura, dificulta su adecuado análisis.

La investigación tiene como **objeto de estudio** los Almacenes de Datos, enmarcado en el **campo de acción** Mercado de Datos para el área de Agricultura, ganadería y silvicultura del Sistema de Información de Gobierno.

El **objetivo general** de este trabajo es desarrollar el Mercado de Datos Agricultura, ganadería y silvicultura para el Sistema de Información de Gobierno que contribuya al almacenamiento homogéneo de los datos, posibilitando un adecuado análisis de la información.

En correspondencia con el objetivo general, se plantean como **objetivos específicos**:

1. Realizar el análisis y diseño del Mercado de Datos.
2. Implementar el Mercado de Datos.
3. Validar el Mercado de Datos.

Para el cumplimiento de estos objetivos se realizaron esencialmente las siguientes **tareas investigativas**:

1. Caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de los Almacenes de Datos para su posterior análisis.
2. Levantamiento de requerimientos para la identificación de requerimientos de información, funcionales y no funcionales.
3. Descripción de los casos de uso del Mercado de Datos para comprender el flujo de actividades entre el actor y el sistema.
4. Definición de los hechos, las medidas y las dimensiones del Mercado de Datos para el diseño del modelo de datos.
5. Diseño del modelo de datos para recoger procesos fundamentales como: tablas de hechos y dimensiones candidatas en la solución del problema.
6. Definición de la arquitectura del Mercado de Datos para establecer una estructura adecuada.
7. Diseño del subsistema de integración para centralizar la información de las distintas fuentes de origen en el Mercado de Datos.
8. Diseño del subsistema de visualización para definir los reportes candidatos que serán de

- utilidad para el usuario.
9. Diseño de los casos de pruebas para que los reportes candidatos respondan a las necesidades del usuario final.
  10. Implementación del modelo de datos para cumplir con el diseño modelado.
  11. Implementación del subsistema de integración para poblar el Mercado de Datos.
  12. Implementación del subsistema de visualización para la administración de los reportes.
  13. Aplicación de las listas de chequeo para validar la solución.
  14. Aplicación de los casos de pruebas para validar que los reportes candidatos respondan a las necesidades del usuario final.

El Trabajo de Diploma está estructurado de la siguiente manera: introducción, cuatro capítulos, conclusiones, recomendaciones, referencias bibliográficas, bibliografía, anexos y glosario de términos.

### **Capítulo 1: Fundamentos teóricos para el desarrollo de un Mercado de Datos**

En el capítulo se hace un análisis del estado del arte del objeto de estudio y la gestión de la información en el área de Agricultura, ganadería y silvicultura del Sistema de Información de Gobierno. Se hace una caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo del Mercado de Datos.

### **Capítulo 2: Análisis y diseño del Mercado de Datos Agricultura, ganadería y silvicultura**

En el capítulo se realiza el análisis y diseño de los datos del área Agricultura, ganadería y silvicultura del Sistema de Información de Gobierno. Para lograr esto se modela el negocio, se hace el levantamiento de los requerimientos y se diseña el modelo de datos.

### **Capítulo 3: Implementación del Mercado de Datos Agricultura, ganadería y silvicultura**

En el capítulo se diseñan e implementan los subsistemas de integración y visualización del área de Agricultura, ganadería y silvicultura del Sistema de Información de Gobierno. Para lograr esto se definen las áreas de análisis con los libros de trabajo correspondientes. Además se diseñan e implementan los reportes candidatos, flujos de transformaciones, trabajos y cubos OLAP.

#### **Capítulo 4: Validación del Mercado de Datos Agricultura, ganadería y silvicultura**

En el capítulo se utiliza el Modelo V para la validación del Mercado de Datos, donde se realizan pruebas unitarias, de integración y de aceptación con el cliente. Se diseñan y aplican casos de prueba y listas de chequeo para validar la solución.

## **Capítulo 1: Fundamentación teórica para el desarrollo de un Mercado de Datos**

### **1.1 Introducción**

En el presente capítulo se hace un análisis del estado del arte del objeto de estudio y la gestión de la información en el área de Agricultura, ganadería y silvicultura del Sistema de Información de Gobierno. Se hace una caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo del Mercado de Datos.

### **1.2 Proceso de digitalización de la información que se almacena en el área de Agricultura, ganadería y silvicultura de la ONE**

La ONE utiliza herramientas de aplicaciones de oficinas como Microsoft Office y Bases de Datos (BD) para el almacenamiento de la información. Los modelos 0336-01 y 0467-06 recogen información referente a la agricultura, ganadería y silvicultura. El modelo 0336-01 o Sector Agropecuario y el modelo 0467-06 o Sector Silvicultura se conforman con los datos enviados por los organismos. Estas herramientas no satisfacen la necesidad existente en la institución debido al gran cúmulo de información almacenada por lo que se hace necesario poseer una herramienta potente. (1)

La respuesta a esto fue la creación de un Mercado de Datos (MD), el cual unirá las distintas fuentes de información de las cuales dispone la institución en un único lugar y al que sólo se le incorporará información relevante sobre la base de una estructura integrada. Esta solución posibilitará una gestión más eficiente, proyectando una visión única de la realidad de la organización y contribuyendo a la competitividad. Otras de las ventajas que traerá consigo es mejorar la productividad y aceptación para el área Agricultura, ganadería y silvicultura por sus disímiles características, así como su enfoque a un área de negocios específica, contribuyendo a la toma de decisiones.

### **1.3 Almacenes de Datos**

ALMACÉN DE DATOS. CONCEPTO.
Un Almacén de Datos (AD), Bodega de Datos o Datawarehouse es una gran colección de datos, que recoge información de múltiples sistemas fuentes u operacionales dispersos y cuya actividad se centra en la toma de decisiones, es decir, en el análisis de la información en vez de en su captura. (2)

**Tabla 1: Concepto de Almacenes de Datos**

## Características de un Almacén de Datos

Según las características que definen a un AD se puede plantear que son: (3)

### ➤ Organizado en torno a temas

Los datos están almacenados por materias o temas. Estos se organizan desde la perspectiva del usuario final, mientras que en las bases de datos operacionales se organizan desde la perspectiva de la aplicación, con vistas a lograr una mayor eficiencia en el acceso a los datos.

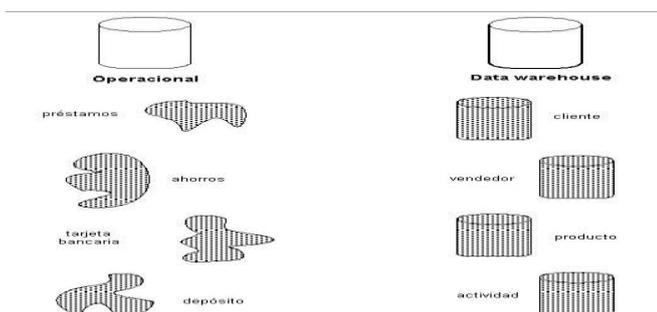


Figura 1: Organizado en torno a temas

### ➤ Integrado

Los datos almacenados en el AD están integrados. La integración de datos consiste en convenciones de nombres, codificaciones consistentes, medida uniforme de variables, entre otros.

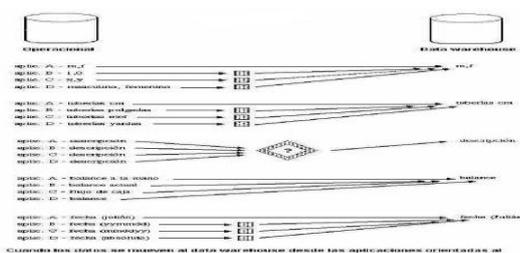


Figura 2: Integrado

### ➤ No volátil

El AD sólo permite cargar nuevos datos y acceder a los ya almacenados, pero no permite borrar ni modificar los datos.

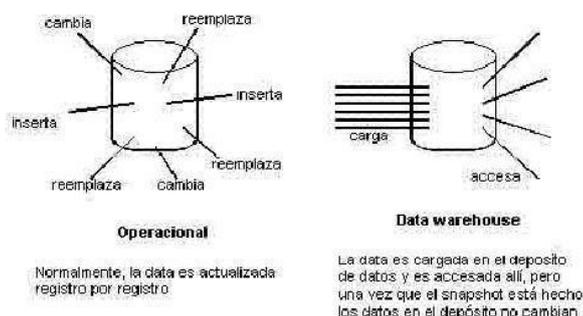


Figura 3: No volátil

➤ **Dependiente del tiempo**

El tiempo debe estar presente en todos los registros contenidos en un AD. El AD contiene información actual y resúmenes de esta en el tiempo. La información una vez registrada correctamente no puede ser actualizada. Cada estructura clave contiene (implícita o explícitamente) un elemento de tiempo (día, semana, mes).



Figura 4: Dependiente del tiempo

**Ventajas y desventajas del uso de un Almacén de Datos**

• **Ventajas**

Con el uso de AD los procesos empresariales pueden ser optimizados, ya que contribuyen a la toma de decisiones estratégicas proporcionando un sentido automatizado para identificar información clave desde volúmenes de datos generados por procesos tradicionales o elementos de software procedentes de múltiples sistemas, permitiendo esto dar a los usuarios prioridad a decisiones y acciones, lo cual facilita reaccionar rápidamente ante los cambios del mercado, aumentando la competitividad en el mercado y las ganancias respectivamente de la empresa. (4)

➤ **Desventajas**

Son grandes los beneficios que proporcionan los AD; pero también acarrear inconvenientes de gran escala ya que éstos no son generalmente estáticos por lo que a lo largo de su vida su mantenimiento es muy elevado. Los AD pueden quedar obsoletos relativamente pronto y a veces ante una petición de información devuelven una información que puede no ser la más óptima posible; lo que también supone una pérdida para la organización. (5)

**1.3.1 Mercado de Datos**

MERCADO DE DATOS. CONCEPTOS.
1. Un Mercado de Datos es una Base de Datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. (6)
2. Según define Meta Group, un MD es una aplicación de Datawarehouse (DWH), construida rápidamente para soportar una línea de negocio simple. (7)

**Tabla 2: Conceptos de Mercado de Datos**

El uso efectivo de los MD es un factor importante para la efectividad del AD y puede también ser determinante en el éxito del proyecto de desarrollo. Los MD son diseñados para satisfacer las necesidades específicas de grupos comunes de usuarios (divisiones organizacionales). Son generalmente subconjuntos del AD, pero pueden también integrar un número de fuentes heterogéneas e inclusive ser más grandes en volumen de datos que el propio AD central. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Como los MD son un factor crítico para el éxito del AD de mayor escala también lo son su creación y mantenimiento. (8)

Los MD representan una estrategia de "divide y vencerás" para ámbitos muy genéricos de un AD y presenta las mismas características que un AD:

Para crear el MD de un área funcional de la empresa es preciso encontrar la estructura óptima para el análisis de su información, estructura que puede estar montada sobre una BD de Procesamiento de Transacciones en Línea (On-line Transaction Processing, OLTP) o sobre una BD de Procesamiento de Análisis en Línea (On-Line Analytical Processing, OLAP). La designación de una u otra depende de los

datos, los requerimientos y las características específicas de cada departamento. De esta forma se pueden plantear dos tipos de MD: (9)

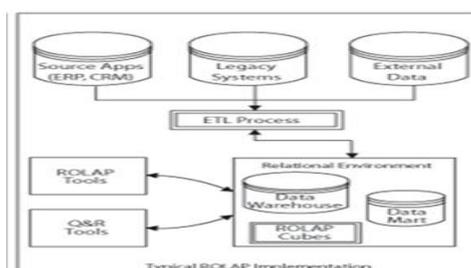
### **Mercado de Datos OLAP**

Solución utilizada en el campo de la llamada Inteligencia de Negocio (BI) basada en estructuras multidimensionales (cubos OLAP) y contienen datos resumidos de grandes BD o sistemas transaccionales (OLTP). Son BD orientadas al procesamiento analítico y este análisis suele implicar generalmente la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil. Se construye agregando según los requerimientos de cada área o departamento las dimensiones y los indicadores necesarios de cada cubo relacional. Provee de una tecnología para el cálculo y análisis requerido por las aplicaciones analíticas para la Inteligencia de Negocio. El modo de creación, explotación y mantenimiento de los cubos OLAP es muy heterogéneo, en función de la herramienta final que se utilice. (10)

### **Clasificación de los sistemas OLAP**

#### ➤ **Procesamiento Analítico Relacional (ROLAP)**

Implementación OLAP que almacena los datos en un motor relacional. Típicamente los datos son detallados, evitando las agregaciones y las tablas que se encuentran normalizadas. Los esquemas más comunes sobre los que se trabaja son estrella o copo de nieve, aunque es posible trabajar sobre cualquier BD relacional. La arquitectura está compuesta por un servidor de banco de datos relacional y el motor OLAP que se encuentra en un servidor dedicado. La principal ventaja de esta arquitectura es que permite el análisis de una enorme cantidad de datos.



**Figura 5: Modelo de almacenamiento ROLAP**

#### ➤ **Procesamiento Analítico Multidimensional (MOLAP)**

Esta implementación OLAP almacena los datos en una BD multidimensional para optimizar los tiempos

de respuesta y el resumen de la información es usualmente calculado por adelantado. Estos valores pre-calculados o agregaciones son la base de las ganancias de desempeño de este sistema. Algunos sistemas utilizan técnicas de compresión de datos para disminuir el espacio de almacenamiento en disco debido a los valores pre-calculados.

