

**Universidad de las Ciencias Informáticas  
Facultad Regional  
“Mártires de Artemisa”**



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

**Título: Mercado de Datos para la Delegación  
Provincial de la Agricultura de Artemisa**

**Autoras:** Adilen Guerra Sanabria  
Ailen Delgado González

**Tutor:** Ing. Ricardo Jorge Hera

**Co-tutor:** Ing. Dennys Lázaro Hernández Quintana

**Artemisa, Junio del 2012**

## Declaración de Autoría

Declaramos ser autoras de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

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

\_\_\_\_\_  
**Autora: Adilen Guerra Sanabria**

\_\_\_\_\_  
**Autora: Ailen Delgado González**

\_\_\_\_\_  
**Tutor: Ing. Ricardo Jorge Hera**

## **Frase**

“La inteligencia consiste no sólo en el conocimiento, sino también en la destreza de aplicar los conocimientos en la práctica”

Aristóteles

## **Dedicatoria**

Dedicada a esa personita que jugó un papel fundamental en mi vida, que mientras mi madre trabajaba me cuidó y me enseñó con mucho amor, quien estuvo pendiente de mí durante 21 años, una persona que hacía hasta lo imposible por verme feliz, a mi abuelita linda que hoy no se encuentra entre nosotros. A mi mamá, mi papá y mi hermano por ser la razón de mí existir. A mi familia, por toda su dedicación y cariño.

**Adilen**

A mi abuelita Elsa y mi tía Tata, dos personas que se entregaron en cuerpo y alma a mi crianza e hicieron de mí la persona que soy. A mi mamá que me ha apoyado siempre en los momentos buenos y malos, solo de saber que está a mi lado me da fuerzas para lograr cualquier cosa. A Félix que siempre ha estado cerca de mí ayudándome en todo, para él nunca ha habido nada imposible cuando se trata de mí. A mi papá que ha representado siempre un ejemplo a seguir y un motivo para esforzarme y llegar a ser universitaria. A mi hermana por acordarse de mí incluso cuando no estoy en la casa y por motivarme a trabajar duro para estar a su altura. A Mayra y Claudia, por ser las mejores amigas que he tenido, darme ánimo, ayudarme y hacerme sentir capaz. A Henry por ayudarme siempre a pensar.

**Ailen**

## **Adilen**

A mis padres por ser las personas más importante en mi vida, quienes me dieron la vida y han guiado mi camino con mucho amor y dedicación, por apoyarme en todo y darme siempre ese voto de confianza. A mi hermano por ser esa persona tan especial en mi vida. A mi familia por estar siempre conmigo en las buenas y en las malas, por darme la fuerza necesaria para levantarme y seguir cada día. A Merli por ser la amiga que nunca tuve y por brindarme toda su ayuda y comprensión cuando más lo necesité. Al Nene que siempre cuidó de mí como si fuera una hermana. A Johanna, Evelyn y Yuslenis por estos lindos 5 años juntas. A Maura y Adrián por estar siempre pendientes de mí. A mi compañera de tesis Ailen por todos los momentos que pasamos juntas trabajando y estudiando. A Sonia y Diana que han sido un gran apoyo para la realización de la tesis. A mi tutor y co-tutor por su colaboración. Al profe Javier por toda su ayuda con paciencia y dedicación. A todas las personas que de una forma u otra me han ayudado a lo largo de mi carrera.

## **Ailen**

A mi familia por todo el apoyo que he recibido durante estos 5 años. Al profesor Javier por la gran cantidad de dudas aclaradas y todo el tiempo que me ha dedicado. A mi profe favorito, Raúl Crespo, por brindarme su ayuda y darme seguridad. A Rolando y los muchachos de UCI-FAR. A Sonia y Diana, dos integrantes más de esta tesis. A Henry por todas las madrugadas de debates y las buenas ideas que me dio. A Orlando y Mayra por estar siempre dispuestos a ayudar y haber puesto su granito de arena. Al profesor Ramón Justiniani por su aporte y buenas respuestas. A mi compañera de tesis Adilen por su paciencia para guiarme por buen camino. A mi tutor y co-tutor por defendernos y colaborar durante la realización. A Yosdani que siempre estuvo presente en los viajes a la Delegación y por el tiempo de estudio juntos. A todas las personas que han colaborado brindando su ayuda.

La presente investigación se desarrolla en el marco de la colaboración existente entre la Facultad Regional Mártires de Artemisa y la Administración Provincial de Artemisa. Este último es el organismo encargado de planificar y proponer la política integral del estado, y una vez aprobada dicha política, dirige, coordina y controla su cumplimiento sobre la base de las estrategias de desarrollo del país. La agricultura es una de las principales ramas que aporta beneficios a la economía de la provincia y la Delegación Provincial de la Agricultura es el centro de control de todas las actividades agropecuarias.

Los datos que se recogen en la Delegación de los distintos procesos que comprende el negocio se han ido acumulando, y actualmente es considerable el volumen de datos históricos almacenado en los ficheros Excel, situación que hace engorroso y difícil el análisis de la información para el proceso de toma de decisiones. El presente trabajo de diploma tiene como objetivo principal desarrollar un mercado de datos para la Delegación Provincial de la Agricultura con el fin de optimizar el tiempo de respuesta del sistema, posibilitando un adecuado análisis de la información. Para dar cumplimiento a ello, se realizó un estudio de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de los almacenes de datos. Se realizó el análisis, diseño e implementación obteniéndose un mercado de datos que cumple con los requerimientos del negocio. Por último, se realizaron pruebas unitarias y funcionales al sistema implementado obteniendo resultados satisfactorios.

<b>INTRODUCCIÓN.....</b>	<b>1</b>
<b>CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA DE LOS MERCADOS DE DATOS .7</b>	
1. MARCO TEÓRICO .....	7
2. TENDENCIAS ACTUALES .....	17
3. METODOLOGÍAS DE DESARROLLO.....	20
4. DESCRIPCIÓN DE LAS HERRAMIENTAS UTILIZADAS .....	23
<b>CAPÍTULO 2. ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS DE LA DELEGACIÓN PROVINCIAL DE LA AGRICULTURA DE ARTEMISA.....</b>	<b>28</b>
2.1. ESTUDIO PRELIMINAR DEL NEGOCIO .....	28
2.2. TEMAS DE ANÁLISIS .....	29
2.3. REGLAS DEL NEGOCIO.....	29
2.4. ESPECIFICACIÓN DE REQUERIMIENTOS .....	30
2.5. CASOS DE USO DEL SISTEMA .....	32
2.6. DIAGRAMA DE CASOS DE USO DEL SISTEMA .....	33
2.7. ESPECIFICACIÓN DE CASOS DE USO DEL SISTEMA .....	35
2.8. NIVEL DE GRANULARIDAD.....	35
2.9. IDENTIFICACIÓN DE LAS DIMENSIONES, HECHOS Y MEDIDAS. ....	36
2.10. MODELO DIMENSIONAL.....	38
2.11. ARQUITECTURA DE LA INFORMACIÓN .....	42
<b>CAPÍTULO 3: IMPLEMENTACIÓN DEL MERCADO DE DATOS PARA LA DELEGACIÓN PROVINCIAL DE LA AGRICULTURA.....</b>	<b>44</b>
3.1. IMPLEMENTACIÓN DE LAS ESTRUCTURAS DE DATOS .....	44
3.2. ROLES Y PRIVILEGIOS DE ACCESO .....	45
3.3. IMPLEMENTACIÓN DEL SUBSISTEMA DE INTEGRACIÓN .....	46
3.4. IMPLEMENTACIÓN DEL SUBSISTEMA DE VISUALIZACIÓN .....	51
3.5. TÉCNICAS DE OPTIMIZACIÓN .....	55
<b>CAPÍTULO 4: VALIDACIÓN DE LAS FUNCIONALES DEL MERCADO DE DATOS PARA LA DELEGACIÓN PROVINCIAL DE LA AGRICULTURA.....</b>	<b>59</b>
4.1. PROCESO DE PRUEBA.....	59
4.2. PRUEBAS APLICADAS AL MERCADO DE DATOS PARA LA DELEGACIÓN PROVINCIAL DE LA AGRICULTURA .....	60
4.3. RESULTADOS OBTENIDOS .....	61
4.4. APOORTE SOCIAL Y ECONÓMICO.....	62
<b>CONCLUSIONES GENERALES .....</b>	<b>63</b>
<b>RECOMENDACIONES .....</b>	<b>64</b>
<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>65</b>
<b>BIBLIOGRAFÍA CONSULTADA .....</b>	<b>67</b>
<b>GLOSARIO DE TÉRMINOS.....</b>	<b>70</b>

<b>ANEXOS .....</b>	<b>73</b>
ANEXO 1. CÓDIGO DE DISPARADOR QUE EJECUTA LA FUNCIÓN PARA ACTUALIZAR LA VISTA MATERIALIZADA. ....	73
ANEXO 2. CÓDIGO DE FUNCIÓN QUE ACTUALIZA LA VISTA MATERIALIZADA .....	73
ANEXO 3. CÓDIGO DE DISPARADOR QUE EJECUTA LA FUNCIÓN QUE ELIMINA LOS REGISTROS DE LA VISTA MATERIALIZADA. ....	74
ANEXO 4. CÓDIGO DE FUNCIÓN QUE ELIMINA LOS REGISTROS DE LA VISTA MATERIALIZADA QUE NO PERTENEZCAN AL ÚLTIMO AÑO. ....	74
ANEXO 5. ENTREVISTA EFECTUADA A LOS TRABAJADORES DE LA DELEGACIÓN PROVINCIAL DE LA AGRICULTURA DE ARTEMISA. ....	74

## Imágenes

Imagen 1: Cubo Multidimensional .....	11
Imagen 2: Diagrama de CU correspondiente al tema de análisis Ganadería .....	34
Imagen 3: Diagrama de CU correspondiente al proceso de ETL .....	34
Imagen 4: Diagrama de CU correspondiente al tema de análisis Cultivos Varios ...	35
Imagen 5 : Esquema dimensional de los procesos siembra, preparación y alimentación para la ganadería. ....	38
Imagen 6 : Esquema dimensional de los procesos producción de carne y leche...	39
Imagen 7: Esquema dimensional de los procesos contratación y demanda .....	39
Imagen 8 : Esquema dimensional de los procesos de preparación de tierra para la siembra de papa, maíz, cultivos varios y entrada de fertilizante.....	40
Imagen 9: Esquema dimensional del balance de existencia en nave y otros destinos. ....	41
Imagen 10: Esquema dimensional de los procesos de siembra de papa, frijol, maíz y entrada de semilla para siembra de papa. ....	42
Imagen 11: Mapa de Navegación .....	43
Imagen 12: Transformación para la carga del hecho existencia en nave .....	48
Imagen 13: Transformación para la carga del hecho siembra de frijol .....	49
Imagen 14: Trabajo poblar mercado de datos .....	50
Imagen 15: Trabajo que agrupa las cargas incrementales .....	51
Imagen 16: Diseño del cubo multidimensional del proceso producción de leche....	52
Imagen 17: Vista de análisis % de cumplimiento del plan de producción de leche.	53
Imagen 18: Reporte parte diario provincial de la preparación de tierras para ganadería .....	54
Imagen 19: Código para la creación de índices.....	56

## Tablas

Tabla 1: Ventajas y desventajas de los modos de almacenamiento. ....	13
Tabla 2: Roles y privilegios de acceso al mercado de datos .....	46
Tabla 3: Roles y privilegios de acceso a la aplicación .....	46
Tabla 4: Detalles de los niveles de pruebas.....	60

## **Introducción**

El gran desarrollo que han tenido las Tecnologías de la Información y las Comunicaciones (TIC) en los últimos tiempos ha propiciado que los avances tecnológicos sean cada vez más accesibles a los países subdesarrollados, en vías de desarrollo y las pequeñas empresas. Esta situación está creando un nuevo escenario que permite una mejor utilización de los recursos productivos y por tanto el aumento de la competencia.