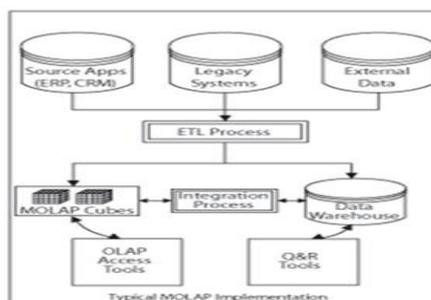


Figura 6: Modelo de almacenamiento MOLAP

#### ➤ Procesamiento Analítico Híbrido (HOLAP)

Es un híbrido entre los métodos ROLAP y MOLAP. Almacena algunos datos en un motor relacional y otros en una BD multidimensional.

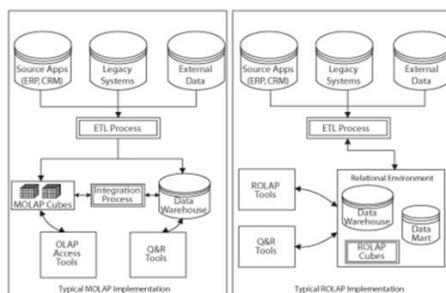


Figura 7: Modelo de almacenamiento HOLAP

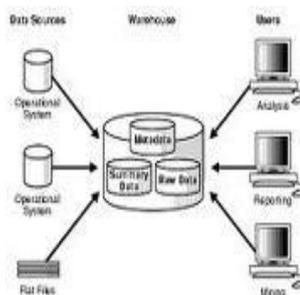
### Mercado de Datos OLTP

Los sistemas OLTP son BD orientadas al procesamiento de transacciones. Una transacción genera un proceso atómico y puede involucrar operaciones de inserción, modificación y borrado de datos. El proceso transaccional es típico de las BD operacionales. (11)

### Características de los sistemas OLTP

- El acceso a los datos está optimizado para tareas frecuentes de lectura y escritura.
- Los datos se estructuran según el nivel aplicación (programa de gestión a medida, Sistemas de Planificación de Recursos Empresariales (ERP) o Administración de Relación de Cliente (CRM) implantado, sistema de información departamental).

- Los formatos de los datos no son necesariamente uniformes en los diferentes departamentos (es común la falta de compatibilidad).
- El historial de datos suele limitarse a los datos actuales o recientes.



**Figura 8: Procesamiento de transacciones**

Se utiliza como modo de almacenamiento de datos al ROLAP debido a que el Sistema Gestor de Base de Datos (SGBD) a utilizar exige que sea este el modo de almacenamiento, además posee un ambiente conocido y disponibilidad de herramientas para su desarrollo, facilitando el desarrollo de la solución.

### Etapas de desarrollo de un Mercado de Datos

#### 1. Análisis y diseño

La función del análisis es dar soporte a las actividades de un negocio o desarrollar un producto que pueda venderse para generar beneficios, mientras que el diseño se encarga de aplicar ciertas técnicas y principios con el propósito de definir un dispositivo, un proceso o un sistema con suficientes detalles como para permitir su interpretación y realización física.

##### ➤ Pasos para el análisis y diseño de un Mercado de Datos:

1. Identificar temas de análisis: define los temas y los indicadores que serán analizados y explotados por los usuarios.
2. Identificar dimensiones de navegación: las dimensiones de navegación es la forma como el usuario podrá agrupar la información. Estas siempre deben responder una de estas seis preguntas: a quién, dónde, cuándo, qué, cómo y quién.
3. Elaboración del modelo multidimensional básico: con ayuda de alguna herramienta CASE deberá diseñar un modelo multidimensional capaz de soportar cualquiera de las consultas que los usuarios puedan realizarle en un futuro al AD.

4. Elaboración del modelo multidimensional complejo: en esta etapa se realiza el proceso de calificación a los indicadores y medidas.
5. Elaboración de las especificaciones de carga de datos: se busca la información en las diferentes fuentes de datos de la organización para cargar el modelo multidimensional creado.
6. Creación de base de datos: se crea una base de datos que en el AD contendrá la información del MD que será explotada.
7. Construcción de la arquitectura del MD dentro del AD: se construye la arquitectura del MD que en el caso de la herramienta ROLAP serían los cubos de información. (12)

## **2. Extracción, Transformación y Carga**

El proceso de Extracción, Transformación y Carga (ETL), permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos, limpiarlos y cargarlos en otra BD, MD o AD, para analizar y apoyar un proceso de negocio, quedando unificados los datos provenientes de las múltiples fuentes y eliminadas las inconsistencias existentes en la fuente, que afectan la calidad de los reportes posteriormente generados por los especialistas. Este proceso es una de las etapas de desarrollo del AD de mayor complejidad, ya que transformar los datos extraídos de los distintos sistemas fuentes de manera que sea más fácil manipularlos una vez cargados en el almacén, es una tarea difícil. Los procesos ETL son los componentes más importantes y de valor añadido de una infraestructura de Inteligencia de Negocio. Aún pueden ser invisibles a los usuarios de la plataforma de BI; pero la exactitud de la plataforma BI entera depende de los procesos ETL. (13)

### ➤ **Características del proceso de ETL**

Las principales características de ETL son:

- Multiplataforma (Windows, Linux).
- Multilenguaje (actualmente soporta español e inglés).
- Multicompañía: administra de manera personalizada y segura los datos de múltiples empresas.
- Interfaz de usuario Web.
- Es un mecanismo de carga muy eficiente y efectivo orientado a los AD.
- Enfocado a migrar y mezclar datos.
- Necesita pocos servicios de administración y mantenimiento. (14)

Una mala planificación o construcción de los programas de ETL pueden ocasionar que:

- La información sea inconsistente.
- Paralización de la carga debido a la identificación de datos inconsistentes.
- El programa no se ejecute en el momento requerido.
- No se reprocese la información y se mantienen datos no certeros. (15)

### **3. Inteligencia de Negocio**

Es la forma en que se extrae información desde la BD operacional de un negocio, para filtrarla, sumarizarla, acotarla, entre otras acciones. Se usan herramientas de reporte, gráficas para la construcción de reportes, lo cual contribuye a la toma de decisiones por parte de las personas directivas o clave. (16)

#### **1.4 Instituciones que utilizan Almacenes de Datos en el mundo**

Los sistemas de AD han surgido como respuesta a la problemática de cómo hacer que la cantidad información almacenada por grandes y medianas empresas en las distintas BD fuera útil. Los AD constituyen un recurso importante en las organizaciones y empresas, gestionando gran cantidad de datos de forma centralizada y manteniendo sus sistemas en línea.

Muchas empresas dependen del uso de los AD como: Coca Cola, Walt Disney, Nike, para la realización de estudios en el mercado internacional y de Inteligencia de Negocio, para realizar consultas de alta complejidad y engorrosos reportes. De esta forma elevan su economía manteniéndose en competencia. También es aplicado a temas de seguridad ciudadana ejemplo de ello en Venezuela y en hospitales peruanos para la sectorización de pacientes en el consumo de medicamentos. (17)

En Cuba existen también centros que hacen uso de los AD como herramienta para la toma de decisiones, como la empresa CIMEX, sobresaliente por su estabilidad financiera y crecimiento constante, utilizando los AD para la gestión de inventarios. La UCI en su colaboración a la informatización del país está desarrollando actualmente un AD para la ONE, cuyo nombre es SIGOB. Este almacén es utilizado fundamentalmente para la toma de decisiones, el cual está conformado por un conjunto de MD que responden a las distintas áreas del centro.

## **1.5 Metodologías para el desarrollo de un Almacén de Datos**

Para desarrollar un AD es necesario hacer un estudio de algunos temas generales de la organización para poder posteriormente analizar las metodologías de arquitectura Kimball e Inmon y metodologías de desarrollo existentes que respondan a las necesidades de la institución.

**Ralph Kimball:** (promotor del enfoque dimensional para el diseño de AD) fue una de las principales personalidades influyentes en el área, consideraba que un AD es una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis. Es reconocido como uno de los padres del concepto de AD y se ha dedicado desde hace más de 10 años al desarrollo de su metodología para que este concepto sea bien aplicado en las organizaciones, asegurando la calidad en el desarrollo de los proyectos que la utilizan.

**William H. Inmon:** conocido como el padre de los AD, plantea que un AD es un conjunto de datos orientados por temas, integrados, variantes en el tiempo y no volátiles, que tienen por objetivo dar soporte a la toma de decisiones.

La diferencia entre ambos estilos se debe a la forma de enfrentar el problema.

### **1.5.1 Metodología Kimball vs Inmon**

La metodología de Kimball está enfocada principalmente en el diseño de la BD que almacenará la información para la toma de decisiones. Este método es iterativo pues plantea que el almacén se debe construir pieza por pieza, es decir, un MD a la vez. De esta forma, realiza la construcción del AD, mediante la unión de todos los MD que se desarrollen. Esta solución es muy eficaz y conduce a una solución completa en un período de tiempo muy corto. Esto cumple con lo que se necesita en una metodología de construcción de AD: dar resultados rápidos para demostrar el valor de la solución al negocio y no perder la confianza de los involucrados.

Sin embargo, el enfoque propuesto por Inmon se conoce como top-down, debido a su arquitectura descendente. Este método propone la creación de un repositorio de datos corporativo como fuente de información consolidada persistente, histórica y de calidad, de esta forma, se crea primero el AD del cual se van a nutrir los distintos MD que van a satisfacer las necesidades específicas de un determinado proceso del negocio. Esta metodología por el contrario puede tener una implementación mucho más tardada, y es recomendada cuando se hace demasiado difícil representar el modelo a través de dimensiones y la complejidad de la solución se hace demasiado grande. (18)

Ambas metodologías pueden implementarse en un mismo AD; si es necesario utilizar esta metodología, se recomienda hacerlo en iteraciones ya avanzadas y siempre empezar con Kimball.

### Ciclo de vida Kimball

A continuación se presenta una imagen donde se muestra el ciclo de vida para el desarrollo de esta metodología.

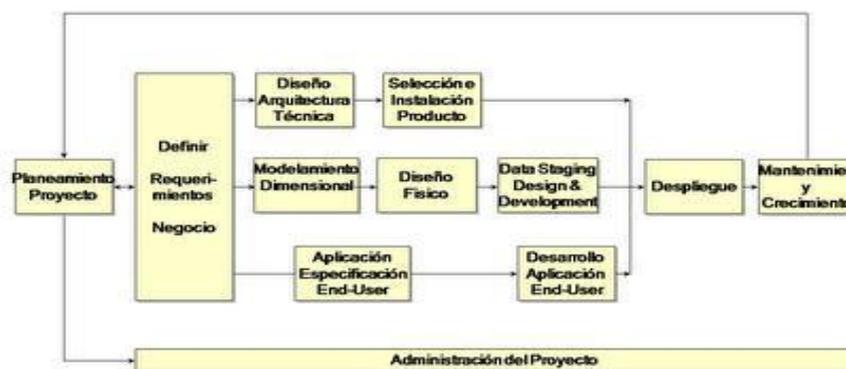


Figura 9: Ciclo de vida Kimball

El Ciclo de vida Kimball comienza con una planificación de proyecto, en la cual se define el alcance, se identifican y programan las tareas, se planifica el uso de los recursos, conformando con todo esto el plan de proyecto. En la segunda etapa de este ciclo se definen los requerimientos del negocio. Luego de definir los requerimientos del negocio, el proyecto se enfoca en tres líneas concurrentes: tecnología, datos y aplicaciones de BI.

El diseño se enfoca principalmente en la creación de la BD que almacenará la información para la toma de decisiones y está enmarcado en la línea de datos; se realiza el modelo dimensional y se analizan los datos del negocio para identificar la granularidad de las tablas de hechos, dimensiones y atributos asociados. Esto se basa en la creación de tablas de hechos, que son tablas que contienen la información numérica de los indicadores a analizar, es decir, la parte cuantitativa de la información. Por último, el ciclo de vida se realiza el despliegue con el objetivo de dejar sentadas las bases de crecimiento y mantenimiento del AD. (19)

➤ **Características del ciclo de vida Kimball:**

- Ilustra el flujo general de implementación de un AD.
- Identifica secuencia de tareas ordenadas y actividades principales que debe suceder concurrentemente.

- Muchas necesidades deben ser acomodadas para lograr una única necesidad de la organización.
- No todos los detalles de las tareas del ciclo de vida deben ser ejecutados en todos los proyectos.

### 1.5.2 Modelo para el desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio en DATEC

El modelo para el desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio en DATEC se basa en el ciclo de vida Kimball, ajustándose a la necesidad de la UCI. Esta adaptación fue lograda gracias a los profesionales del grupo de desarrollo de DATEC. A lo largo de la investigación se utiliza este modelo ya que brinda un conjunto de facilidades y elementos tales como:

- Existe abundante documentación sobre esta metodología y la respuesta a todas las dudas y preguntas que puedan surgir se pueden encontrar en la web, a través de los servicios que brindan el grupo creador de la metodología.
  - Es una metodología madura y reconocida por el resto de la comunidad dedicada al tema. Tiene bien definidas las etapas, actividades, artefactos y roles.
  - La solución completa se puede implementar en poco tiempo.
  - El desarrollo es iterativo e incremental donde se construye una pieza a la vez (MD). Se basa en el paradigma de construcción adoptado por Kimball (“Bottom-Up”), que define la creación de los mercados de datos departamentales y su posterior integración en un AD de la organización. Esta estrategia coincide con la división lógica que se tiene en una organización.
  - Cuenta con mayor velocidad de respuesta al cliente.
  - Los productos son más comprensibles para los usuarios.
  - Es resistente y tolerante ante los cambios.
- **La metodología cuenta con 4 fases:**
- Requerimientos y gestión de proyectos.
  - Arquitectura técnica.
  - Implementación.
  - Implantación y crecimiento.

### ➤ **Flujos de trabajo del ciclo de vida:**

- Estudio preliminar o planificación: se realiza el estudio de la entidad cliente, la planificación 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, definiendo 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 ETL y se configuran e implementan las herramientas de BI para la obtención de los elementos (reportes, gráficos) 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, la capacitación y la 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.

### **1.6 Herramienta de modelado**

Las herramientas de modelado de sistemas informáticos se emplean para modelar los artefactos necesarios durante el proceso de desarrollo de un producto. De esta forma, se enfatiza en ciertas propiedades críticas del sistema, mientras que simultáneamente se ven otros de sus aspectos. Son el conjunto de métodos, utilidades y técnicas que facilitan el mejoramiento del ciclo de vida del desarrollo de un software. (20)

Se utiliza Visual Paradigm 6.4 como herramienta CASE por su integración a UML. Soporta el ciclo de vida completo del desarrollo de software. Utiliza UML como lenguaje de modelado posibilitando una rápida construcción de las aplicaciones con alta calidad y a un menor coste, además porta con las características de ser multiplataforma, amigable en su uso, poseer interoperabilidad con otras aplicaciones e integración con distintos Ambientes de Desarrollo Integrado (IDE) y permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. (21)

### 1.7 Técnicas de captura de requerimientos

La captura de requerimientos es la pieza fundamental del proyecto marcando el punto de partida para las siguientes actividades, sirviendo de base para verificar si se alcanzaron los objetivos establecidos en el proyecto. Se enfoca en el área fundamental: la producción, siendo su tarea la generación de especificaciones correctas que describan con claridad, sin ambigüedades y en forma compacta las necesidades del cliente, cumpliendo lo antes expresado se obtendrá un proyecto que minimizará los problemas relacionados con la gestión de requerimientos. (22)

Una obtención de requerimientos con calidad demuestra que el trabajo realizado culminará con éxito y esto se debe a dos factores:

- La utilización adecuada de las técnicas de captura de requerimientos con los clientes.
- La experiencia de las analistas del proyecto.

#### ➤ **Técnicas más utilizadas en el proceso:**

- Entrevistas y cuestionarios.
- Sistemas existentes.
- Lluvia de ideas.

#### ➤ **Otras técnicas utilizadas para la captura de requerimientos:**

- Observaciones.
- Discusiones.
- Casos de uso.

Ninguna técnica es efectiva por sí misma, requiere ser combinada con otras para lograr una correcta especificación de requerimientos. Para esta investigación se utilizaron principalmente las entrevistas, lluvia de ideas, observaciones y discusiones por las características que presenta la gestión de la información en el área de negocio.

### 1.8 Herramientas informáticas para el proceso de Extracción, Transformación y Carga

Se utiliza Pentaho Data Integration 4.0.1 el cual abre, limpia e integra la información y la pone en manos del usuario proporcionando consistencia, una sola versión de todos los recursos de información y una poderosa Extracción, Transformación y Carga por las siguientes características:

- Entorno gráfico de desarrollo.
- Uso de tecnologías estándar: Java, XML, JavaScript.
- Fácil de instalar y configurar.
- Multiplataforma: Windows, Linux.
- Basado en dos tipos de objetos: transformaciones (colección de pasos en un proceso ETL) y trabajos (colección de transformaciones).
- Código abierto.
- Sin costes de licencia.

#### Incluye cuatro herramientas fundamentales:

- Spoon: para diseñar transformaciones ETL usando el entorno gráfico.
- PAN: para ejecutar transformaciones diseñadas con Spoon.
- CHEF: para crear trabajos.
- Kitchen: para ejecutar trabajos.

También se utiliza Data Cleaner 1.5.3 para el perfilado de los datos y de esta forma realizar un análisis inicial de los datos de las fuentes, con el objetivo entender su estructura, formato y nivel de calidad. El perfilado de los datos permite interactuar con archivos en formato XML y ficheros creados en Microsoft Excel. Además, brinda la opción de generar informes y gráficos que permiten a los usuarios determinar rápidamente el nivel de calidad de los datos. Herramienta de código abierto y sin coste de licencia. (23)

### 1.9 Herramientas informáticas para la inteligencia de negocios, aplicando técnicas OLAP

#### ➤ Pentaho Schema Workbench v3.2

Herramienta de análisis caracterizada por su potencia gráfica y su capacidad multitarea. Puede utilizarse para ingeniería inversa a una BD o a un modelo para visualizar mejor y realizar su mantenimiento.

Es un entorno visual para el desarrollo y prueba de cubos OLAP Mondrian, los cuales proveen un

mecanismo para buscar datos con rapidez y tiempo de respuesta uniforme independientemente de la cantidad de datos en el cubo o la complejidad del procedimiento de búsqueda. Con esta aplicación, se puede configurar una conexión JDBC como el modelo físico, para luego elaborar el esquema lógico de manera simple y efectiva. (24)

### ➤ **Mondrian OLAP Server v3.0.4**

Es un servidor OLAP de código abierto que gestiona la comunicación entre una aplicación OLAP y la BD con los datos fuentes. La herramienta OLAP Mondrian, permite crear cubos de información para análisis multidimensional. Dichos cubos se componen de archivos XML y en ellos se definen las dimensiones y las conexiones de los datos. Los archivos XML por lo general son complejos de realizar manualmente por lo que es común utilizar herramientas gráficas para realizar la edición de estos. (25)

### ➤ **Apache Tomcat v5.5**

Herramienta robusta y altamente eficiente de código abierto, fácil de instalar, multiplataforma, compatible con las API más recientes de Java. Fue creado por el proyecto Jakarta de la Fundación Apache, siendo mantenido por una comunidad de desarrollo. Producto que se ejecuta rápidamente ya que su código binario posee un tamaño total de un poco más de un megabyte, no ocupando mucho espacio en memoria. (26)

### ➤ **Pentaho BI Server**

Es una herramienta de código abierto perteneciente a la familia de aplicaciones de Inteligencia de Negocio. La plataforma Pentaho BI provee la arquitectura y la infraestructura necesarias para crear soluciones de BI. Proporciona servicios básicos incluyendo la autenticación, registro, auditoría, servicios web y motores de reglas. La plataforma también incluye un motor de solución que integra la presentación de informes, análisis, cuadros de mando y los componentes de minería de datos. Integra no sólo las herramientas Mondrian (motor ROLAP), Reporting, Kettle (ETL) y Weka (algoritmos de aprendizaje automático para resolver problemas de data mining), sino que incluye todas las herramientas necesarias para poder tener al completo un entorno para el análisis de datos. (27)

Se utiliza el Pentaho Workbench 3.2.0 para el diseño los cubos OLAP, el Mondrian OLAP Server 3.0.4 y Pentaho BI Server 3.6.0 para la visualización de los cubos OLAP por las características planteadas y por ser herramientas:

- Fácil de instalar y configurar.
- Multiplataforma: Windows, Linux.
- Uso de tecnologías estándar como Java.

- Código abierto.
- Sin costes de licencia.

### **1.10 Sistema Gestor de Bases de Datos**

Un Sistema Gestor de Bases de Datos (SGBD) es un conjunto de programas que ayudan a administrar la información que está contenida en una BD. Estos programas se encargan de la integridad y la seguridad de los datos y garantizan la interacción con el sistema operativo. Entre los tipos de gestores de BD que existen se pueden mencionar los relacionales, jerárquico y de red; pero el más utilizado es el relacional. (28)

#### **➤ Características de los Sistema Gestor de Bases de Datos**

- Son capaces de controlar la concurrencia y las operaciones que implican la recuperación de fallos.
- Definen usuarios y sus restricciones de acceso.
- Respetan la integridad y seguridad de los datos.
- Toleran definiciones de esquemas y vistas.

La UCI se encuentra en franca migración hacia la independencia tecnológica. Actualmente está exportando todos sus dispositivos de almacenamiento hacia plataforma PostgreSQL por lo que queda seleccionado como SGBD. Esta decisión ha sido previamente colegiada y aceptada por parte del cliente final debido a que dentro de sus políticas de migración se encuentran las de llevar a todas sus BD hacia dicha plataforma. En este sentido, la versión que se utilizará es la 8.4.0 por ser lo suficientemente estable y segura.

#### **➤ Características del PostgreSQL 8.4.0:**

- Funciona en la mayoría de los sistemas operativos actuales.
- Posee puntos de recuperación en un momento dado.
- Replicación asincrónica.
- Copia de seguridad en línea.
- Altamente extensible pues soporta operadores, funciones, métodos de acceso y tipos de datos definidos por el usuario.
- Soporta la integridad referencial, la cual es utilizada para garantizar la validez de los datos en una BD.

- Posee un API flexible.

### **1.11 Conclusiones del capítulo**

A lo largo de este capítulo se ha explicado y detallado la base sobre la que se trabajará en aras de desarrollar el MD para darle solución a la problemática existente en el área Agricultura, ganadería y silvicultura de la ONE. Después de haber realizado el estado del arte concerniente a ellos se concluye lo siguiente:

- Para el desarrollo del MD se utilizó el modelo para el desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio en DATEC la cual facilitará y agilizará el trabajo durante el desarrollo de la solución propuesta.
- Para una clara obtención de requerimientos se utilizaron técnicas como las entrevistas, tormenta de ideas, observaciones y discusiones.
- Para desarrollar el diseño de la solución se decide utilizar la herramienta de modelado Visual Paradigm versión 6.4.
- Como herramienta de ETL se escogió el Pentaho Data Integration 4.0.1.
- Como herramienta de BI se escogió Pentaho BI Server 3.0.4.
- Se utilizará el SGBD PostgreSQL en su versión 8.4.0.

## **Capítulo 2: Análisis y diseño del Mercado de Datos Agricultura, ganadería y silvicultura**

### **2.1 Introducción**

Se realiza el análisis y diseño de los datos del área Agricultura, ganadería y silvicultura del Sistema de Información de Gobierno. Para lograr esto se modela el negocio, se hace el levantamiento de los requerimientos y se diseña el modelo de datos.

### **2.2 Caracterización de las áreas de la organización**

La ONE, órgano rector de la estadística en Cuba, actúa como un almacén que gestiona y almacena la información estadística del país. La información recopilada proviene de diferentes fuentes, la diferencia entre estas está dada fundamentalmente en las formas de captación, períodos de captura (mensual, trimestral, semestral, anual), entre otras. Esta información es registrada en modelos organizados por organismos y sus respectivos centros informantes. Los modelos están compuestos por un conjunto de indicadores ajustados a cada centro informante en dependencia al trabajo que se realice en cada uno. Los modelos 0336-01 y 0467-06 recogen información referente a la Agricultura, ganadería y silvicultura. El proceso de recogida y gestión de sus datos es el siguiente (ver Figura 10):

1. El modelo 0336-01 o Sector Agropecuario y el modelo 0467-06 o Sector Silvicultura se conforman con los datos enviados por los organismos, los cuales se recogen mensualmente, salvo el modelo 0336-01 que la cantidad de superficie sembrada y en producción que se recoge trimestralmente por la Oficina Municipal de Estadísticas (OME) de las provincias para luego enviarlos a la Oficina Provincial de Estadísticas (OPE).
2. En la OPE se realiza un resumen de los datos por municipio, los cuales se guardan en un fichero de extensión “dbf” para colocarlos en un servidor, que tiene una carpeta definida para cada provincia.
3. El fichero “dbf” es enviado a la ONE, donde se realiza la limpieza de los datos por parte de los especialistas y se genera un archivo general con la información.
4. Luego esta información es publicada en el sitio de la ONE, quedando disponible a los interesados.

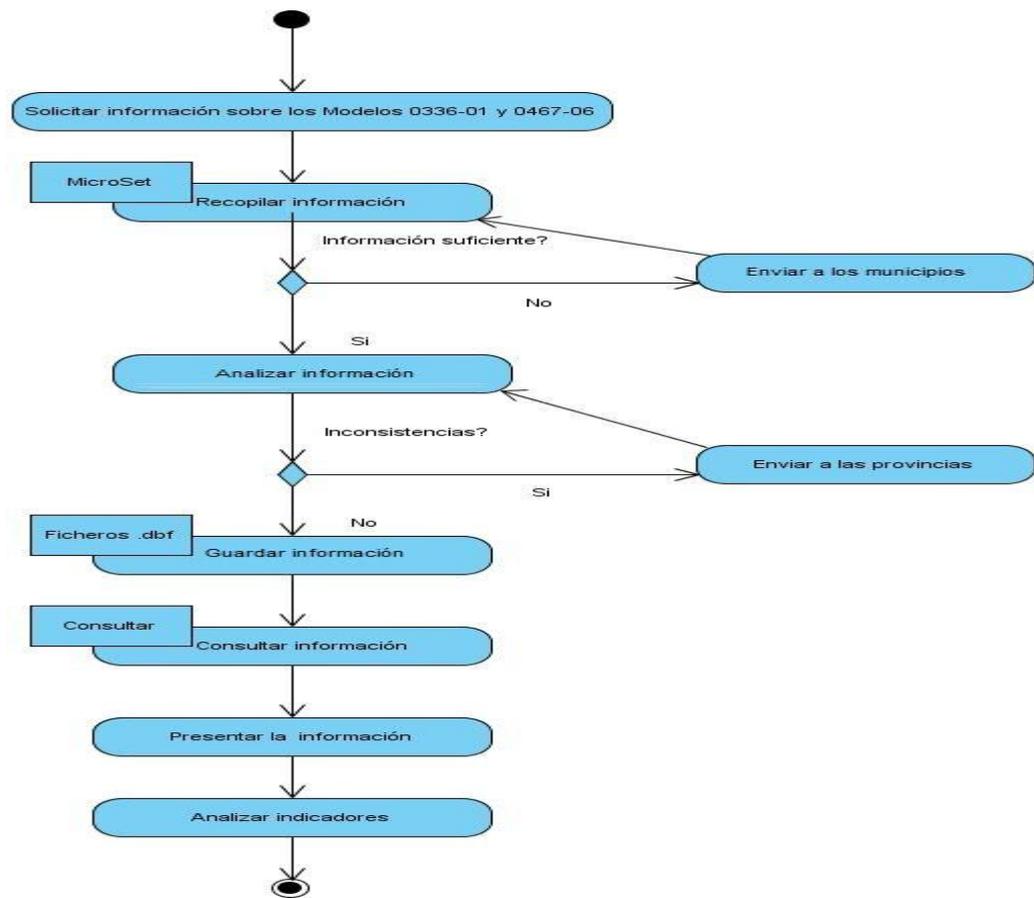


Figura 10: Proceso del negocio

### 2.3 Necesidades de información

Es vital conocer qué buscan y necesitan los clientes, por esta razón en esta etapa se realizan reuniones con los especialistas de la ONE para analizar la información referente al área de Agricultura, ganadería y silvicultura, identificando cuáles son las necesidades que tiene la organización. Las necesidades de información se recogen realizando un estudio donde se especifica cuáles son los objetivos de la organización, redactándose en forma de requerimientos para luego analizar los indicadores que serán tomados para el desarrollo del MD.