En el actual contexto competitivo, los modelos de negocio se encuentran en constante cambio y desarrollo dada la necesidad de mantenerse acorde a las exigencias del mercado. Frente a esta gran inestabilidad, el único recurso que da a las organizaciones una fuente segura para obtener ventaja en la competencia, es el conocimiento.

A raíz de esta situación la economía a nivel mundial avanza hacia un nuevo modelo basado en el conocimiento, situándolo por encima de los demás factores económicos. Se concibe que la formulación de estrategias de éxito se deba basar en la gestión de la información más que en la gestión de los activos tangibles, por tal razón cuanto más información útil tenga una empresa en su poder y mejor calidad de análisis, mayores serán las probabilidades de alcanzar el éxito.

Desde este punto de vista, la disponibilidad de la información es un factor que adquiere vital importancia, ya que las decisiones se suelen tomar sobre la base de dicha información. Reflejar evoluciones, hechos históricos en el negocio y posibilidades futuras, son procesos que la alta dirección de las organizaciones necesita manejar de forma habitual, más aún cuando se trata de una institución que influye directamente en la economía de una nación.

En la actualidad, las organizaciones poseen grandes volúmenes de datos que han sido almacenados con el propósito de satisfacer sus necesidades a la hora de tomar decisiones de negocio. En paralelo al proceso de informatización de la

sociedad, estos volúmenes tienden al crecimiento y se hace cada vez más difícil extraer información realmente útil haciendo uso de los métodos tradicionales.

Para lograr explotar al máximo las ventajas que ofrece todo este cúmulo de información se requiere proporcionar al usuario no solo la capacidad de acceso fácil, sino también de realizar análisis desde diferentes perspectivas para evaluar comportamientos, prever los resultados y poder anticipar tendencias. Este hecho ciertamente condiciona la necesidad de avanzar en conocimientos y medios para desarrollar soluciones que cumplan con las nuevas necesidades.

Cuba ha seguido a grandes pasos el progreso constante de las tecnologías, proponiéndose día a día nuevos retos. A pesar de ser un país en vías de desarrollo que cuenta con escasos recursos económicos, se han logrado grandes avances en la informatización de sectores como la salud y la educación. La Universidad de las Ciencias Informáticas (UCI) ha jugado un papel fundamental en el desarrollo de la industria cubana del software aportando grandes ingresos a la economía. Para ello se han creado diferentes centros de desarrollo de software, entre ellos el Centro de Tecnologías de Gestión de Datos (DATEC) que se especializa en el desarrollo de productos y prestación de servicios relacionados con las bases de datos y el análisis de datos.

La Facultad Regional Mártires de Artemisa colabora de forma activa en el progreso del país con el desarrollo de proyectos que tienen como fin llevar a cabo la informatización de diversos sectores económicos. Una de las numerosas metas que se ha trazado dicha facultad, es llevar a cabo la realización del Sistema Informativo de la Administración Provincial (AP) de Artemisa.

La AP es el organismo encargado de planificar y proponer la política integral del estado, y una vez aprobada dicha política, dirige, coordina y controla su cumplimiento sobre la base de las estrategias de desarrollo del país. Su objetivo fundamental es formular las estrategias que logren hacer de los procesos económicos de la provincia una importante fuente de ingresos para el país.

En Artemisa la economía es fundamentalmente ganadera y agrícola, posicionándose de esta forma la agricultura como una de las principales ramas generadora de recursos para la provincia. La Delegación Provincial de la Agricultura es el centro de control de todas las actividades agropecuarias. Tiene una estructura organizativa compuesta por 3 direcciones, entre las que se encuentra la dirección de análisis, que tiene como función principal la captación y almacenamiento de los datos de producción.

Debido a la amplia gama de procesos que comprende el negocio y la periodicidad con que le llegan los datos, la Delegación Provincial de la Agricultura recibe gran cantidad de información estadística que se ha ido acumulando. Con el paso del tiempo, el volumen de estos datos históricos ha aumentado considerablemente, situación que hace del análisis de la información mediante los métodos tradicionales una tarea difícil, larga y tediosa que trae consigo pérdidas de tiempo.

En reiteradas ocasiones los directivos necesitan realizar análisis de comportamiento y tendencias con datos históricos. Generalmente el tiempo con que se cuenta para estas tareas es limitado, debido a la premura con que se necesita dicha información en niveles superiores. Los retrasos en la entrega de los informes influyen de forma negativa en el proceso de toma de decisiones.

Ante tal situación se plantea el siguiente **problema a resolver**: ¿Cómo agilizar el análisis de la información como parte del proceso de toma de decisiones en la Delegación Provincial de la Agricultura de Artemisa?

Para establecer los límites de la investigación se define como **objeto de estudio** a los almacenes de datos, reduciendo el **campo de acción** a los mercados de datos para el sector de la agricultura.

El **objetivo general** de la investigación es: Desarrollar un mercado de datos para la Delegación Provincial de la Agricultura de Artemisa que agilice el análisis de la información como parte del proceso de toma de decisiones.

Teniendo como **idea a defender** del presente trabajo: Con el desarrollo de un mercado de datos se agiliza el análisis de la información en la Delegación Provincial de la Agricultura de Artemisa.

Para dar cumplimiento al objetivo general se presentan las siguientes **tareas de investigación**:

1. Fundamentación del marco teórico de los almacenes de datos.
2. Caracterización del proceso de análisis de la información en la Delegación Provincial de la Agricultura.
3. Desarrollo del mercado de datos de la Delegación Provincial de la Agricultura.
4. Validación de las funcionalidades del mercado de datos de la Delegación Provincial de la Agricultura.

Al concluir el trabajo se espera obtener como **posible resultado**:

Mercado de datos para la Delegación Provincial de la Agricultura de Artemisa.

Los **métodos científicos** de investigación empleados son:

### **Métodos Teóricos**

**Histórico lógico:** Este método se utilizó para profundizar en las distintas etapas del desarrollo de los almacenes de datos teniendo en cuenta las metodologías y herramientas utilizadas, además de analizar soluciones similares en caso de existir alguna. Con dicha información se pudo conformar un marco teórico y la idea a defender del problema.

**Análisis-Síntesis:** Se empleó con el objetivo de analizar las características de las metodologías existentes para el desarrollo de los almacenes de datos y finalmente poder seleccionar entre todas las opciones la metodología adecuada según las particularidades del proyecto. Sirvió además para diagnosticar y analizar el problema desde la información obtenida en los diferentes métodos empíricos que se emplearon.

**Modelación:** Fue útil para lograr reproducir la realidad del negocio, los objetos y relaciones que lo componen. Su uso se pone de manifiesto en el diseño de la estructura del almacén de datos que incluye las tablas de hecho y las dimensiones.

## **Métodos Empíricos**

**Observación:** Permitió la realización de un estudio profundo de las actividades que se llevan a cabo en la Delegación Provincial de la Agricultura de Artemisa, lo que posibilitó obtener un amplio entendimiento del negocio e identificar los factores que afectan la toma de decisiones, las necesidades de información de los usuarios, los indicadores que se necesitan medir y establecer con claridad los objetivos que se persiguen con la realización del sistema.

**Entrevista:** Por medio de la entrevista se estableció una comunicación directa con el cliente. Las respuestas obtenidas a través del cuestionario aplicado (Ver Anexo 5) permitieron obtener información más completa sobre las necesidades de los usuarios, qué información se considera más importante, los pedidos más frecuentes, la periodicidad de obtención de los datos y que espera el usuario del producto.

## **Declaración de variables**

**Variable independiente:** Mercado de Datos para la Delegación Provincial de la Agricultura de Artemisa.

**Variables dependientes:** Análisis de la información.

La **estructura del documento** se compone de un resumen, introducción, 4 capítulos, conclusiones, recomendaciones, bibliografía, anexos y un glosario de términos.

**Capítulo 1:** Fundamentación teórica de los almacenes de datos: Se realiza una investigación sobre el desarrollo que han tenido los almacenes y mercados de datos a nivel internacional y nacional. Se estudian las diferentes metodologías y

herramientas que van a ser empleadas para solucionar el problema, así como las definiciones y características generales de este tipo de soluciones.

**Capítulo 2:** Análisis y diseño del mercado de datos para la Delegación Provincial de la Agricultura: Se realiza un estudio sobre las características del negocio, se identifican las necesidades de información, se definen los requerimientos del sistema, se diseña el diagrama de casos de uso del sistema, el modelo de datos, los subsistemas de integración y visualización.

**Capítulo 3:** Implementación del mercado de datos para la Delegación Provincial de la Agricultura: En este capítulo se presentan los aspectos relacionados con la implementación de los subsistemas de integración y visualización. Se especifican los detalles del proceso de extracción, transformación y carga de los datos, así como de la implementación de los cubos, los reportes candidatos y el establecimiento de la política de seguridad de los usuarios.

**Capítulo 4:** Validación de las funcionalidades del mercado de datos para la Delegación Provincial de Artemisa: En este capítulo se exponen los aspectos relacionados con las pruebas realizadas al mercado de datos y los resultados obtenidos.

## **Capítulo 1. Fundamentación teórica de los mercados de datos**

En este capítulo se realizó un estudio valorativo en el que se tuvo en cuenta la evolución que han tenido los almacenes y mercados de datos a nivel internacional y nacional. Para exponer los fundamentos teóricos que respaldan la solución se tuvo en cuenta las definiciones de almacén de datos y mercado de datos que han sido concebidas por los especialistas del tema, las características que los diferencian, las metodologías que existen para su desarrollo, definiendo cual será usada para la construcción del mercado de datos que se desarrollará. Se abordaron temas relacionados con los diferentes modos de almacenamiento y los esquemas multidimensionales, así como las herramientas que existen para el desarrollo de esta tecnología, de ellas cuáles se usarán y las razones de las elecciones realizadas.

### **1.Marco Teórico**

La información referente a las actividades de negocios constituyen una fuente invaluable de conocimiento, sin embargo en muchos de los casos las organizaciones disponen de grandes volúmenes de datos que ocupan espacio sin que se les de la utilización adecuada. En la búsqueda del aprovechamiento de recursos, surge la idea de utilizar los datos históricos que se van acumulando con el paso de los años para generar conocimiento que sirva de apoyo en la toma de decisiones. Esta nueva necesidad condiciona el surgimiento del concepto de almacén de datos.

#### **1.1.Almacén de datos**

Cuando se habla de almacenamiento de datos, es imposible dejar de mencionar a dos grandes personalidades que han incursionado en este campo: William Harvey Bill Inmon (reconocido por muchos como el padre del almacén de datos) y Ralph Kimball. Cada uno define el almacén de datos desde una concepción diferente y aunque ambos persiguen el mismo objetivo, difieren en la forma de conseguirlo.

# Capítulo 1: Fundamentación teórica de los mercados de datos

Esta situación ha derivado en dos filosofías diferentes que aún en la actualidad son utilizadas como guía en el proceso de desarrollo de los almacenes:

Según William H. Inmon, los almacenes de datos “Son un conjunto de datos orientados a un tema, integrados, de tiempo variante y no volátiles usados en la estrategia de toma de decisiones administrativas”. (Inmon, 2005)

Inmon concibe a los mercados de datos como subconjunto del almacén de datos, que son construidos para satisfacer las necesidades de análisis de cada departamento, siempre a partir del almacén.

En contradicción con la definición anterior, Ralph Kimball plantea:

“...los almacenes de datos son una copia de los datos de la transacción estructurados específicamente para la pregunta y el análisis.” (Kimball, y otros, 2002)

También fue Kimball quien señaló que un almacén de datos es la unión de todos los mercados de datos departamentales de una entidad.

A partir de estos dos conceptos se ha creado una gran polémica sobre cual enfoque es el mejor, tratando de definir la alternativa adecuada para la correcta creación de un almacén de datos. No hay bien o mal entre estas dos ideas, simplemente cada una tiene un punto de vista diferente sobre lo que debe prevalecer a la hora de diseñar un almacén de datos. Kimball representa la relación de los datos con el usuario final, la flexibilidad y la rápida explotación de la información. Por otra parte Inmon representa la pulcritud en el diseño y el respeto por una serie de normas que garanticen la exactitud de los datos, su integración y su coherencia.

Durante la presente investigación se sigue la teoría de Kimball, ya que son los procesos de negocio los que marcan la forma en la que se diseña el almacén de datos y propone emplear las estructuras dimensionales, lo que mejora el rendimiento a la hora de realizar consultas. Además permite organizar los datos de una forma más intuitiva y natural para los usuarios.

# Capítulo 1: Fundamentación teórica de los mercados de datos

## 1.2.1. Características de los almacenes de datos

Según Inmon, los almacenes de datos se caracterizan por 4 aspectos fundamentales:

Orientado al tema: Una primera característica del almacén de datos es que la información se clasifica en base a los aspectos que son de interés para la empresa. Siendo así, los datos tomados están en contraste con los clásicos procesos orientados a las aplicaciones. (Inmon, 2005)

No volátiles: La información es útil sólo cuando es estable, es por eso que los datos no serán modificados o cambiados de ninguna manera una vez ellos han sido introducidos en el almacén de datos, solamente podrán ser cargados, leídos o accedidos. (Inmon, 2005)

Tiempo variante: Como la información en el almacén de datos es solicitada en cualquier momento (es decir, no "ahora mismo"), los datos encontrados en el depósito se llaman de "tiempo variante". Los datos históricos son de poco uso en el procesamiento operacional. En cambio la información del almacén de datos, debe incluir los datos históricos para usarse en la identificación y evaluación de tendencias así como en comparaciones y previsiones. (Inmon, 2005)

Integrado: El aspecto más importante del ambiente de almacén de datos es que la información encontrada al interior está siempre integrada. La integración de datos se muestra de muchas maneras: en convenciones de nombres consistentes, en la medida uniforme de variables, en la codificación de estructuras consistentes, en atributos físicos de los datos consistentes, fuentes múltiples y otros. (Inmon, 2005)

## 1.2. Mercado de datos

Existen autores que se refieren al término almacén de datos y mercado de datos de manera similar. Esto se debe a que aunque no representan exactamente el mismo concepto, existe gran similitud en ambos casos.

# Capítulo 1: Fundamentación teórica de los mercados de datos

Los mercados de datos son almacenes de datos orientados a temas específicos o aplicaciones específicas y contienen datos de sólo una línea del negocio como puede ser ventas o marketing. (Hobbs, y otros, 2005)

Los mercados de datos son generalmente subconjuntos del almacén de datos, diseñados para satisfacer las necesidades específicas de grupos comunes de usuarios. (Torres, 2007)

Partiendo de las semejanzas que existen surge la interrogante: ¿Qué diferencia existe entonces entre un mercado de datos y un almacén de datos?

Su alcance. El mercado de datos está hecho para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización. Es el almacén natural para los datos departamentales. En cambio, el ámbito del almacén de datos es la organización en su conjunto. Es el almacén natural para los datos corporativos comunes. (Curto, 2007)

Para la realización del presente trabajo de diploma se adoptará el punto de vista de Hobbs, quien señala que los mercados de datos son almacenes de datos orientados a temas específicos. Por lo tanto, todos los fundamentos teóricos expuestos en lo adelante: conceptos asociados, características generales, metodologías de desarrollo y herramientas resultan aplicables de igual forma tanto a los mercados como a los almacenes.

## **1.2.2. Características de los mercados de datos**

1. Se centran en los requerimientos de los usuarios asociados a un departamento o área de negocios concreta.
2. Son sencillos a la hora de utilizarlos y comprender sus datos.
3. El diseño del mercado de datos, se realiza siguiendo una estructura consistente para las necesidades de los usuarios.
4. Contiene sólo el mínimo de información histórica.
5. Facilita el acceso de las herramientas de consulta y divide los datos para

controlar mejores accesos, debido a que frecuentemente hay grupos de usuarios que sólo acceden a un subconjunto concreto de datos. (Kimball, 2009)

## 1.3.Cubos Multidimensionales

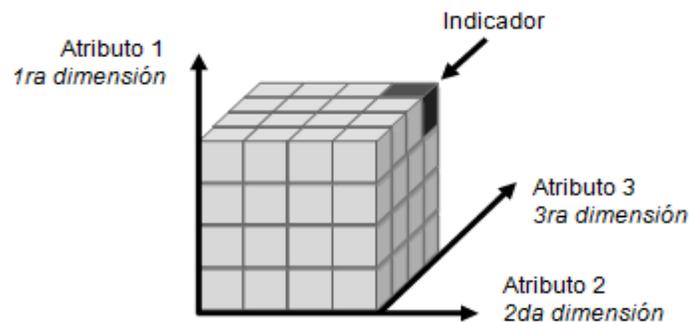
Un cubo multidimensional o hipercubo, representa los datos planos que se encuentran en las filas y columnas en una matriz de n dimensiones. Los objetos más importantes que se pueden incluir en un cubo multidimensional, son los siguientes:

Indicadores: Son las sumalizaciones que se efectúan sobre algún hecho, perteneciente a una tabla de hechos.

Atributos: Son campos o criterios de análisis, que pertenecen a las tablas de dimensiones.

Jerarquías: Representa una relación lógica entre dos o más atributos.

De esta forma, en un cubo multidimensional los atributos existen a lo largo de varios ejes o dimensiones, y la intersección de las mismas representa el valor que tomará el indicador que se está evaluando. (Bernabeu, 2009)



**Imagen 1: Cubo Multidimensional**

## 1.4.Procesamiento Analítico en Línea (OLAP)

Tecnología que permite manejar una serie de consultas de forma interactiva sobre estructuras multidimensionales (Cubos OLAP) cargadas previamente con los datos almacenados en las bases de datos corporativas tradicionales. Permiten realizar

# Capítulo 1: Fundamentación teórica de los mercados de datos

informes y obtener grandes cantidades de información a partir de lo que resultaría ser a modo rutinario una serie de complejas consultas sobre una base de datos de forma sencilla. Al estar los datos precompilados sobre una estructura intermedia, el tiempo de respuesta de las consultas es menor, posee una enorme potencia de cálculo y técnicas de indexación especializadas. (Zapata, 2009)

## **1.4.1. Procesamiento Analítico Relacional en Línea (ROLAP)**

Sistema OLAP que mantiene los datos almacenados en una base de datos relacional. Para esta implementación se realiza un cubo virtual o tablas en forma de estrella con lo que se consigue una mayor capacidad de almacenamiento sacrificando tiempo de respuesta. (Zapata, 2009)

## **1.4.2. Procesamiento Analítico Multidimensional en Línea (MOLAP)**

Sistema OLAP que posee los datos almacenados en una base de datos multidimensional. Esta implementación mejora los tiempos de acceso a los datos ya que están precalculados a costa de necesitar mayor espacio de almacenamiento, aunque algunos sistemas utilizan la compresión. (Zapata, 2009)

## **1.4.3. Procesamiento Analítico Híbrido en Línea (HOLAP)**

Combinación de los dos sistemas anteriores donde los datos se almacenan repartidos en implementaciones MOLAP y ROLAP. Esta combinación permite obtener ventajas de ambas implementaciones según donde se almacene el dato y las operaciones que se vayan a realizar sobre él. (Zapata, 2009)

Posee dos tipos de particionamiento:

**Particionamiento Vertical:** Almacena las agregaciones como un MOLAP para mejorar la velocidad de las consultas, y los datos se detallan en ROLAP para optimizar el tiempo en que se procesa el cubo. (Ortiz, 2009)

**Particionamiento Horizontal:** En este modo HOLAP almacena una sección de los datos, normalmente los más recientes (por ejemplo particionando por la dimensión

# Capítulo 1: Fundamentación teórica de los mercados de datos

tiempo) en modo MOLAP para mejorar la velocidad de las consultas, y los datos más antiguos en ROLAP. (Ortiz, 2009)

Modos de almacenamiento	Ventajas	Desventajas
MOLAP	Mejoras en cuanto al tiempo de respuesta de las consultas.	Ocupa más espacio de almacenamiento en disco.
ROLAP	Ahorro de espacio de almacenamiento en disco.	El tiempo de respuesta de las consultas es mayor.
HOLAP	Mejoras en cuanto al tiempo de respuesta de las consultas solamente para información resumida.	Más grandes volúmenes de datos en la base de datos relacional

**Tabla 1: Ventajas y desventajas de los modos de almacenamiento.**

Teniendo en cuenta las características de cada modo de almacenamiento y las ventajas y desventajas que supone su uso en cada caso, se decide implementar los cubos usando el modo de almacenamiento ROLAP. Este tipo de arquitectura se implementa mejor sobre bases de datos relacionales, permite ahorrar espacio de almacenamiento y soporta análisis OLAP contra grandes volúmenes de datos.