### 2.4 Reglas del negocio

Las reglas del negocio como su nombre lo indica son reglas, políticas, acuerdos o regulaciones del negocio para el funcionamiento del sistema. Por esta razón la identificación de estas constituye un eslabón esencial dentro del sistema. Las reglas del negocio definidas en la investigación fueron:

- Enviar la información en una fecha establecida por la sede central, para lograr un análisis continuo y sistemático.
- La dimensión DPA tiene cuatro cifras, las dos primeras corresponden al municipio y las restantes a la provincia.
- En el modelo 0336-01 el  $\text{total\_ventas} = \text{ventas\_contratadas} + \text{ventas\_trabajadores} + \text{ventas\_mercado\_agropecuario}$ .
- En la serie agricultura, ganadería, silvicultura y pesca el rendimiento de la caña de azúcar =  $\text{producción/superficie cosechada} * 100$ .

## **2.5 Especificación de requerimientos**

### **2.5.1 Requerimientos de información**

Los requerimientos de información representa toda la información que debe estar disponible para el usuario final. Estos deben estar disponibles a la hora de hacer las consultas, y constituyen una entrada principal para el proceso de inteligencia de negocio y para futuros reportes bases. A continuación se especifican los requerimientos de información que fueron identificados agrupados por dominios informativos.

#### **Sector agropecuario**

- RI1.** Obtener la superficie existente sembrada de cultivos del sector agropecuario, dado una entidad, indicadores de la agricultura, un sector, un organismo, un DPA y un mes.
- RI2.** Obtener superficie sembrada de cultivos del sector agropecuario, dado una entidad, indicadores de la agricultura, un sector, un organismo, un DPA y un mes.
- RI3.** Obtener superficie cosechada y en producción del sector agropecuario, dado una entidad, indicadores de la agricultura, un sector, un organismo, un DPA y un mes.
- RI4.** Obtener la producción agrícola del sector agropecuario, dado una entidad, indicadores de la agricultura, un sector, un organismo, un DPA y un mes.
- RI5.** Obtener autoconsumo del sector agropecuario, dado una entidad, indicadores de la agricultura, un sector, un organismo, un DPA y un mes.
- RI6.** Obtener el total de ventas, dado una entidad, indicadores de la agricultura, un sector, un organismo, un DPA y un mes.

- RI7. Obtener ventas totales a contratadas del sector agropecuario, dado una entidad, indicadores de la agricultura, un sector, un organismo, un DPA y un mes.
- RI8. Obtener ventas totales al mercado agropecuario del sector agropecuario, dado una entidad, indicadores de la agricultura, un sector, un organismo, un DPA y un mes.
- RI9. Obtener ventas totales a trabajadores del sector agropecuario, dado una entidad, indicadores de la agricultura, un sector, un organismo, un DPA y un mes.
- RI10. Obtener el rendimiento agrícola por cultivos seleccionados del sector agropecuario, dado una entidad, indicadores de la agricultura, un sector, un organismo, un DPA y un mes.

### **Agricultura Cañera**

- RI11. Obtener la superficie cosechada de la caña de azúcar, dado una zafra, un DPA y un sector.
- RI12. Obtener la producción de la caña de azúcar, dado una zafra, un DPA y un sector.
- RI13. Obtener el rendimiento por hectárea de la caña de azúcar, dado una zafra, un DPA y un sector.

### **Ganadería**

- RI14. Obtener ganado vacuno, dado una entidad, indicadores de ganadería, un sector, un organismo, un DPA, y un mes.
- RI15. Obtener ganado equino, dado una entidad, indicadores de ganadería, un sector, un organismo, un DPA, y un mes.
- RI16. Obtener ganado ovino, dado una entidad, indicadores de ganadería, un sector, un organismo, un DPA, y un mes.
- RI17. Obtener ganado caprino, dado una entidad, indicadores de ganadería, un sector, un organismo, un DPA, y un mes.
- RI18. Obtener ganado porcino, dado una entidad, indicadores de ganadería, un sector, un organismo, un DPA, y un mes.
- RI19. Obtener ganado cunícula, dado una entidad, indicadores de ganadería, un sector, un organismo, un DPA, y un mes.
- RI20. Obtener ganado avícola, dado una entidad, indicadores de ganadería, un sector, un organismo, un DPA, y un mes.
- RI21. Obtener ganado búfalo, dado una entidad, indicadores de ganadería, un sector, un organismo, un DPA, y un mes.

## **Silvicultura**

- RI22. Obtener estatal flora y fauna, dado una provincia, indicadores de silvicultura y un mes.
- RI23. Obtener estatal EMA y EFI, dado una provincia, indicadores de silvicultura y un mes.
- RI24. Obtener estatal otras MINAG, dado una provincia, indicadores de silvicultura y un mes.
- RI25. Obtener estatal MINAZ, dado una provincia, indicadores de silvicultura y un mes.
- RI26. Obtener no estatal UBPC MINAZ, dado una provincia, indicadores de silvicultura y un mes.
- RI27. Obtener no estatal UBPC MINAG, dado una provincia, indicadores de silvicultura y un mes.
- RI28. Obtener no estatal CPA y CSS MINAZ, dado una provincia, indicadores de silvicultura y un mes.
- RI29. Obtener no estatal CPA y CSS MINAG, dado una provincia, indicadores de silvicultura y un mes.
- RI30. Obtener no estatal Camp. Disp. MINAZ, dado una provincia, indicadores de silvicultura y un mes.
- RI31. Obtener no estatal Camp. Disp. MINAG, dado una provincia, indicadores de silvicultura y un mes.
- RI32. Obtener no estatal MINFAR, dado una provincia, indicadores de silvicultura y un mes.
- RI33. Obtener no estatal MININT, dado una provincia, indicadores de silvicultura y un mes.
- RI34. Obtener otros por las masas, dado una provincia, indicadores de silvicultura y un mes.
- RI35. Obtener otros, dado una provincia, indicadores de silvicultura y un mes.

### **2.5.2 Requerimientos funcionales**

Los requerimientos funcionales son capacidades o condiciones que el sistema debe cumplir para dar respuesta a todos los requerimientos de información listados anteriormente. Fueron identificados los siguientes requerimientos funcionales:

**RF 1.1** Insertar usuario.

**RF 1.3** Eliminar usuario.

Se definieron 22 requerimientos funcionales los cuales se encuentran contemplados en el artefacto Especificación de requerimientos.

### **2.5.3 Requerimientos no funcionales**

Los requerimientos no funcionales son propiedades o cualidades que el producto debe tener y están vinculados directamente a los requerimientos funcionales. Estos requerimientos determinan cómo debe comportarse el producto, haciéndolo atractivo, usable, rápido y confiable.

#### **➤ Software**

#### **RNF 1. Instalar en las estaciones de trabajo el software necesario para el correcto funcionamiento del sistema**

Las configuraciones de software de las máquinas clientes deben contar al menos el siguiente elemento:

- Firefox 2.0 o superior.
- Java Virtual Machine 6.0 y Schema Workbench 3.2.1 en caso de que un usuario capacitado requiera la construcción de esquemas multidimensionales para el diseño de nuevos reportes.

Se identificaron además requerimientos no funcionales de Usabilidad, Fiabilidad, Eficiencia, Soporte, Restricciones de diseño, Requerimientos para la documentación de usuarios y ayuda del sistema, Interfaz, Requerimientos legales, de Derecho de autor, Estándares aplicables, Requerimientos de seguridad, Requerimientos de software y los Requerimientos de hardware. Estos requerimientos se encuentran contemplados en el artefacto Especificación de requerimientos.

### **2.6 Modelo de casos de uso del sistema**

#### **2.6.1 Actores del sistema**

Los actores del sistema son personas o programa que interactúa con el sistema, es decir, no son parte de este, sino que pueden intercambiar información con él.

- **Especialista de información:** consulta y analiza los reportes recibidos de las oficinas.
- **Administrador:** encargado de gestionar usuarios, reportes, roles, consulta y analiza los reportes recibidos de las oficinas.
- **Administrador de datos:** realiza el proceso de Extracción, Transformación y Carga de datos.

### 2.6.2 Diagrama de Casos de Uso del sistema

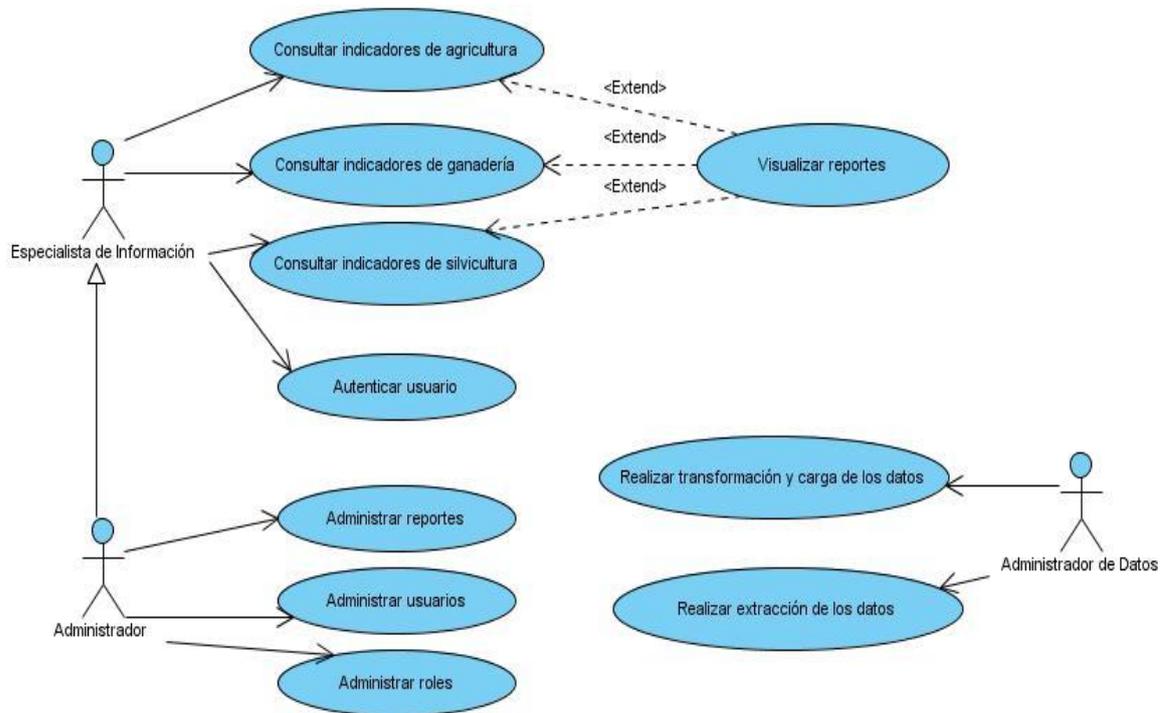


Figura 11: Diagrama de Casos de Uso del Sistema

### 2.6.3 Especificación de Casos de Uso del Sistema

<b>Caso de Uso:</b>	Consultar indicadores de ganadería.
<b>Tipo:</b>	Información
<b>Actores:</b>	Especialista de información, Administrador
<b>Resumen:</b>	El CU inicia cuando el Especialista de Información o el Administrador seleccionan la opción “Área de análisis de Ganadería (A.A Ganadería)”, la cual muestra los libros de trabajo referentes al tema de ganadería. El CU finaliza una vez que el Especialista de Información o el administrador terminan de analizar los reportes.
<b>Precondiciones:</b>	Compleitud del MD. Los datos tienen que estar cargados. El usuario tiene que estar

	autenticado.	
<b>Referencias</b>	RI 14, RI 15, RI 16, RI 17, RI 18, RI 19, RI 20, RI 21	
<b>Prioridad</b>	Crítico	
<b>Flujo Normal de Eventos</b>		
<b>Acción del Actor</b>	<b>Respuesta del Sistema</b>	
1. El Especialista de información entra al sistema.	2. El sistema muestra la interfaz para insertar usuario y contraseña.	
3. El Especialista de información se autentica.	4. El sistema muestra el área de análisis general A.A.G SIGOB.	
5. El Especialista de información selecciona el área de análisis de Agricultura, ganadería y silvicultura (A.A Agricultura, ganadería y silvicultura).	6. El sistema muestra las áreas de análisis asociados al área de análisis de Agricultura, ganadería y silvicultura.	
7. El Especialista de información selecciona el área de análisis de Ganadería (A.A Ganadería)	8. El sistema muestra los libros de trabajo asociados al área de análisis de Ganadería.	
9. El Especialista de información selecciona el libro de trabajo deseado.	10. El sistema muestra los reportes asociados al libro de trabajo seleccionado y brinda la posibilidad de configurar dichos reportes (ver caso de uso Visualizar reporte).	
<b>Opciones de reportes del área de análisis Ganadería ‘Libro de trabajo Ganadería’</b>		
<b>Perspectivas de análisis</b>	<b>Posibles resultados</b>	
	<b>Medidas</b>	<b>Periodicidad</b>
<u>Dimensiones disponibles relacionadas con el hecho Ganadería.</u>	<u>Medidas disponibles en el hecho Ganadería.</u>	<u>Rango de tiempo en que se solicitan las medidas.</u>
<ul style="list-style-type: none"> <li>Sector</li> <li>Temporal mes</li> </ul>	<ul style="list-style-type: none"> <li>VACUNO</li> <li>EQUINO</li> <li>OVINO</li> </ul>	