## 1.5. Metadatos

No son más que datos relativos a datos. Estos permiten mantener información acerca de la procedencia de los datos, su fiabilidad, forma de cálculo y otros. (Humantumba, 2007)

Los metadatos brindan ayuda para la comprensión de la información del almacén por los usuarios finales. Indican el significado de cada valor que toman los datos en el lenguaje del negocio, además de ayudar en la construcción de consultas, informes y en el análisis en general.

## 1.6. Modelo dimensional

Los almacenes de datos necesitan manejar grandes cantidades de información con rapidez, por lo que se utiliza el modelo dimensional que tiene como ventaja fundamental la optimización del tiempo de respuesta de las consultas gracias a la estructura que presenta.

El modelo dimensional divide el mundo de los datos en dos grandes tipos: las medidas y las descripciones del entorno de estas medidas. Las medidas, que generalmente son numéricas, se almacenan en las tablas de hechos y las descripciones de los entornos que son textuales se almacenan en las tablas de dimensiones. (Ponniah, 2001)

Existen tres alternativas de modelación dimensional caracterizadas de acuerdo a su estructura física:

### 1.6.1. Esquema en estrella

El esquema estrella o *StarScheme* en inglés, es la arquitectura más simple de un almacén de datos. El esquema estrella consta de una tabla de variables o tabla de hechos central y de varias tablas de dimensiones relacionadas a esta a través de sus claves respectivas. Este modelo debe estar totalmente desnormalizado, es decir, no puede presentarse en tercera forma normal. (Bernabeu, 2009)

### 1.6.2. Esquema copo de nieve

El esquema copo de nieve o *SnowflakeScheme* en inglés representa una extensión del modelo en estrella cuando las tablas se organizan en jerarquías de dimensiones. Existe una tabla de hechos central que está relacionada con una o más tablas de dimensiones, las cuales a su vez pueden estar relacionadas o no con una o más tablas de dimensiones. (Bernabeu, 2009)

Este tipo de esquema es el más cercano a un modelo entidad relación, ya que en este caso en particular las dimensiones se encuentran normalizadas.

### 1.6.3. Esquema constelación

El esquema constelación o *StarflakeScheme* está compuesto por una serie de esquemas en estrellas. Está formado por una tabla de hechos principal y por una o más tablas de hechos auxiliares, las cuales pueden ser sumalizaciones de la principal. Estas tablas están relacionadas con sus respectivas tablas de dimensiones. (Bernabeu, 2009)

Este tipo de esquema es muy flexible pero cuando el número de relaciones es elevado se hace difícil de implementar y mantener.

Para la selección del esquema dimensional que se va a utilizar para la construcción del mercado de datos, se debe tener en cuenta las ventajas y desventajas que ofrece cada alternativa. El esquema constelación resulta mucho más complejo de implementar que las otras dos alternativas, por tener varias tablas de hechos relacionadas con dimensiones en común. Utilizar el esquema copo de nieve ahorraría espacio de almacenamiento y eliminaría redundancias, pero el rendimiento se vería afectado por la necesidad de realizar *JOINS* adicionales para acceder a los datos de una dimensión, mientras que el esquema en estrella supone la necesidad de mayor espacio de almacenamiento pero a su vez mayor velocidad en la ejecución de las consultas.

Como la prioridad establecida es que el sistema responda en un tiempo óptimo y teniendo en cuenta las características del negocio, la mejor opción sería el esquema constelación que aunque es el más complejo resulta una alternativa viable dado que el número de relaciones no es elevado y es necesario relacionar las tablas de hechos con varias dimensiones en común.

### 1.7. Conceptos asociados al modelo dimensional

Para realizar un correcto modelado dimensional con vistas a obtener la estructura física de un mercado de datos, es necesario conocer los conceptos y componentes básicos de dicha estructura:

## **1.7.1.Tablas de hechos**

Las tablas de hechos son las tablas primarias en el modelo dimensional, ya que almacenan los valores del negocio, cada tabla representa una relación de muchos a muchos, y contiene dos o más llaves extranjeras que se enlazan con sus respectivas tablas de dimensiones. La principal condición que debe cumplir las tablas de hechos es que el hecho debe almacenarse de tal forma que su valor sea numérico y a su vez sea aditivo para así poder realizar cálculos sobre él, ya sea por ciento, sumas o igualdades. (msdn, 2011)

## **1.7.2.Tablas de dimensiones**

Las tablas de dimensiones son las compañeras integrales de las tablas de hechos, ellas contienen la descripción textual del negocio. En el modelo dimensional, las tablas de dimensiones 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 relaciones de las consultas, agrupaciones y las etiquetas de los reportes. (msdn, 2011)

## **1.7.3.Medidas**

La medida o hecho es un dato numérico que representa una actividad específica de un negocio, mientras que una dimensión representa una perspectiva de los datos. Una medida contiene una propiedad numérica y una fórmula. (Herrera, 2007)

## **1.7.4.Granularidad**

Una característica importante que define a una tabla de hechos es el nivel de granularidad de los datos que en ella se almacenan, entendiéndose por “granularidad” el nivel de detalle de dichos datos, es decir, la granularidad de la tabla de hechos representa el nivel más atómico por el cual se definen los datos. (Guijarro Riera, y otros, 2010)

La granularidad establecida influye de forma directa tanto en la cardinalidad de la tabla de hechos como en la cardinalidad de las dimensiones. Mientras mayor sea el nivel de granularidad, mayor será el número de filas de la tabla de hechos.

## **1.7.5. Jerarquía**

Es un conjunto de atributos descriptivos que se rigen por un orden preestablecido. Una jerarquía implica una organización de niveles dentro de una dimensión, con cada nivel representando el total agregado de los datos del nivel inferior. Además, definen como los datos son sumados desde los niveles más bajos hasta los más altos. Una jerarquía puede contener todos los valores que se encuentran en una dimensión, pero no es necesario. (Herrera, 2007)

## **2. Tendencias Actuales**

Desde su surgimiento en los años 80 hasta la actualidad se han venido perfeccionando las metodologías de desarrollo de los almacenes de datos. Ha tomado más de 20 años obtener la madurez, estabilidad y aceptación de dicha tecnología, que ha llegado a situarse en el centro de atención de las grandes instituciones, dado que proporciona un entorno en el que las organizaciones pueden hacer un mejor uso de aquella información que antes no podían explotar al máximo.

Los almacenes de datos han resultado aplicables a diversas ramas tanto de la economía como de la sociedad. Desde el punto de vista médico, tienen gran utilidad para la realización de estudios de los patrones de comportamiento de pacientes con diferentes síntomas y como resultado se obtienen diagnósticos más rápidos y precisos. En el sector del transporte los almacenes de datos contribuyen en el análisis de los datos históricos referentes al estudio sobre transportaciones, clientes, monitoreo de ganancias y futuras proyecciones.

En los estudios de mercado orientados hacia el comercio electrónico y el marketing es donde más incidencia ha tenido esta novedosa forma de utilizar los datos. Juegan un papel determinante en la monitorización de la información que caracteriza a los clientes para la realización de estudios que permitan a las empresas mantenerse en la preferencia y evite a su vez la pérdida de clientes potenciales.

## Capítulo 1: Fundamentación teórica de los mercados de datos

Las organizaciones gubernamentales han encontrado en los almacenes y mercados de datos un gran apoyo, que ofrece entre otras ventajas más seguridad al verse en la posibilidad de tomar decisiones mejor informadas. Este tipo de solución se ha convertido en una necesidad primaria para los altos directivos, que cargan con gran responsabilidad y no disponen del tiempo suficiente para dedicarlo al análisis de la información.

A nivel mundial existen gran cantidad de organismos que hacen uso de los mercados de datos en aras de alcanzar el mejoramiento empresarial. Existen evidencias del desarrollo de estas tecnologías enfocadas al sector de la agricultura, como se puede observar en los ejemplos estudiados:

1. SIAP<sup>1</sup> tiene a su disposición un mercado de datos, denominado Data Mart Pecuario. Dicho mercado contiene información de la producción agrícola nacional (para los años comprendidos entre 1980 y 2010) por entidad federativa, con las variables de superficie sembrada, cosechada, valor de la producción de cultivos anuales y perennes, por ciclo, modalidad de riego y temporal. Se puede visualizar la información en diferentes periodos de tiempo, hacer filtros dinámicos, creación de consultas (agregación de columnas calculadas, agregación de cortes por dimensión, operaciones en una columna de indicador, tendencias, máximos y mínimos).

Desde el año 2010 se encuentra disponible en la web (<http://www.siap.gob.mx>) un portal que permite acceder a los servicios antes mencionados. Según los resultados de las encuestas realizadas y las opiniones de los usuarios se evidencia la gran aceptación que tuvo el producto.

2. CECAV<sup>2</sup> dispone de una solución de inteligencia de negocios que incluye un mercado de datos y herramientas de exploración y análisis OLAP

---

1 Servicio de Información Agroalimentaria y Pesquera de México. Tiene como misión proveer a los productores agropecuarios, pesqueros y agentes económicos que participan en las cadenas agroalimentarias, de información confiable y oportuna para la toma de sus decisiones

2 Centro de Calidad Avícola y alimentación animal de la Comunidad Valenciana (España). Su propósito es garantizar una serie de productos de calidad que generen riqueza para la comunidad.

## Capítulo 1: Fundamentación teórica de los mercados de datos

desarrolladas por la empresa *Lite Internet Solutions* “Litebi”<sup>3</sup> S.L en el periodo 2009-2010. Ofrece múltiples métodos de análisis de la información mediante cubos con tecnología OLAP y desarrollo de informes simples en poco tiempo, todo esto sin necesidad de amplios conocimientos informáticos por parte del usuario.

La existencia de estas posibilidades fue muy útil para los usuarios operativos que tienen que tratar con información de forma inmediata y tienen necesidad de analizarla en profundidad para la toma de decisiones del negocio. Los usuarios destacaron sobre todo la rapidez y la calidad de los informes generados. La implantación de la solución ha servido para observar y corroborar de primera mano la gran utilidad de estos sistemas en las organizaciones.

3. Empresas PMC<sup>4</sup>, es un grupo de empresas del sector agropecuario-industrial involucradas en todo el ciclo del negocio azucarero, desde la recolección y cosecha de la caña hasta la producción y distribución de azúcar refinada y sus derivados. Una de las necesidades fundamentales para PMC era la incorporación de herramientas analíticas que permitiesen transformar los datos del negocio en información que fuese útil para la toma de decisiones. Cuenta con 2 mercados de datos, denominados el Data Mart del Negocio Agrícola y el Data Mart del Negocio de Fábrica, un cubo para agrícola, más de 300 indicadores y una familia de 38 reportes.

En Cuba se ha evidenciado el uso de los mercados de datos en diversas esferas, aunque todavía no se han alcanzado los niveles de perfeccionamiento y apropiación que tienen otros países, se han realizado notables avances en el desarrollo y utilización de los almacenes de datos. Organismos estatales como la Oficina Nacional de Estadísticas e Información (ONEI), en su momento presentó dificultades con el análisis de los datos recogidos a lo largo y ancho de todo el país, obstruyendo el flujo de información necesario para una correcta toma de

---

<sup>3</sup> Plataforma de inteligencia de negocios que utiliza el modelo de distribución de Software como Servicio (*Software as a Service*, SaaS)

<sup>4</sup> Siglas de Plan de Mejora Competitiva

# Capítulo 1: Fundamentación teórica de los mercados de datos

decisiones. En la actualidad cuenta con un almacén de datos que mejora la disponibilidad y calidad de análisis requerido así como la disminución de errores estadísticos.

Como subconjunto del almacén y enfocado en el sector de la agricultura, se encuentra el mercado de datos indicadores relacionados con el Comercio Agropecuario, que forma parte del almacén de datos para la ONEI, almacena las ventas de productos en el mercado agropecuario con el fin de conocer los niveles del consumo de productos agropecuarios de la población.

Incluido en el Sistema de Información del Gobierno desarrollado por la UCI, el mercado de datos para el área de Agricultura, Ganadería y Silvicultura (en la capital) persigue como objetivo fundamental el análisis de los procesos de siembra y las actividades ganaderas con el objetivo de trazar estrategias que permitan incrementar los niveles productivos.

Cada uno de estos mercados fue desarrollado con el propósito de resolver una problemática similar a la que da origen a la presente investigación. Luego de haber realizado un estudio de las soluciones antes mencionadas se puede afirmar que no existe aún una solución que se ajuste a las necesidades del cliente. Los ejemplos antes mencionados no abarcan los indicadores requeridos ni las perspectivas de análisis necesarias. Tratar de adaptarlos a las exigencias del negocio supondría emplear mayor cantidad de tiempo y esfuerzo en su modificación que en la implementación de una nueva solución desde cero.

### **3. Metodologías de Desarrollo**

Existen dos tendencias bien marcadas que se han generalizado a nivel mundial y sirven de guía sobre este tema: La Metodología de Kimball y la Metodología de Immon, en representación a sus creadores Ralph Kimball y William H. Inmon respectivamente.

Para Kimball el almacén de datos es el conjunto de todos los mercados de datos existentes en la empresa, aunque normalmente se almacenen de forma separada.

## Capítulo 1: Fundamentación teórica de los mercados de datos

Bajo esta concepción la información está almacenada en un modelo dimensional y por tanto está lista para ser consultada. Esta aproximación se conoce como “*bottom-up*”, una metodología ascendente a la hora de diseñar un almacén de datos. (Zapata, 2009)

Para Inmon el almacén de datos es una parte del sistema de la inteligencia de negocios. Los mercados de datos se crean después de diseñar el almacén de datos y obtienen la información de éste, no siendo posible la consulta directa de información (se dice que la información no está almacenada de forma dimensional). Esta aproximación es conocida como “*top-down*”. (Zapata, 2009)

En determinadas situaciones una metodología puede ofrecer ventajas sobre otra, sin embargo es posible que una combinación permita una mejor respuesta en dependencia de la problemática que se presente. Metodologías como DM2, Hefesto y la metodología para el Diseño Conceptual Almacenes de Datos presentada por Leopoldo Zenaido Zepeda Sánchez en su tesis de doctorado, no siguen estrictamente uno de los enfoques antes mencionados, sino que seleccionan lo mejor de cada una para definir una nueva metodología que se ajuste a un grupo de intereses específicos:

DM2: Se basa en las necesidades de información a nivel gerencial, donde la información debe ser encarada como patrimonio de la empresa, accesible a quien la necesite. Por la propia naturaleza del ambiente, el modelo cumple con su objetivo (atender las necesidades de información del nivel gerencial y ejecutivo de una empresa), esta metodología se asemeja a la forma *top-down* que propone Inmon, y acorta en función razonable el tiempo entre el inicio del análisis y la implantación. Esta rapidez no solo es buena para el cliente sino que también es exigida y necesaria por el propio ambiente que lo rodea. (González Hernández, 2011)

Hefesto: Entre sus principales directrices plantea que la construcción e implementación de un almacén de datos puede adaptarse muy bien a cualquier ciclo de vida de desarrollo de software, con la salvedad de que para algunas fases

# Capítulo 1: Fundamentación teórica de los mercados de datos

en particular, las acciones que se han de realizar serán muy diferentes. Lo que se debe tener en cuenta, es no entrar en la utilización de metodologías que requieran fases extensas de reunión de requerimientos y análisis, fases de desarrollo monolítico que conlleve demasiado tiempo y fases de despliegue muy largas. Lo que se busca, es entregar una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas del almacén de datos y motivar a los usuarios. (González Hernández, 2011)

Metodología para el Diseño Conceptual de Almacenes de Datos presentada en la Tesis de Doctorado de Leopoldo Zenaido Zepeda Sánchez: Aporta como aspecto novedoso con respecto a las anteriores la incorporación de una serie de transformaciones para llevar un diagrama relacional a uno dimensional y así obtener las estructuras candidatas que conformarán el repositorio de datos. (González Hernández, 2011)

## **3.1. Metodología utilizada**

La metodología que se utilizará en el desarrollo del mercado de datos es la propuesta desarrollada para la Línea de Soluciones de Almacenes de Datos e Inteligencia de Negocios por la Universidad de las Ciencias Informáticas.

Toma como base la metodología de Kimball, con el objetivo de aprovechar las ventajas que ofrece:

1. Crea los conceptos de hechos y dimensiones, lo que indudablemente es muy eficaz en el proceso de la toma de decisiones y proporciona mayor agilidad en el proceso de desarrollo.
2. Propone ir construyendo el almacén de datos a través de la construcción de los mercados de datos departamentales, lo que constituye una estrategia buena y coincide con la división lógica de las empresas, entidades, organismos y otros.
3. Existe abundante documentación sobre la misma, 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. (González Hernández, 2011)

Para dar complemento a la metodología de Kimball y mitigar los inconvenientes que presenta aplicada al entorno productivo de la facultad se fusiona con la metodología planteada por Leopoldo Zenaido Zepeda Sánchez en su tesis de doctorado, para orientar el trabajo hacia casos de uso y lograr ajustarse mejor a las normas y patrones de la universidad.

## 4.Descripción de las herramientas utilizadas

### 4.1. Herramienta de modelado

**Visual Paradigm 6.4 (*Enterprise Edition*<sup>5</sup>):** Es una herramienta que utiliza UML<sup>6</sup> como lenguaje de modelado. Soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. Su uso posibilita una construcción de aplicaciones rápida, con buena calidad, y a un menor coste. Proporciona abundante documentación y tutoriales. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. Esta herramienta es colaborativa, soporta múltiples usuarios trabajando sobre el mismo proyecto; genera la documentación automáticamente en varios formatos como web o PDF<sup>7</sup>, y permite el control de versiones. Cabe destacar su robustez, usabilidad y portabilidad. Es un producto de calidad que soporta aplicaciones Web, genera código para Java y es fácil de instalar y actualizar. Acelera el desarrollo de aplicaciones, sirviendo de puente visual entre arquitectos, analistas y diseñadores de sistemas de información, haciendo el trabajo más fácil y dinámico. Es multiplataforma. Esta herramienta posee tanto licencia comercial como gratuita. (Visual Paradigm, 2012)

---

5 Edición Comunitaria

6 Siglas de Lenguaje Unificado de Modelado. Es un lenguaje gráfico para visualizar, especificar y documentar cada una de las partes que comprende el desarrollo de software.

7 (Formato de Documento Portátil), es un formato de almacenamiento de documentos

## 4.2. Sistema gestor de base de datos

**PostgreSQL 9.1:** Es un Sistema Gestor de Bases de Datos (SGBD) relacional orientado a objetos y libre, está considerado como la base de datos de código abierto más avanzada del mundo, se caracteriza por ser un sistema estable, de alto rendimiento, gran flexibilidad. Tienen la función de mantener la integridad y consistencia en la base de datos. Soporta operadores, funciones, métodos de acceso y tipos de datos definidos por el usuario. Incorpora funciones de diversa índole: manejo de fechas, geométricas, orientadas a operaciones con redes. Es una herramienta de software libre. (ArPUG, 2009)

## 4.3. Cliente para la Administración de bases de datos

**PgAdmin III (1.12):** Es una herramienta de código abierto muy avanzada y se utiliza para administrar y desarrollar bases de datos en PostgreSQL. Se diseña para responder a las necesidades de la mayoría de los usuarios, desde escribir simples consultas SQL, hasta desarrollar bases de datos complejas. La interfaz gráfica soporta todas las características de PostgreSQL y facilita su administración. Está disponible en más de una docena de idiomas y para varios sistemas operativos. (ArPUG, 2009)

## 4.4. Herramienta de integración

**Pentaho Data Integration 3.2:** Conocido también como *Kettle*, es el componente de Pentaho responsable de la extracción, transformación y carga de procesos. Este se encarga de abrir, limpiar e integrar la información disponible en diferentes fuentes de datos y ponerla en manos del usuario. Provee una consistencia, una sola versión de todos los recursos de información, que es uno de los más grandes desafíos para las organizaciones. Posee un poderoso proceso de Extracción, Transformación y Carga (ETL). El uso de *Kettle* evita grandes cargas de trabajo manual frecuentemente difícil de mantener y de desplegar. Es una de las más antiguas herramientas ETL de código abierto y sin costos de licencia, cuenta con

# Capítulo 1: Fundamentación teórica de los mercados de datos

una gran comunidad de usuarios y su interfaz gráfica permite un aumento de la productividad.