<ul style="list-style-type: none"> <li>Indicadores ganadería</li> <li>DPA</li> <li>Entidad</li> <li>Organismo</li> </ul>	<ul style="list-style-type: none"> <li>CAPRINO</li> <li>PORCINO</li> <li>CUNÍCULA</li> <li>AVÍCOLA</li> <li>BÚFALO</li> </ul>	<ul style="list-style-type: none"> <li>Mensual</li> </ul>
--	---	---

**Prototipo de Interfaz**

Indicadores ganadería		Temporal mes				
		Enero				
		Medidas				
Indicadores ganadería	DPA	VACUNO	EQUINO	OVINO	CAPRINO	PORCINO
Ventas totales de carne	+ Cuba	20204,00	55,30	404,30	133,10	23792,30
Recuperación de cueros	+ Cuba	907,90	0,10	34,90	6,30	1,10
Venta de cueros a la industria	+ Cuba	517,60	0,00	16,10	0,00	0,00
Hembras en ordeño	+ Cuba	557956,00	0,00	0,00	11166,00	0,00
Producción de leche	+ Cuba	350685,00	0,00	0,00	757,20	0,00
Ventas totales de leche	+ Cuba	295402,00	0,00	0,00	482,60	0,00
Insumo animal	+ Cuba	297,50	0,00	0,00	182,60	0,00
Insumo productivo	+ Cuba	290,30	0,00	0,00	0,80	0,00

**Postcondiciones** Disponibilidad de reportes relacionados con el CU “Consultar indicadores de ganadería”.

**Tabla 3: Especificación de Casos de Uso del Sistema Consultar indicadores de ganadería**

El resto de las descripciones de los Casos de Uso se encuentran en el artefacto Modelos de Casos de Uso del Sistema.

**2.7 Especificación del modelo dimensional**

Para el desarrollo en esta etapa del MD se diseña el modelo de datos donde se recogen procesos fundamentales como son las tablas de hechos independientes para realizar un mejor análisis. También este modelo recoge las dimensiones candidatas para la solución del problema. Es importante conocer la composición de ambos y la relación que existe entre ellos para la eficiente función del MD.

### 2.7.1 Matriz dimensional

La Matriz dimensional representa las relaciones existentes entre los hechos y las dimensiones del sistema diseñado. Para los modelos propuestos existen dimensiones comunes y específicas de cada modelo que se relacionan directamente con las dos tablas de hechos existentes (ver Tabla 3).

➤ **Dimensiones**

- D1.dim\_dpa
- D2.dim\_entidad
- D3.dim\_zafra
- D4.dim\_sector
- D5.dim \_organismo
- D6.dim\_temporal\_mes
- D7.dim\_indicadores\_silvicultura
- D8.dim\_provincia
- D9.dim\_indicadores\_ganaderia
- D10. dim\_indicadores\_agricultura

➤ **Hechos**

- H1.hech\_canna\_azucar
- H2.hech\_sector\_agropecuario
- H3.hech\_informacion\_ganado
- H4.hech\_silvicultura

Hechos/ Dimensiones	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
H1			X	X						
H2	X	X		X	X	X				X
H3	X	X		X	X	X			X	
H4						X	X	X		

**Tabla 4: Matriz dimensional**

### 2.7.2 Tablas de dimensiones

Las tablas de dimensiones contienen la descripción textual del negocio. En el modelo dimensional poseen varios atributos que en su conjunto definen una fila en la tabla de dimensión. Los atributos de las dimensiones sirven como fuente primaria de las restricciones de las consultas, agrupaciones y las etiquetas de los reportes. Estos atributos son las llaves de entrada a los hechos o medidas almacenadas. La calidad del AD se mide por la definición de los atributos de las dimensiones. Un ejemplo de ello se representa en la Figura 12, el resto se encuentra en el artefacto Especificación del Modelo de Datos Dimensional.

- **dim\_indicadores\_ganaderia**: dimensión que contiene los indicadores referentes al área de ganadería.

dim_indicadores_ganaderia		
<b>+dim_indicadores_ganaderia_id</b>	<b>int4</b>	<b>Nullable = false</b>
indicador_nombre	varchar(45)	Nullable = false
indicador_descripcion	varchar(100)	Nullable = false
indicador_codigo	varchar(11)	Nullable = false
indicador_um	char(25)	Nullable = false
indicador_agregable	char(1)	Nullable = false
seguridad_codigo	varchar(1)	Nullable = false
seguridad_nombre	varchar(25)	Nullable = false
codigo_fila	varchar(8)	Nullable = false

Figura 12: Dimensión indicadores ganadería

### 2.7.3 Tablas de hechos

Las tablas de hechos son las fuentes de almacenamiento de las medidas numéricas. Son las tablas que servirán como fuente de información principal para la realización de las estructuras que soporten los diversos reportes más comunes de la institución. Siendo estas tablas las que van a residir como repositorio central, para toda la información existente en los Modelo 0334-01 y 0467-06. Mientras que las medidas representan las variables de salida del almacén, estas pueden estar o no en las tablas hechos. Se clasifican en variables físicas las cuales se encuentran en las tablas de hecho mientras que las calculables no son más resultados que el almacén devuelve tras algún análisis realizado. Un ejemplo de ello se representa en la Figura 13, el resto se encuentra en el artefacto Especificación del Modelo de Datos Dimensional.

- **hech\_informacion\_ganado**: almacena la información referente a la ganadería.

hech_informacion_ganado		
<b>+#id_dim_sector</b>	<b>int4</b>	<b>Nullable = false</b>
<b>+#dim_dpa_id</b>	<b>int4</b>	<b>Nullable = false</b>
<b>+#dim_organismo_id</b>	<b>int4</b>	<b>Nullable = false</b>
<b>+#dim_indicadores_ganaderia_id</b>	<b>int4</b>	<b>Nullable = false</b>
# dim_temporal_mesdim_temporal_mes_id	int4	Nullable = false
cant_vacuno	float8	Nullable = false
cant_equino	float8	Nullable = false
cant_porcino	float8	Nullable = false
cant_caprino	float8	Nullable = false
cant_bufalo	float8	Nullable = false
cant_cunicula	float8	Nullable = false
cant_avicola	float8	Nullable = false
cant_ovino	float8	Nullable = false
# dim_entidaddim_entidad_id	int4	Nullable = false

Figura 13: Hecho información ganado

#### 2.7.4 Modelo dimensional

Un Modelo Dimensional es una representación de un diseño de datos, que tiene en cuenta las facilidades y restricciones de los sistemas de base datos. Permite describir las estructuras de datos y las restricciones de integridad. Organiza la información para garantizar la velocidad y eficiencia en la recuperación de esta. Una de las características principales es que no necesita una predefinición de los reportes debido a que se diseñan de forma tal que cubren el universo de variantes que los usuarios necesitan para consultar la información almacenada. Los modelos dimensionales están compuestos por hechos y dimensiones relacionados entre sí. Una correcta modelación de este modelo es vital para éxito del MD.

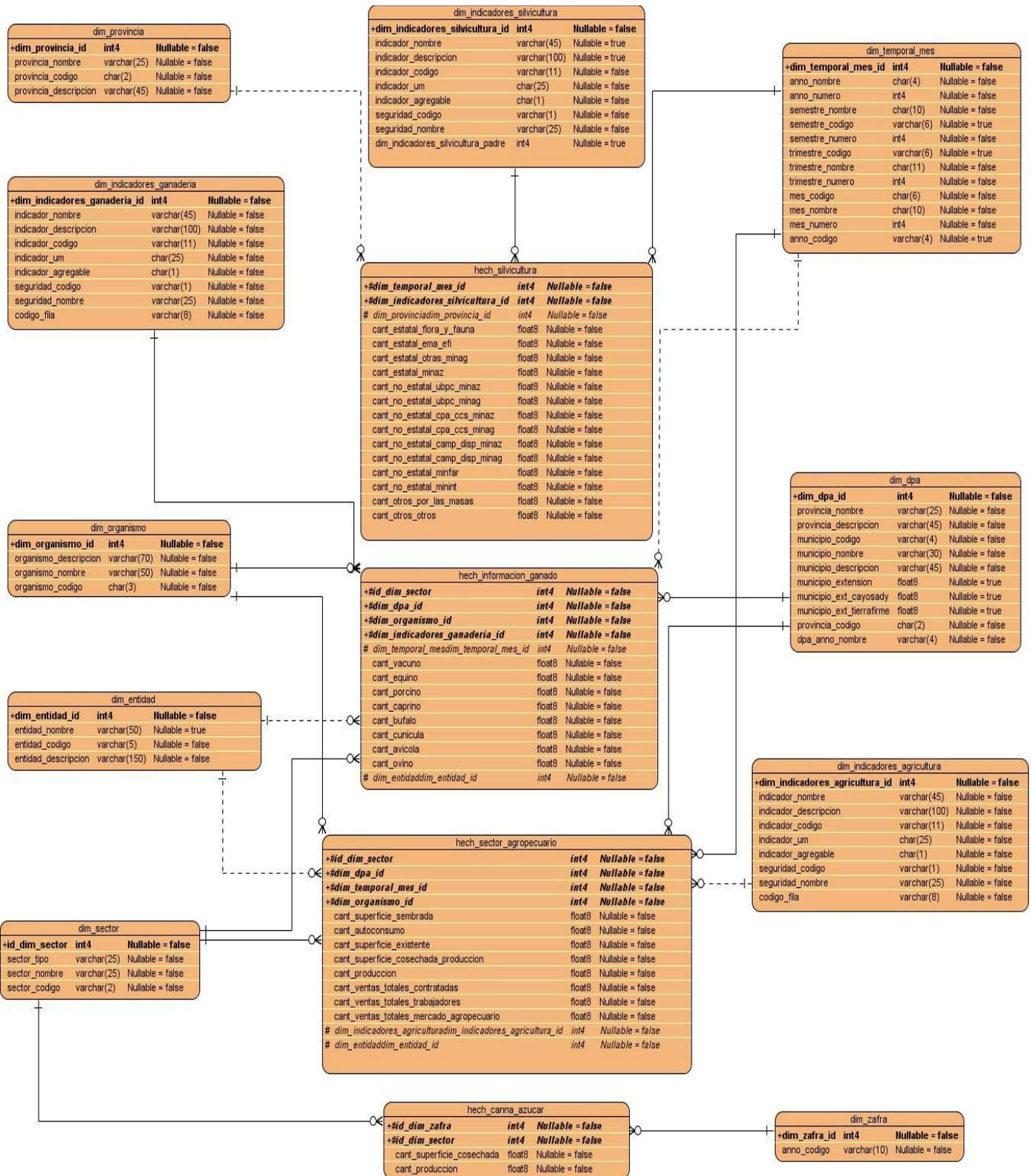


Figura 14: Modelo dimensional

## **2.8 Especificación del modelo físico**

El modelo de datos físico proporciona conceptos que describen los detalles de cómo se almacenan los datos en el sistema. Los conceptos de los modelos físicos están dirigidos al personal informático, no a los usuarios finales. La representación del modelo físico es la misma que la del modelo dimensional ya que las relaciones entre las dimensiones y los hechos son de uno a mucho y no se generan nuevas tablas.

## **2.9 Conclusiones del capítulo**

A lo largo de este capítulo se realizó una propuesta de diseño de un MD, aplicándose al área de Agricultura, ganadería y silvicultura del SIGOB. Posteriormente a los procedimientos realizados en el capítulo se obtuvieron los siguientes resultados:

- Se realizó una caracterización de los objetivos de las áreas de la organización.
- Se identificaron las necesidades de información de la organización según sus objetivos y funciones.
- Se identificaron cuatro reglas del negocio.
- Se identificaron 35 requerimientos de información que recogen los datos informativos para el usuario final.
- Se identificaron 22 requerimientos funcionales que deben dar respuesta a los requerimientos de información.
- Se identificaron 33 requerimientos no funcionales para el funcionamiento del sistema.
- Se realizó el diseño el Diagrama de Casos de Uso del Sistema a partir de los requerimientos funcionales identificados.
- Se modelaron 10 tablas de dimensiones según las perspectivas del análisis identificadas.
- Se modelaron cuatro tablas de hechos que recogen los procesos fundamentales.
- Se diseñó un modelo lógico luego de haber establecido las relaciones entre las tablas de dimensiones y las tablas de hechos.
- Se hace una breve descripción del modelo físico generado previamente a partir del modelo lógico.

## **Capítulo3: Implementación del Mercado de Datos Agricultura, ganadería y silvicultura**

### **3.1 Introducción**

En el capítulo se diseñan e implementan los subsistemas de integración y visualización del área de Agricultura, ganadería y silvicultura del Sistema de Información de Gobierno. Para lograr esto se definen las áreas de análisis con los libros de trabajo correspondientes. Además se diseñan e implementan los reportes candidatos, flujos de transformaciones, trabajos y cubos OLAP.

### **3.2 Diseño del subsistema de integración**

#### **3.2.1 Registro de sistemas fuentes**

El registro de sistemas fuente está enfocado a la documentación del estado físico de los datos y de los responsables de estos en los sistemas fuentes.

##### **➤ Estructura de los datos**

Se caracterizan por las funciones de acceso que utilizan para almacenar y acceder a elementos individuales y es una forma de organizar los datos elementales con el objetivo de facilitar su manipulación. Por debajo de estas estructuras hay un nivel, en el cual se sitúan los archivos, discos, particiones y espacios de tablas.

##### **• Esquemas**

Representan una manera de tener organizada toda la información contenida en las BD. En diferencia de las BD los esquemas no se encuentran separados entre sí, esto permite que el usuario pueda tener acceso a ellos siempre que posea los permisos adecuados y estos pueden contener funciones, operadores y tipos de datos.

Para el desarrollo del Mercado de Datos se definieron 2 esquemas:

- dimensiones: contiene las tablas de las dimensiones propuestas que son comunes con la BD de la ONE.
- mart\_agropecuario: contiene todas las tablas de hechos y las dimensiones propias del mercado de datos.

• **Tablas**

La solución cuenta con 16 tablas en total, descomponiéndose en 10 tablas de dimensiones, y 4 tablas de hechos distribuidos por los 2 esquemas.

<b>Esquemas</b>	<b>Tablas que lo componen</b>
<b>dimensiones</b>	<b>dim_dpa</b> <b>dim_entidad</b> <b>dim_organismo</b> <b>dim_provincia</b> <b>dim_temporal_mes</b>
<b>mart_agropecuario</b>	<b>dim_indicadores_agricultura</b> <b>dim_indicadores_ganaderia</b> <b>dim_indicadores_silvicultura</b> <b>dim_sector</b> <b>dim_zafra</b> <b>hech_canna_azucar</b> <b>hech_sector_agropecuario</b> <b>hech_informacion_ganado</b> <b>hech_silvicultura</b>

**Tabla 5: Estructura de los datos**

### **3.3 Implementación del subsistema de integración**

El proceso de Integración de los Datos consta de tres etapas fundamentales relacionadas entre sí. Estas etapas son: (29)

➤ **Extracción de los datos**

La primera parte del proceso ETL consiste en extraer los datos desde los sistemas de origen. La mayoría de los proyectos de almacenamiento de datos fusionan datos provenientes de diferentes sistemas de origen. Cada sistema separado puede usar una organización diferente de los datos o formatos distintos. Los formatos de las fuentes normalmente se encuentran en BD relacionales o ficheros planos, pero pueden incluir BD no relacionales u otras estructuras diferentes. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación.

### ➤ **Transformación de los datos**

La fase de transformación aplica una serie de transformaciones basadas precisamente en las reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. Algunas fuentes de datos requerirán alguna pequeña manipulación de los datos. No obstante en otros casos pueden ser necesarias aplicar algunas de las siguientes transformaciones:

- Seleccionar sólo ciertas columnas para su carga, un ejemplo de esto es que las columnas con valores nulos no se carguen.
- Traducir códigos, esto se puede evaluar si la fuente almacena una “H” para Hombre y “M” para Mujer pero el destino tiene que guardar “1” para Hombre y “2” para Mujer.
- Unir datos de múltiples fuentes como búsquedas y combinaciones.
- Calcular totales de múltiples filas de datos, mostrándose valores de ventas totales de cada región.
- Transponer o pivotar (girando múltiples columnas en filas o viceversa).

### ➤ **Carga de los datos**

La fase de carga es el momento en el cual los datos de la fase anterior (transformación) son cargados en el sistema de destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes. En algunas BD se sobrescribe la información antigua con nuevos datos, y en otros se mantienen un historial de los registros de manera que se pueda hacer una auditoría.

#### **3.3.1 Implementación de los flujos de transformación**

La transformación de los datos se realiza de acuerdo a las reglas que se definieron en el negocio. Para su desarrollo se ejecutaron 15 transformaciones, un ejemplo de ello se representa en la Figura 15, el resto se encuentra en el expediente de proyecto dentro de la carpeta ETL en los (.ktr).

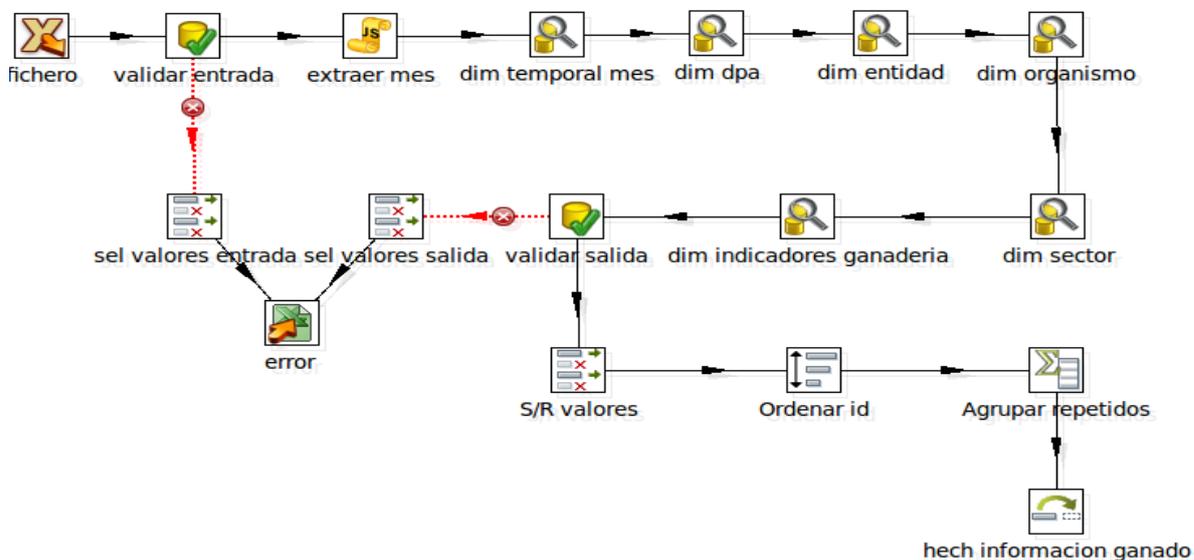


Figura 15: Transformación del hecho información ganado

### 3.3.2 Implementación de los trabajos

Los trabajos son una secuencia de tareas que se ejecutan secuencialmente. Un ejemplo de ello es el que se muestra en la Figura 16, el resto se encuentra en el expediente de proyecto dentro de la carpeta ETL en los (.ktr).



Figura 16: Trabajo

### 3.4 Diseño del subsistema de visualización

En correspondencia con los requerimientos de información identificados en el área de Agricultura, ganadería y silvicultura, se definieron tres áreas de análisis y seis reportes, agrupados en cuatro libros de trabajo y asociados a las áreas de análisis ubicadas dentro del Área de Análisis General SIGOB, ver Figura 17.

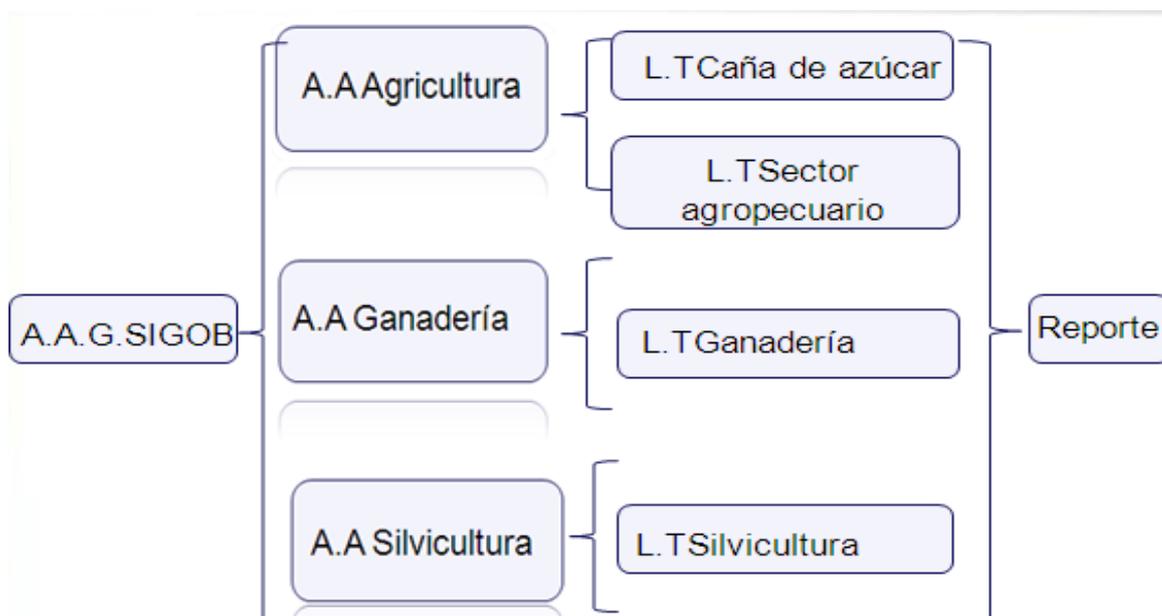


Figura 17: Mapa de navegación

### 3.4.1 Arquitectura de información

Ofrece un entorno de análisis, monitoreo y control de la información del área de Agricultura, ganadería y silvicultura. La arquitectura de información ayuda a obtener una descripción de los elementos que estructuran el sistema así como también los elementos que serán mostrando en la capa del MD.

### 3.4.2 Diseño de los reportes candidatos

Con el diseño de los reportes candidatos se puede obtener una comprensión precisa de los elementos que lo componen. También se puede diseñar la frecuencia en que se va a ejecutar un reporte así como las funciones que realiza. Un ejemplo de su diseño se muestra a continuación:

<b>Área de análisis (A.A)</b>	<b>Ganadería</b>
<b>Libro de Trabajo (L.T)</b>	LT1 –Ganadería
<b>Reporte (Tabla de Salida – TS)</b>	TS4 – Indicadores de ganadería por DPA y mes
<b>Descripción</b>	Reporte que muestra la información referente al ganado en Cuba dado un DPA y un mes.
<b>Elementos del reporte</b>	Indicadores ganadería: dim_indicadores_ganaderia Tiempo: dim_temporal_mes Sector: dim_dpa
<b>Frecuencia de emisión</b>	Mensual
<b>Funciones</b>	
<b>Gráfico</b>	Gráfico de barras en 3D

**Tabla 6: Diseño de Reporte Candidato Indicadores de ganadería por DPA y mes**

### **3.5 Implementación del subsistema de visualización**

#### **3.5.1 Implementación de los cubos OLAP**

Los hechos definidos para un MD indica la cantidad de cubos OLAP que deben ser creados para la visualización de la información. En correspondencia con los hechos definidos para el área de Agricultura, ganadería y silvicultura se implementaron cuatro cubos OLAP, el resto se encuentra en el expediente de proyecto dentro de la carpeta de BI en los (.xml). Los cubos OLAP están compuestos por las dimensiones y medidas definidas en Modelo de Datos, ver Figura 18.



Figura 18: Cubo del hecho información ganado

### 3.5.2 Implementación de los reportes candidatos

Los reportes son la unidad básica utilizada para mostrar información almacenada. Un ejemplo de reporte es Indicadores ganadería por DPA y mes, asociado al Área de Análisis Ganadería, el cual muestra la información referente al ganado en Cuba y permite analizar la información a través de los perfiles de análisis Indicadores ganadería, DPA y Temporal mes.

Indicadores ganaderia	DPA	Temporal mes				
		Enero				
		VACUNO	EQUINO	OVINO	CAPRINO	PORCINO
Ventas totales de carne	Cuba	20204,00	55,30	404,30	133,10	23792,30
Recuperación de cueros	Cuba	907,90	0,10	34,90	6,30	1,10
Venta de cueros a la industria	Cuba	517,60	0,00	16,10	0,00	0,00
Hembras en ordeño	Cuba	557956,00	0,00	0,00	11166,00	0,00
Producción de leche	Cuba	350685,00	0,00	0,00	757,20	0,00
Ventas totales de leche	Cuba	295402,00	0,00	0,00	482,60	0,00
Insumo animal	Cuba	297,50	0,00	0,00	182,60	0,00
Insumo productivo	Cuba	290,30	0,00	0,00	0,80	0,00

Figura 19: Reporte Candidato

### 3.6 Configurar la seguridad de los usuarios

En Postgres se ofrece la opción de agrupar los usuarios según los roles y permisos de acceso que ellos necesitan para trabajar con el sistema.

➤ **Usuarios y roles:**

A continuación se describen los roles de los usuarios del sistema para garantizar la seguridad de este:

- **Especialista de información:** su rol se basa en consultar la información de la BD.
- **Administrador de datos:** su rol consiste en el acceso para desarrollar el proceso de Extracción, Transformación y Carga de los datos.
- **Administrador:** tiene el total acceso a la BD, tanto para la administración como la configuración del sistema.

➤ **Privilegios:**

Los privilegios se conceden a los usuarios según el rol que desempeñan:

- **Especialista de información:** se le es asignado los privilegios de Select los datos almacenados en el Almacén.
- **Administrador de Datos:** se le es asignado los privilegios de Select, Insert, Update, Delete, Refresh y Trigger de los datos almacenados en el Almacén.
- **Administrador:** se le asigna los derechos de Owner, Select, Update, Insert, Delete, Refresh y Trigger sobre la estructura de la base de datos.