Incluye cuatro herramientas:

**Spoon:** utilizado para diseñar transformaciones ETL usando el entorno gráfico.

**Pan:** se usa para ejecutar transformaciones diseñadas con *spoon*.

**Chef:** específicamente para crear trabajos.

**Kitchen:** específicamente para ejecutar trabajos. (Gravitar , 2012)

## 4.5. Herramientas de inteligencia del negocio

**Pentaho Metadata Editor 3.7:** Esta herramienta de Pentaho permite definir modelos de negocio integrando datos de distintas fuentes para posteriormente ser utilizado en reportería de auto servicio (ad-hoc), tableros de control, gráficos y tablas. Pentaho Metadata Editor (PME) es una herramienta que permite crear dominios de metadatos y modelos. El objetivo es mapear la estructura física de la base de datos a un modelo lógico de negocio. ( Comunidad de Pentaho, 2010)

**Pentaho BI Server 3.2:** Provee el soporte y la infraestructura necesaria para crear soluciones de inteligencia empresarial a problemas de negocios. Brinda diversas funcionalidades como autenticación, registro, auditoría, servicios web y motor de reglas, también incluye un motor de solución que integra reportes, análisis, tableros de comandos y componentes de minería de datos. Funciona como un sistema basado en administración web de informes y está diseñado para utilizarse fácilmente en cualquier proceso de negocio. (Productive, 2007)

**Pentaho Schema Workbench 3.2:** Es una herramienta para el desarrollo y prueba de esquemas multidimensionales. La definición del XML no es extremadamente compleja, pero en la práctica resulta engorroso recordar cada uno de los elementos junto a sus atributos y sub-elementos, tal y como se encuentran en el almacén. Esta aplicación permite la ejecución de las consultas MDX4. (Sinnexus, 2011)

# Capítulo 1: Fundamentación teórica de los mercados de datos

**Pentaho Report Designer 3.7:** Esta aplicación de Pentaho proporciona todas las herramientas necesarias para que estos sean diseñados y publicados para ser consultados fácilmente vía web por quienes lo requieran. Dentro de sus características se destacan:

1. Provee plantillas de alto nivel, predefinidas para agilizar el proceso de diseño del reporte sin mermar la consistencia de este mismo.
2. Permite crear parámetros para que el usuario final personalice la información que quiere desplegar en el reporte
3. Permite que los reportes sean publicados al BI Server desde la misma interfaz para que los usuarios seleccionados puedan ver los reportes. (Comunidad de Pentaho, 2010)

## 4.6. Servidor web

**Apache Tomcat 5.5:** Es una herramienta que funciona como un contenedor de *servlets* desarrollado en el proyecto Jakarta en el *Apache Software Foundation*<sup>8</sup>. Es un servidor web y de aplicaciones que gestiona solicitudes y respuestas Http (incluye el servidor Apache) gracias a sus conectores Http. Es un servidor de aplicaciones o contenedor de Servlets/JSP. Permite escribir y desplegar aplicaciones web complejas de forma sencilla, proporcionando el soporte para características de desarrollo que de otra forma tendrían que ser desplegadas manualmente. (Apache Software Foundation, 2012)

## Conclusiones del capítulo

Después de haber realizado un análisis a profundidad de la problemática existente en la Delegación Provincial de la Agricultura de Artemisa se puede concluir que la alternativa más eficiente para eliminar las actuales deficiencias en cuanto al análisis

---

<sup>8</sup> Es una organización no lucrativa creada para dar soporte a los proyectos de software bajo la denominación Apache, incluyendo el popular servidor HTTP Apache.

## Capítulo 1: Fundamentación teórica de los mercados de datos

de la información es la creación de un mercado de datos que permita el almacenamiento, procesamiento, obtención y visualización de toda la información histórica que se posee.

Al concluir el capítulo quedan establecidos los fundamentos teóricos y seleccionada la metodología que va a servir de guía durante el desarrollo del producto, orientando el trabajo hacia el logro de los resultados esperados. Se definieron las herramientas a utilizar, selección que se basó principalmente en la necesidad de hacer uso de alternativas de software libre y multiplataforma que brinde buenas prestaciones.

### **Capítulo 2. Análisis y Diseño del mercado de datos de la Delegación Provincial de la Agricultura de Artemisa.**

A lo largo del presente capítulo se abordaron una serie de elementos que posibilitan un mayor entendimiento del negocio. Se muestra de manera general los resultados arrojados durante la fase de levantamiento de requisitos, arquitectura y diseño: los temas de análisis, las reglas del negocio, los requisitos de información, funcionales y no funcionales, diagrama de casos de uso y descripción. Se muestra el modelo de datos de la solución, detallando las dimensiones, hechos y medidas que lo componen. De igual forma se presenta la arquitectura de la información definida para el mercado.

#### **2.1. Estudio preliminar del negocio**

La AP de Artemisa es el organismo gubernamental encargado de llevar el control de todos los procesos y actividades de la provincia, tanto económicos como de índole social. Necesita por lo tanto registrar el comportamiento de numerosos indicadores para las diferentes esferas, que representan en su totalidad grandes cantidades de datos. Cada una de las 32 direcciones que conforman su estructura organizativa tiene a su cargo la recepción y procesamiento de los datos de un sector de la economía o tema específico de interés para el gobierno.

La Delegación Provincial de la Agricultura tiene como responsabilidad la captación y el análisis diario de los datos relacionados con el comportamiento de las actividades agropecuarias de las diferentes empresas productoras. La información que se procesa está dividida en dos categorías generales: ganadería y cultivos varios.

Los valores de los indicadores se obtienen a través de partes diarios emitidos por cada una de las empresas productoras de la provincia. Para su almacenamiento la delegación cuenta con varios modelos que permiten recoger los datos en cada caso: Los partes de matadero, leche, cumplimiento de la demanda, siembra y preparación de tierras para la ganadería, existencia en nave, alimentación animal,

siembra de cultivos varios, siembra de la campaña de papa, siembra de la campaña del frijol y siembra de la campaña del maíz.

A partir de las estadísticas recopiladas mediante los modelos se realizan los cortes diarios y acumulados a nivel provincial. Luego son enviados al delegado provincial que se encarga de emitir el resultado del análisis de los datos a los directivos de la AP.

### **2.2. Temas de Análisis**

Uno de los pasos más importantes en el proceso de desarrollo de un mercado de datos es la definición de los Temas de Análisis (TA). A través de estos se realiza la agrupación de las necesidades de los usuarios en áreas de información, cada una enfocada a un objetivo en específico, que sirve al propósito general que persigue el mercado de datos. De esta forma se garantiza un alto nivel organizativo que orienta la realización de las actividades hacia el cumplimiento de las metas establecidas.

La presente propuesta de solución está orientada hacia las principales categorías de la información con la que se trabaja en la Delegación Provincial de la Agricultura, como son: ganadería y cultivos varios, obteniendo como resultado los temas de análisis:

TA 1. Análisis de los indicadores específicos de la ganadería.

TA 2. Análisis de los indicadores específicos de cultivos varios.

### **2.3. Reglas del Negocio**

Las Reglas del Negocio (RN) describen las normas, operaciones, definiciones y restricciones presentes en una organización. Son de gran importancia para el logro de los objetivos en la misma. Se identificaron 46 reglas del negocio que se detallan en el artefacto Reglas del Negocio (FRA-Government-0116\_RNeg), disponible en el expediente de proyecto del mercado de datos de la Delegación Provincial de la Agricultura, entre las cuales se encuentran:

RN 1. Los campos de llenado opcional que se encuentren vacíos se rellenan con el valor cero.

RN 2. Los campos de llenado obligatorio que se encuentren vacíos se rellenan con el valor cero.

RN 3. Porcentaje de cumplimiento: Porcentaje que representa el valor real obtenido con respecto al plan que ha sido establecido. El resultado se redondea por exceso dos lugares después de la coma.

$\%_{\text{cumplimiento}} = \text{real} / \text{plan} * 100.$

### **2.4. Especificación de requerimientos**

Después de realizar un estudio de los principales indicadores y temas de análisis que han sido establecidos para el mercado de datos, se procede a identificar los requerimientos necesarios para satisfacer las necesidades del cliente. En el caso particular de las soluciones de almacenes de datos, los requerimientos se dividen en 3 categorías fundamentales:

#### **2.5.1. Requerimientos Funcionales**

Los Requerimientos Funcionales (RF) son capacidades o condiciones que el sistema debe cumplir para dar respuesta a las necesidades del usuario final. La solución debe permitir:

RF 1. Autenticar usuario.

RF 2. Adicionar usuario.

RF 3. Modificar usuario.

RF 4. Eliminar usuario.

RF 5. Adicionar rol.

RF 6. Eliminar rol.

RF 7. Modificar rol.

RF 8. Crear reporte Ad-hoc.

RF 9. Adicionar reporte.

RF 10. Eliminar reporte.

RF 11. Extraer datos de indicadores de la Delegación Provincial de la Agricultura.

RF 12. Transformar y Cargar datos de indicadores de la Delegación Provincial de la Agricultura.

RF 13. Perfilar datos de indicadores de la Delegación Provincial de la Agricultura.

### **2.5.2.Requerimientos No Funcionales**

Los Requerimientos No Funcionales (RNF) son propiedades o cualidades con las que debe contar el producto y se encuentran directamente vinculados con los requerimientos funcionales. Determinan cómo debe comportarse el producto, haciéndolo atractivo, usable, rápido y confiable. Además incluyen el nivel de conocimiento que debe tener el usuario final para garantizar la correcta utilización del producto. Para la implementación de la solución se proponen 26 requisitos que se detallan en el artefacto Especificación de Requisitos de Software (FRA-Government- 0113\_ERS), disponible en el expediente de proyecto del mercado de datos de la Delegación Provincial de la Agricultura, entre los cuales se encuentran los siguientes:

RI 1. Los reportes se deben poder diseñar de manera sencilla y ágil.

RI 2. Agilizar el acceso a los reportes mediante la distribución de la información por áreas de análisis.

### **2.5.3.Requerimientos de Información**

Los Requerimientos de Información (RI) son las principales funcionalidades que el sistema debe tener disponible a la hora de realizar el análisis sobre los datos. Se definen a partir de la comparación entre las necesidades de información y las reglas del negocio con los elementos disponibles en las fuentes. Constituyen una entrada fundamental para el proceso de inteligencia del negocio y para futuros reportes. En la presente investigación fueron identificados 110 requerimientos que se detallan en

el artefacto Especificación de Requisitos de Software (FRA-Government-0113\_ERS), disponible en el expediente de proyecto del mercado de datos de la Delegación Provincial de la Agricultura, entre los cuales se encuentran los siguientes:

RI 1. Obtener el nombre de la empresa, cantidad de animales a suplementar, cantidad suplementados, % de cumplimiento del plan de suplemento, plan per cápita, real per cápita, % de cumplimiento del plan per cápita, diario/mensual/semestral/trimestral/ anual por empresas.