### **3.7 Conclusiones del capítulo**

A lo largo de este capítulo se realizó la implementación la propuesta de un MD, sobre el área de Agricultura, ganadería y silvicultura del SIGOB. Posteriormente de haber realizado todos los procedimientos desarrollados en el capítulo se obtuvieron los siguientes resultados:

- Se realizó el diseño del subsistema de integración desarrollando los registro de sistemas fuentes, el diseño un diccionario de datos y de un mapa lógico de datos.
- Se implementaron 15 flujos de transformaciones.
- Se realizó la implementación de tres trabajos.
- Se definieron tres áreas de análisis, cuatro libros de trabajo y seis reportes.
- Para la implementación del subsistema de visualización se implementaron cuatro cubos OLAP.
- Se realizó la configuración de la seguridad del sistema definiendo usuarios y roles.

## **Capítulo 4: Validación del Mercado de Datos Agricultura, ganadería y silvicultura**

### **4.1 Introducción**

En el capítulo se utiliza el Modelo V para la validación del Mercado de Datos, donde se realizan pruebas unitarias, de integración y de aceptación con el cliente. Se diseñan y aplican casos de prueba y listas de chequeo para validar la solución.

### **4.2 Gestión de la Calidad de Software**

A través de los años el hombre siempre ha trabajado en función del perfeccionamiento, pero resulta imposible lograr tal perfección, el término “Gestión de la Calidad de Software” ilustra la perfección, ya que se basa en la determinación y aplicación de las políticas de calidad de la empresa (objetivos y directrices generales), es decir, técnicas y actividades de carácter operativo, utilizadas para satisfacer los requerimientos relativos a la calidad. La Gestión o Administración de la Calidad se aplica normalmente a nivel empresa, aunque puede haber gestión de la calidad dentro de la gestión de cada proyecto. El aspecto a considerar en el Control de la Calidad del Software es la Prueba del Software.

#### **Pruebas de software**

Las pruebas de software son los procesos que se le realizan al producto para verificar su correcto funcionamiento y calidad. Es el proceso de ejecutar un programa con intención de encontrar defectos, en estas se integra las técnicas de diseño de casos de prueba en una serie de pasos bien planificados. Estos casos de prueba dan como resultado una correcta construcción del software, con el objetivo principal de identificar dificultades en temas de implementación, o de manera general en la usabilidad del sistema. Las pruebas demuestran hasta qué punto las funciones del software parecen funcionar de acuerdo con las especificaciones y parecen alcanzarse los requerimientos de rendimiento. La definición de cuándo y cómo hacerlas varía en dependencia de la metodología de desarrollo utilizada.

Para mejorar el desarrollo de software existen varios modelos que ayudan al proceso de aplicación de pruebas. En el desarrollo de Soluciones de Almacenes de Datos se utiliza el modelo V, el cual es una variación del modelo en cascada que demuestra cómo se relacionan las actividades de prueba con las de análisis y diseño. Se puede ver que la punta de la V es la codificación, con el análisis y el diseño a la izquierda y la prueba y el mantenimiento a la derecha. La vinculación entre la parte izquierda y derecha del modelo V implica que si se encuentran problemas durante la verificación y la validación

entonces el lado izquierdo de la V puede ser ejecutado nuevamente para solucionar el problema y mejorar los requerimientos, el diseño y el código antes de retomar las pruebas del lado derecho.

El modelo V sugiere que la prueba unitaria y de integración sea hecha también para verificar el diseño, verificando que todos los aspectos del diseño del programa se han implementado correctamente en el código.

- **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 interface de los módulos y su comportamiento real. (30)
- **Prueba de integración:** es el proceso en el cual los componentes son agregados para crear componentes más grandes. Es el testing realizado para mostrar que aunque los componentes hayan pasado satisfactoriamente las pruebas unitarias, la combinación de componentes es incorrecta o insatisfactoria. (31)

El modelo V plantea que la prueba del sistema debe verificar el diseño del sistema, asegurando que todos los aspectos del diseño están correctamente implementados.

La mayoría de las faltas funcionales deben haber sido identificadas ya durante las pruebas de unidad e integración. La prueba del sistema generalmente se considera apropiada para probar los requerimientos no funcionales del sistema, tales como seguridad, desempeño, exactitud, y confiabilidad. Las interfaces externas, los dispositivos de hardware, o el ambiente de funcionamiento también se evalúan a este nivel. (32)

Mediante el modelo V se define que la prueba de aceptación que es dirigida por el cliente, valida los requerimientos. Es el proceso de comparar el programa contra sus requerimientos iniciales y las necesidades reales de los usuarios. Es realizado generalmente por el cliente o el usuario final. En el caso de un programa por contrato, se realiza la prueba de aceptación comparando la operación del programa contra el contrato original. (33)

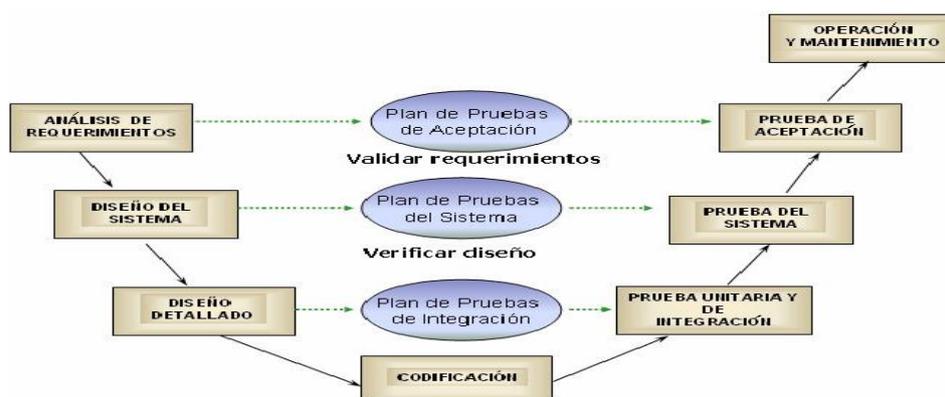


Figura 20: Modelo V

Después de haber aplicado las pruebas de modelo V a la aplicación se obtuvieron los siguientes resultados:

- Los especialistas de análisis del Centro realizaron las pruebas unitarias, donde no se encontraron no conformidades.
- La especialista de la ONE realizó las pruebas de aceptación y encontró dos no conformidades.
- Actualmente se encuentra la aplicación en las pruebas del sistema, llevada a cabo por Calidad UCI y las de integración por los especialistas del Centro.

#### 4.3 Casos de prueba

Los casos de prueba son un conjunto de valores de entrada, precondiciones de ejecución, resultados esperados y postcondiciones de ejecución, desarrollados con un objetivo particular o condición de prueba, tal como ejercitar un camino de un programa particular o para verificar que se cumple un requerimiento específico. (34)

Los casos de pruebas fueron diseñados con el objetivo de validar la solución propuesta. Para llevar a cabo la validación se realizó el diseño de tres casos de prueba para el área de Agricultura, ganadería y silvicultura.

La aplicación de los casos de prueba se encuentra reflejada en los artefactos Caso de prueba Agricultura, Caso de prueba Ganadería y Caso de prueba Silvicultura. En la siguiente imagen, se muestra un fragmento del Caso de Prueba correspondiente al Caso de Uso “Consultar indicadores de ganadería”:

Escenario	Descripción	Perfiles de análisis	Indicadores a medir	Respuesta del sistema	Flujo central
EC1.1: Indicadores de ganadería por DPA y mes	Permite visualizar reportes con las variables presentes en el mismo.	<ul style="list-style-type: none"> <li>• DPA</li> <li>• Temporal mes</li> <li>• Indicadores ganadería</li> </ul>	<ul style="list-style-type: none"> <li>• VACUNO</li> <li>• EQUINO</li> <li>• OVINO</li> <li>• CAPRINO</li> <li>• PORCINO</li> <li>• CUNÍCULA</li> <li>• AVÍCOLA</li> <li>• BÚFALO</li> </ul>	El sistema muestra todas las variables disponibles para los análisis, ubicados en las filas y las columnas que pueden ser visualizadas para cada reporte.	Se autentifica. Se entra al sistema. Se despliega hacia la derecha el componente ubicado en el lateral izquierdo que contiene el navegador. Se selecciona el área de análisis de <b>A.A Agricultura, ganadería y silvicultura.</b>

Figura 21: Casos de prueba Ganadería

#### 4.4 Elaboración y evaluación de las listas de chequeo

También, con el propósito de validar los artefactos desarrollados a partir de un grupo de indicadores, se decide utilizar una lista de chequeo general, previamente definida por el centro. Para una mayor organización, la lista de chequeo se divide en tres secciones que agrupan un total de 16 indicadores de los cuales 8 tienen un peso crítico. Estas secciones son:

- **Estructura del documento:** abarca todos los aspectos definidos por el expediente de proyecto o el formato establecido por el proyecto.
- **Indicadores definidos por la etapa:** abarca todos los indicadores a evaluar durante la etapa de análisis de datos.
- **Semántica del documento:** contempla todos los indicadores a evaluar respecto a la ortografía, redacción y demás.

#### 4.5 Resultados de las pruebas

Seguidamente se detallará de forma breve cada una de las pruebas aplicadas, así como los resultados obtenidos en cada una de ellas:

- **A nivel de departamento**

Se realizaron pruebas internas por parte de los especialistas del departamento Almacenes de Datos, donde no se identificaron no conformidades.

➤ **A nivel de centro**

Se realizaron pruebas por parte del grupo de especialistas de calidad del centro DATEC, donde no se identificaron no conformidades.

➤ **A nivel de universidad**

Se realizaron pruebas por parte del grupo de calidad de la UCI (CALISOFT), donde no se identificaron no conformidades.

➤ **Pruebas de aceptación**

El cliente realizó pruebas al sistema para verificar no sólo el cumplimiento de los requerimientos definidos, sino también el funcionamiento de la aplicación desde el punto de vista de un usuario final. Se detectaron dos no conformidades relacionadas con la vista de la aplicación, las cuales fueron resueltas.

#### **4.6 Conclusiones del capítulo**

A lo largo de este capítulo se realizó la validación de la propuesta del MD, sobre el área de Agricultura, ganadería y silvicultura del SIGOB. Posteriormente de haber realizado todos los procedimientos desarrollados en el capítulo se obtuvieron los siguientes resultados:

- Se configuró la seguridad del sistema definiendo los roles y los privilegios de los usuarios que van a interactuar con la aplicación según las funciones que estos realizan.
- Se realizaron las pruebas unitarias.
- Se aplicaron las pruebas de integración
- Se realizó la prueba de aceptación con el cliente.
- Se diseñaron y aplicaron tres casos de prueba para validar la solución.

## **Conclusiones**

Las principales deficiencias que presenta la ONE en la gestión de los datos dieron lugar a la elaboración del presente trabajo de diploma donde se realizó el desarrollo de un MD para el área de Agricultura, ganadería y silvicultura. Este mercado arrojó resultados satisfactorios para el aumento de la disponibilidad de información y el proceso de toma de decisiones, algunos de estos resultados son:

- Se realizó el análisis y diseño de los datos del área Agricultura, ganadería y silvicultura.
- Se diseñaron e implementaron los subsistemas de integración y visualización para lograr la implementación del sistema.
- Se elaboró una lista de chequeo que fue aplicada a la solución para validar el desarrollo.
- Se realizaron las pruebas de aceptación, unidad y pruebas pilotos.
- Se diseñaron y aplicaron los casos de pruebas con el objetivo de validar la solución propuesta.

## **Recomendaciones**

- Utilización de la información generada y la documentación para proyectos futuros similares, siempre cumpliendo con las normas y políticas de confidencialidad requeridos.
- Tomar como referencia la presente investigación con el objetivo de realizar un trabajo de consultoría a los distintos organismos del país para capacitar a los clientes sobre las facilidades y ventajas de los almacenes de datos.
- Integrar el Almacén de Datos al Sistema de Información de Gestión Estadística.

---

**Bibliografía**

- [www.one.cu](http://www.one.cu). [En línea]
- Data Warehouse vs Data Mart. [En línea] [Citado el: 3 de Noviembre de 2010.] <http://www.dataprix.com/datawarehouse-vs-datamart.7>.
- Data\_warehouse. [En línea] [Citado el: 3 de Noviembre de 2010.] [http://www.worldlingo.com/ma/enwiki/es/Data\\_warehouse](http://www.worldlingo.com/ma/enwiki/es/Data_warehouse). 5.
- Almacenes-datos. [En línea] [Citado el: 3 de Noviembre de 2010.] <http://www.monografias.com/trabajos31/almacenes-datos/almacenes-datos.shtml>. 4.
- Roberto Hernando Velasco. AD (Datawarehouse). Rhernando. [En línea] [Citado el: 3 de Noviembre de 2010.] <http://www.rhernando.net/modules/tutorials/doc/bd/dw.html>. 3.
- Kimball. *Kimball(2002)*. 2002.
- Velasco, Roberto Hernando. Rhernando. [En línea] [Citado el: 3 de Noviembre de 2010.] <http://www.rhernando.net/modules/tutorials/doc/bd/dw.html>. 2..
- Sinnexus. [En línea] [Citado el: 3 de Noviembre de 2010.] [http://www.sinnexus.com/business\\_intelligence/datamart.aspx](http://www.sinnexus.com/business_intelligence/datamart.aspx) bibliografía. 8.
- Data Warehouse vs Data Mart. [En línea] [Citado el: 10 de Noviembre de 2010.] <http://www.dataprix.com/datawarehouse-vs-datamart>. 9..
- Bases de datos OLTP y OLAP. [En línea] [Citado el: 10 de Noviembre de 2010.] [http://www.sinnexus.com/business\\_intelligence/olap\\_vs\\_oltp.aspx](http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx). 10..
- Sinnexus. [En línea] [Citado el: 10 de Noviembre de 2010.] [http://www.sinnexus.com/business\\_intelligence/olap\\_vs\\_oltp.aspx](http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx). 11.
- Sousa, Ing. Jose Zarate. *Datamarting Enfocado al Usuario*.
- ETL para Analytics. [En línea] [Citado el: 4 de diciembre de 2010.] <http://es.talend.com/solutions-data-integration/etl-for-analytics.php>.
- <http://www.jatun.com/web/company/products/etl>. *ETL*. [En línea] [Citado el: ]
- Herramientas. [En línea] [Citado el: 25 de Febrero de 2011.] <http://www.dataprix.com/blogs/respinosamilla/herramientas-etl-que-son-para-que-valen-productos-mas-conocidos-etl-s-open-sour>. 13..
- OrlandoOlguin. [En línea] [Citado el: 25 de Febrero de 2011.] <http://orlandoolguin.wordpress.com/2008/08/12/inteligencia-de-negocio-%C2%BFque-es/>. 14.

- Falcon Rodriguez, Yolanda y Leyva Osorio, Reynaldo. Mercado de Datos Estadístico de Inmigración y Extranjería para el Departamento de Turismo y Comercio de la Oficina Nacional de Estadísticas. [En línea]
- Sinnexus. [En línea] [Citado el: 3 de Noviembre de 2010.] [http://www.sinnexus.com/business\\_intelligence/datamart.aspx](http://www.sinnexus.com/business_intelligence/datamart.aspx) bibliografía. 6..
- DATAMART PASO A PASO. [En línea] [Citado el: 10 de Febrero de 2011.] <http://www.u-cursos.cl/ingenieria/2008>. 15..
- Definición de herramientas de modelado. [En línea] [Citado el: 15 de Febrero de 2011.] <http://www.alegsa.com.ar/.../herramienta%20de%20modelado.php>. 16..
- Visual Paradigm for UML. [En línea] [http://www.freedownloadmanager.org/es/downloads/Paradigma\\_Visual\\_para\\_UML\\_%28M%C3%8D%29\\_14720\\_p/](http://www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_%28M%C3%8D%29_14720_p/). 17..
- Importancia de la captura de requisitos. [En línea] [Citado el: 6 de Noviembre de 2010.] <http://www.articuloz.com/otro-articulos/importancia-de-la-captura-de-requerimientos-en-el-proceso-productivo-del-proyecto-ciscop-de-la-facultad-regional-de-artemisa-importance-of-the-captu-636738.html>. 18..
- Datacleaner. [En línea] [Citado el: 3 de diciembre de 2010.] <http://datacleaner.eobjects.org>.
- Gravatar. [En línea] <http://www.gravatar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>. 19..
- Mondrian. [En línea] [Citado el: 4 de diciembre de 2010.] <http://mondrian.pentaho.com>.
- *Professional Apache Tomcat 6*. Chopra, Vivek, Li, Sing y Genender, Jeff. Indianapolis : s.n., 2007.
- Business Intelligence. [En línea] [Citado el: 16 de Febrero de 2011.] <http://www.gestiopolis.com/recursos5/docs/ger/buconce.htm>. 20.
- Lamancha, Beatriz Pérez. Proceso de Testing Funcional Independiente. Montevideo, Uruguay : s.n., 2006, pág. 164.
- Perdomo, Yeniset León. Estrategia de Pruebas Exploratorias para mejorar el rendimiento del Laboratorio Industrial de Pruebas de Software de Calisoft. Ciudad de la Habana : s.n., 2011.
- ALVAREZ. S. Sistemas gestores de bases de datos. [En línea] [Citado el: 20 de Septiembre de 2010.] <http://www.desarrolloweb.com/articulos/sistemas-gestores-bases-datos.html>. 21.
- Kimball, Ralph y Caserta, Joe. The Data Warehouse ETL Toolkit Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. s.l.: Wiley Publishing. ISBN: 0-764-57923-1.

- E, Kit. *Software Testing In The Real World : Improving The Process* . s.l. : Addison Wesley , 1995.
- B., Beizer. “*Software testing techniques (2nd ed.)*”. s.l. : Van Nostrand , 1990.
- Guide to the Software Engineering Body of Knowledge (SWEBOK). [En línea] [Citado el: 31 de mayo de 2011.] <http://www.swebok.org>.
- Sons, John Wiley &. *The art of software testing, 2nd edition* . New Jersey : s.n., 2004.
- **EEE-Std.** *IEEE Standard Glossary of Software Engineering Terminology*. 2002.

## Referencias bibliográficas

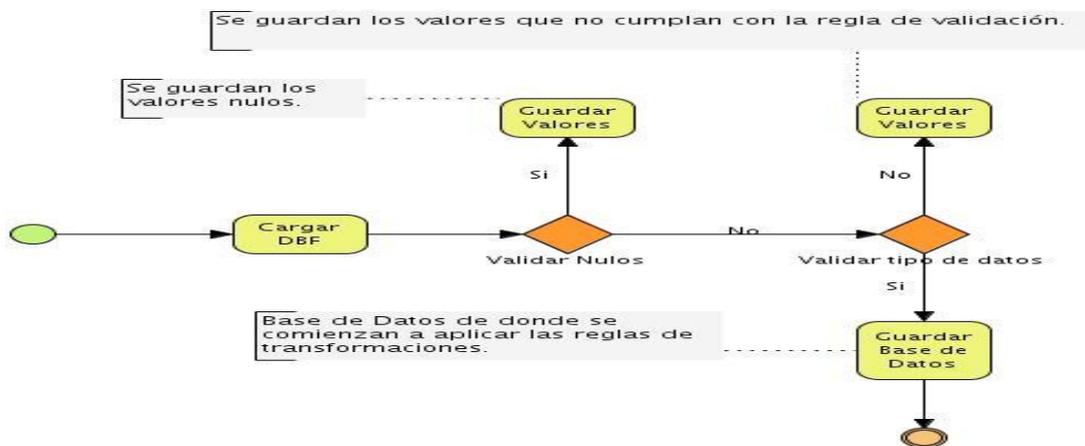
1. www.one.cu. [En línea]
2. Data Warehouse vs Data Mart. [En línea] [Citado el: 3 de Noviembre de 2010.] <http://www.dataprix.com/datawarehouse-vs-datamart.7>.
3. Data\_warehouse. [En línea] [Citado el: 3 de Noviembre de 2010.] [http://www.worldlingo.com/ma/enwiki/es/Data\\_warehouse](http://www.worldlingo.com/ma/enwiki/es/Data_warehouse). 5.
4. Almacenes-datos. [En línea] [Citado el: 3 de Noviembre de 2010.] <http://www.monografias.com/trabajos31/almacenes-datos/almacenes-datos.shtml>. 4.
5. Roberto Hernando Velasco. AD (Datawarehouse). Rhernando. [En línea] [Citado el: 3 de Noviembre de 2010.] <http://www.rhernando.net/modules/tutorials/doc/bd/dw.html>. 3.
6. Kimball. *Kimball(2002)*. 2002.
7. Velasco, Roberto Hernando. Rhernando. [En línea] [Citado el: 3 de Noviembre de 2010.] <http://www.rhernando.net/modules/tutorials/doc/bd/dw.html>. 2..
8. Sinnexus. [En línea] [Citado el: 3 de Noviembre de 2010.] [http://www.sinnexus.com/business\\_intelligence/datamart.aspx](http://www.sinnexus.com/business_intelligence/datamart.aspx) bibliografía. 8.
9. Data Warehouse vs Data Mart. [En línea] [Citado el: 10 de Noviembre de 2010.] <http://www.dataprix.com/datawarehouse-vs-datamart>. 9..
10. Bases de datos OLTP y OLAP. [En línea] [Citado el: 10 de Noviembre de 2010.] [http://www.sinnexus.com/business\\_intelligence/olap\\_vs\\_oltp.aspx](http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx). 10..
11. Sinnexus. [En línea] [Citado el: 10 de Noviembre de 2010.] [http://www.sinnexus.com/business\\_intelligence/olap\\_vs\\_oltp.aspx](http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx). 11.
12. Sousa, Ing. Jose Zarate. *Datamarting Enfocado al Usuario*.
13. ETL para Analytics. [En línea] [Citado el: 4 de diciembre de 2010.] <http://es.talend.com/solutions-data-integration/etl-for-analytics.php>.
14. <http://www.jatun.com/web/company/products/etl>. *ETL*. [En línea] [Citado el: ]
15. Herramientas. [En línea] [Citado el: 25 de Febrero de 2011.] <http://www.dataprix.com/blogs/respinosamilla/herramientas-etl-que-son-para-que-valen-productos-mas-conocidos-etl-s-open-sour>. 13..

16. Orlandoolguin. [En línea] [Citado el: 25 de Febrero de 2011.] <http://orlandoolguin.wordpress.com/2008/08/12/inteligencia-de-negocio-%C2%BFque-es/>. 14.
17. Falcon Rodriguez, Yolanda y Leyva Osorio, Reynaldo. Mercado de Datos Estadístico de Inmigración y Extranjería para el Departamento de Turismo y Comercio de la Oficina Nacional de Estadísticas. [En línea]
18. Sinnexus. [En línea] [Citado el: 3 de Noviembre de 2010.] [http://www.sinnexus.com/business\\_intelligence/datamart.aspx](http://www.sinnexus.com/business_intelligence/datamart.aspx) bibliografía. 6..
19. DATAMART PASO A PASO. [En línea] [Citado el: 10 de Febrero de 2011.] <http://www.u-cursos.cl/ingenieria/2008>. 15..
20. Definición de herramientas de modelado. [En línea] [Citado el: 15 de Febrero de 2011.] <http://www.alegsa.com.ar/.../herramienta%20de%20modelado.php>. 16..
21. Visual Paradigm for UML. [En línea] [http://www.freedownloadmanager.org/es/downloads/Paradigma\\_Visual\\_para\\_UML\\_%28M%C3%8D%29\\_14720\\_p/](http://www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_%28M%C3%8D%29_14720_p/). 17..
22. Importancia de la captura de requisitos. [En línea] [Citado el: 6 de Noviembre de 2010.] <http://www.articuloz.com/otro-articulos/importancia-de-la-captura-de-requerimientos-en-el-proceso-productivo-del-proyecto-ciscop-de-la-facultad-regional-de-artemisa-importance-of-the-captu-636738.html>. 18..
23. Datacleaner. [En línea] [Citado el: 3 de diciembre de 2010.] <http://datacleaner.eobjects.org>.
24. Gravatar. [En línea] <http://www.gravatar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>. 19..
25. Mondrian. [En línea] [Citado el: 4 de diciembre de 2010.] <http://mondrian.pentaho.com>.
26. *Professional Apache Tomcat 6*. Chopra, Vivek, Li, Sing y Genender, Jeff. Indianapolis : s.n., 2007.
27. Business Intelligence. [En línea] [Citado el: 16 de Febrero de 2011.] <http://www.gestiopolis.com/recursos5/docs/ger/buconce.htm>. 20.
28. ALVAREZ. S. Sistemas gestores de bases de datos. [En línea] [Citado el: 20 de Septiembre de 2010.] <http://www.desarrolloweb.com/articulos/sistemas-gestores-bases-datos.html>. 21.
29. Kimball, Ralph y Caserta, Joe. *The Data Warehouse ETL Toolkit Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. s.l.: Wiley Publishing. ISBN: 0-764-57923-1.
30. E, Kit. *Software Testing In The Real World : Improving The Process* . s.l. : Addison Wesley , 1995.
31. B., Beizer. *“Software testing techniques (2nd ed.)”*. s.l. : Van Nostrand , 1990.
32. Guide to the Software Engineering Body of Knowledge (SWEBOK). [En línea] [Citado el: 31 de mayo de 2011.] <http://www.swebok.org>.

33. Sons, John Wiley &. *The art of software testing, 2nd edition* . New Jersey : s.n., 2004.
34. EEE-Std. *IEEE Standard Glossary of Software Engineering Terminology*. 2002.



**Anexo3. Diseño de las transformaciones**



## Glosario

**Área de Análisis:** Se le denomina Área de análisis a la 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.

**Cubo:** Colección de dimensiones y medidas en un área temática particular.

**Diccionario de datos:** artefacto que se utiliza para definir la correspondencia que existe entre las variables de las tablas, bases de datos o modelos determinados con sus perspectivas.

**Dimensión:** característica de un hecho que permite su análisis posterior en el proceso de toma de decisiones y brinda una perspectiva adicional a un hecho dado.

**Granularidad:** representa el nivel de detalle con el que se desea almacenar la información sobre el negocio analizado.

**Hecho:** operación que se realiza en el negocio la cual está estrechamente relacionada con el tiempo y es objeto de análisis para la toma de decisiones.

**Indicador:** generalmente es un valor numérico y representan lo que se desea analizar concretamente.

**Jerarquía:** implica una organización de niveles dentro de una dimensión, donde cada nivel representa el total agregado de los datos del nivel inferior.

**Libro de Trabajo:** Estructura organizativa que agrupa los reportes generados dentro de las áreas de análisis. Puede ser creado teniendo en cuenta criterio que permitan organizar la información: Emisor de los reportes, receptor del reporte, contenido, entre otros.

**Lista de chequeo:** instrumento de medición y evaluación que consiste básicamente en un formulario de preguntas referentes al atributo de calidad que se está probando y de las características del documento en el caso de la documentación.

**Multidimensional:** La información multidimensional se puede encontrar en hojas de cálculo, bases de datos, etc. Una herramienta de Inteligencia de Negocios debe de ser capaz de reunir información dispersa en toda la empresa e incluso en diferentes fuentes para así proporcionar a los departamentos la accesibilidad, poder y flexibilidad que necesitan para analizar la información.