RI 2. Obtener el plan, real, % de cumplimiento con respecto al plan y cantidad de tierras lista de la preparación de tierras para la ganadería diario/mensual/semestral/trimestral/anual por empresa y tipo de cultivo.

### **2.5. Casos de Uso del Sistema**

Los Casos de Uso del Sistema (CUS) son elementos que forman parte del análisis dado que ayudan a describir qué es lo que el sistema debe hacer. Representan las acciones del mismo desde el punto de vista del usuario, en pocas palabras, describen el uso del sistema y cómo este interactúa con el usuario. Se dividen en dos clasificaciones:

#### **2.5.1. Casos de uso funcionales**

CUS 1. Autenticar usuario.

CUS 2. Administrar usuarios.

CUS 3. Administrar roles.

CUS 4. Administrar reportes.

CUS 5. Crear reporte Ad-hoc.

CUS 6. Extraer datos de indicadores de la Delegación Provincial de la Agricultura.

CUS 7. Transformar y cargar datos los indicadores de la Delegación Provincial de la Agricultura.

CUS 8. Perfilar datos de indicadores de la Delegación Provincial de la Agricultura.

### **2.5.2.Casos de uso informativos**

CUI 1. Visualizar indicadores de la alimentación para la ganadería.

CUI 2. Visualizar indicadores de la siembra para la ganadería

CUI 3. Visualizar indicadores de la preparación de tierras para la ganadería.

CUI 4. Visualizar indicadores de producción de leche.

CUI 5. Visualizar indicadores de producción de carne.

CUI 6. Visualizar indicadores del Balance de existencia en nave de cultivos varios.

CUI 7. Visualizar indicadores de la siembra de cultivos varios.

CUI 8. Visualizar indicadores de la preparación de tierras para cultivos varios.

CUI 9. Visualizar indicadores de la demanda de cultivos varios.

### **2.6. Diagrama de casos de uso del sistema**

El diagrama de casos de uso del sistema resulta de gran utilidad para obtener una visión general de cómo los diferentes actores interactúan con los casos de uso. Se componen de actores, casos de uso y las relaciones entre ellos. Se muestra dividido por temas de análisis para mejorar la visibilidad y así facilitar su comprensión.

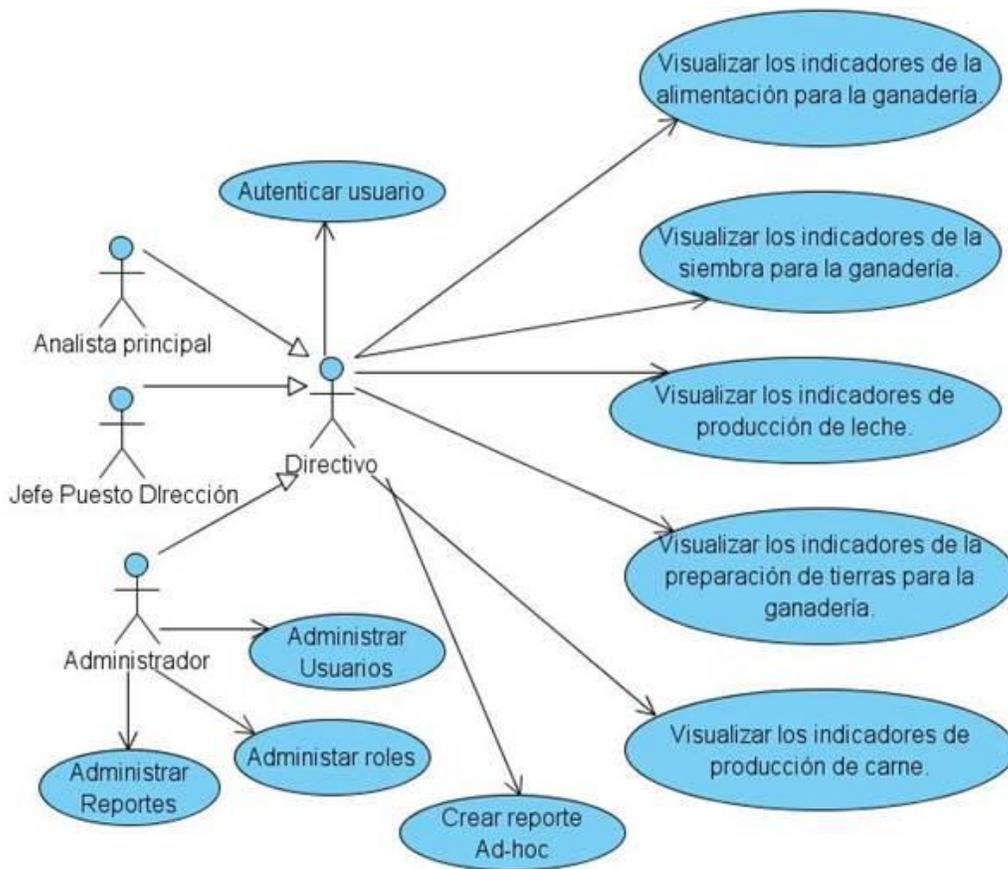
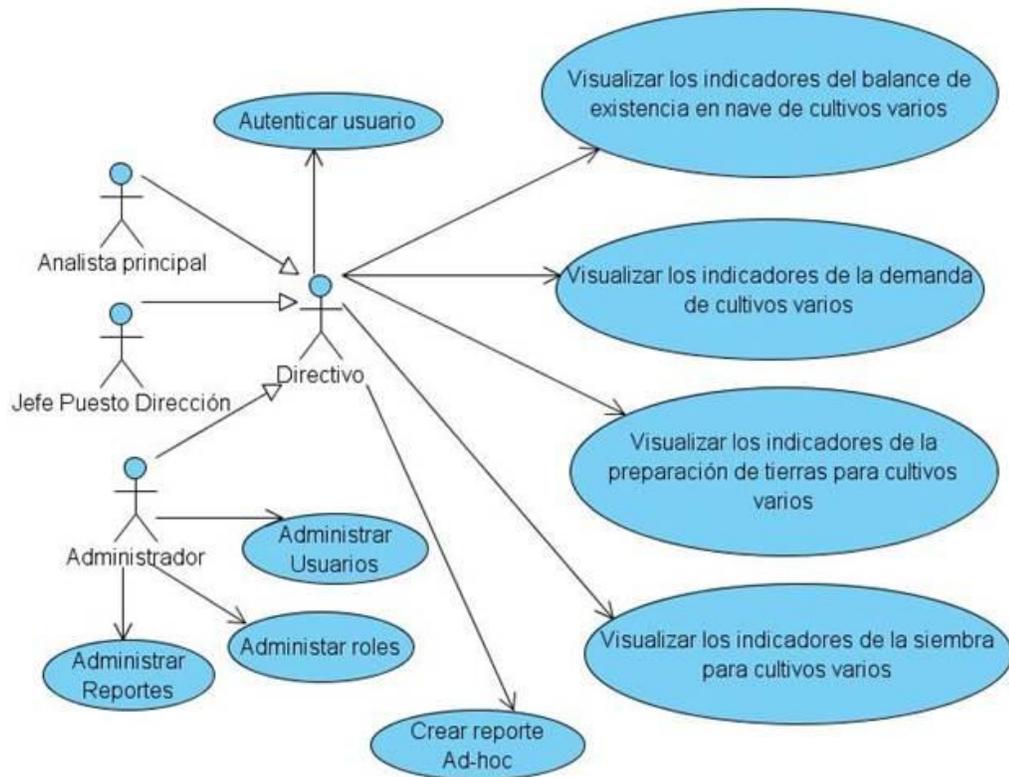


Imagen 2: Diagrama de CU correspondiente al tema de análisis Ganadería



Imagen 3: Diagrama de CU correspondiente al proceso de ETL



**Imagen 4: Diagrama de CU correspondiente al tema de análisis Cultivos Varios**

### 2.7. Especificación de casos de uso del sistema

La especificación de casos de uso del sistema es la parte más importante de los casos de uso, debido a que describe el rol que desempeña un actor en su interacción con el sistema y es de gran utilidad para el entendimiento de la forma en que se realiza paso a paso. Para una mayor comprensión de los casos de uso presentes en las Imágenes 2, 3 y 4 consultar el artefacto Modelo de Casos de Usos del Sistema (CFRA-Government-MCUSist), disponible en el expediente de proyecto del mercado de datos de la Delegación Provincial de la Agricultura.

### 2.8. Nivel de Granularidad

En el caso del mercado de datos cada tabla de hecho posee un nivel de granularidad que está dado por el grano de la información presente en el hecho y

de la información presente en cada una de las dimensiones con las que se relaciona.

A continuación se muestra la granularidad establecida para los hechos correspondientes a los procesos siembra y preparación de tierras para la ganadería. Para un mayor entendimiento consultar el artefacto Especificación del Modelo de Datos Dimensional, disponible en el expediente de proyecto del mercado de datos de la Delegación Provincial de la Agricultura.

### **H1 - hech\_siembra\_ganaderia**

La información estadística diaria perteneciente a la Delegación Provincial de la Agricultura, registrada bajo la estructura de empresas y tipo de cultivo, captada en los indicadores de los modelos de la ganadería específicamente la siembra.

### **H2- hech\_preparacion\_ganaderia**

La información estadística diaria perteneciente a la Delegación Provincial de la Agricultura, registrada bajo la estructura de empresas y tipo de cultivo, captada en los indicadores de los modelos de la ganadería específicamente la preparación.

## **2.9. Identificación de las dimensiones, hechos y medidas.**

Luego de realizar el análisis de los sistemas fuentes en el área de la agricultura se procede a identificar las dimensiones, hechos y medidas que tendrá el mercado de datos para a partir de estos poder realizar el modelo de datos.

### **Medidas**

Las medidas se refieren a los valores o datos cuantificables, normalmente numéricos que miden algunos aspectos del negocio. Las medidas numéricas definidas se almacenarán en las tablas de hechos del modelo dimensional.

### **Hechos**

En la realización del diseño del mercado de datos se identificaron 18 tablas de hechos correspondientes a cada proceso identificado en el negocio. A continuación se mostrarán los hechos identificados, para mayor detalle consultar el documento

## Capítulo 2: Análisis y Diseño del mercado de datos

Especificaciones del Modelo de Datos Dimensional que se encuentra disponible en el expediente de proyecto del mercado de datos de la Delegación Provincial de la Agricultura.

H1. hech_contratación.	H10. hech_siembra_papa.
H2. hech_demanda.	H11. hech_preparación_papa.
H3. hech_existencia_nave.	H12. hech_entrada_fertilizante.
H4. hech_otros_destinos.	H13. hech_entrada_semilla.
H5. hech_producción_leche.	H14. hech_siembra_cv.
H6. hech_producción_carne.	H15. hech_preparación_cv.
H7. hech_siembra_ganadería.	H16. hech_siembra_maiz.
H8. hech_preparación_ganadería.	H17. hech_preparacion_maíz.
H9. hech_alimentación_ganadería.	H18. hech_siembra_frijol.

### Dimensiones

Para almacenar la descripción de los elementos del negocio se identificaron 7 tablas de dimensiones, incluyendo la dimensión temporal que permitirá al usuario visualizar la información de diferentes periodos y moverse con facilidad y rapidez a través del espacio de tiempo definido para el mercado. A continuación se mostrarán las dimensiones identificadas, para mayor detalle consultar el documento Especificaciones del Modelo de Datos Dimensional que se encuentra disponible en el expediente de proyecto del mercado de datos de la Delegación Provincial de la Agricultura.

D1. dim_empresa.	D5. dim_tierras_movimiento.
D2. dim_destino.	D6. dim_producto.
D3. dim_tipo_cultivo.	D7. dim_generico

D4. dim\_temporal\_diario.

### 2.10. Modelo dimensional

Una vez definido todos los aspectos importantes para la construcción del modelo dimensional como son: las dimensiones, las medidas y la granularidad, se procede al modelado de las estructuras dimensionales. Cada dimensión es definida por su llave primaria, que sirve para mantener la referencia en las tablas de hechos a las que se acopla.

Para garantizar la comprensión del modelo dimensional de la solución se dividió en pequeñas partes que corresponden a los diferentes procesos que se modelaron para estructurar el diseño del mercado de datos.

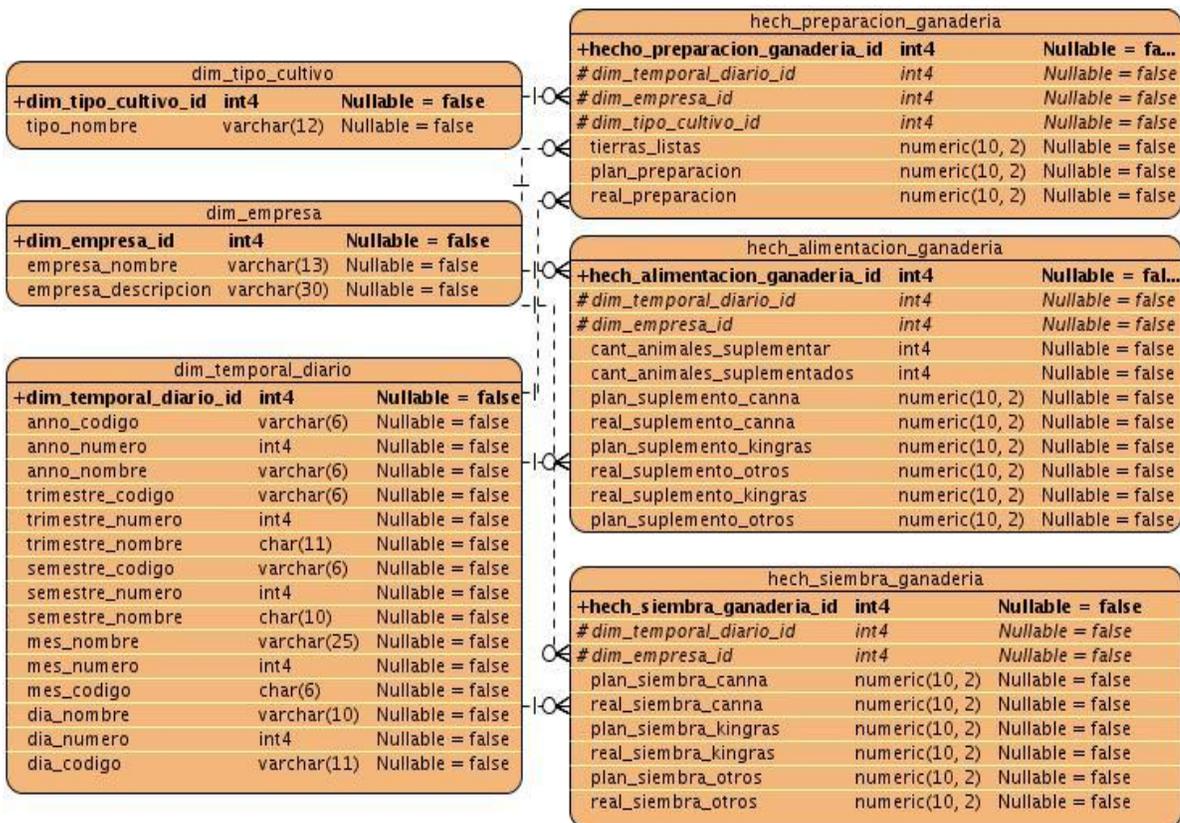


Imagen 5 : Esquema dimensional de los procesos siembra, preparación de tierras y alimentación para la ganadería.

## Capítulo 2: Análisis y Diseño del mercado de datos

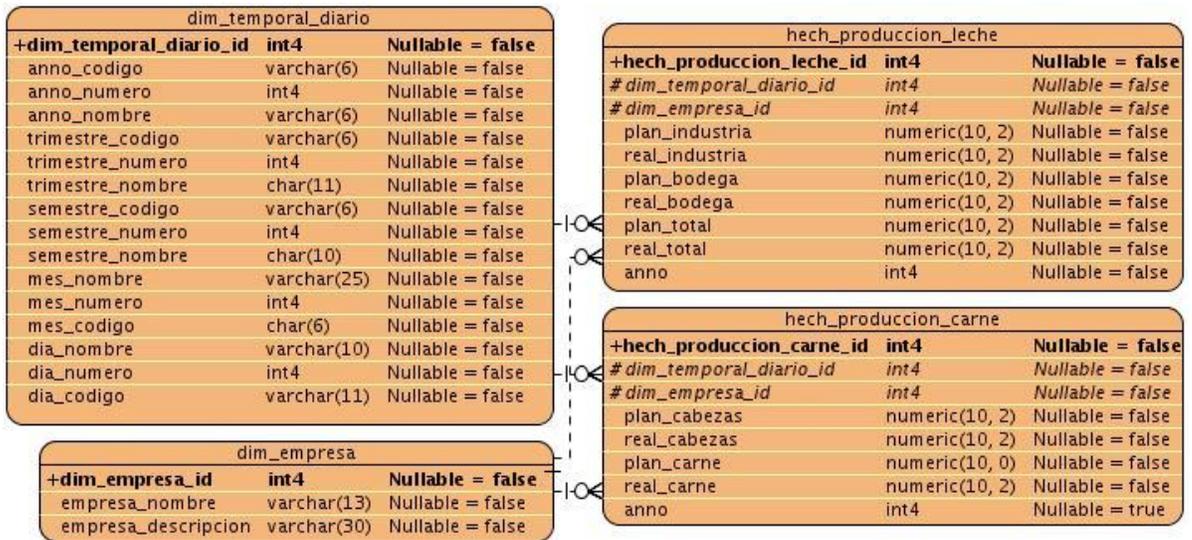


Imagen 6 : Esquema dimensional de los procesos producción de carne y leche.

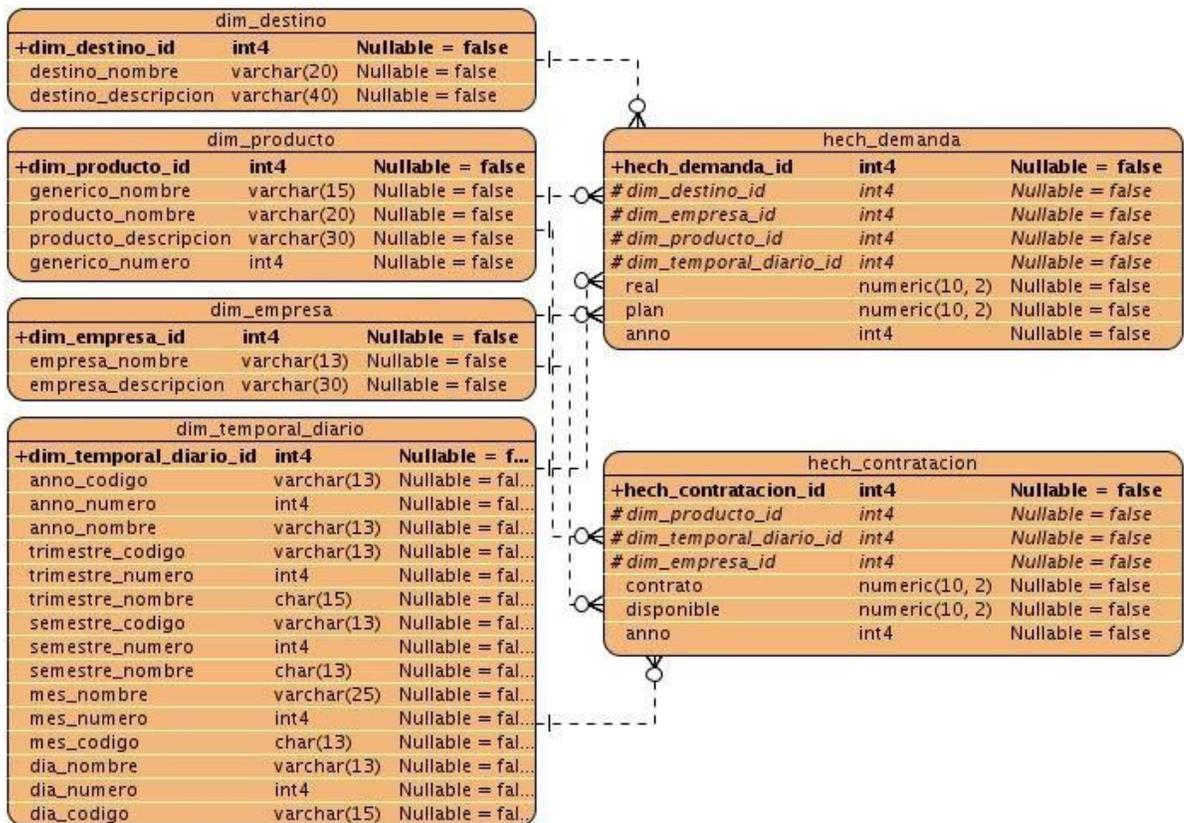


Imagen 7: Esquema dimensional de los procesos contratación y demanda

## Capítulo 2: Análisis y Diseño del mercado de datos

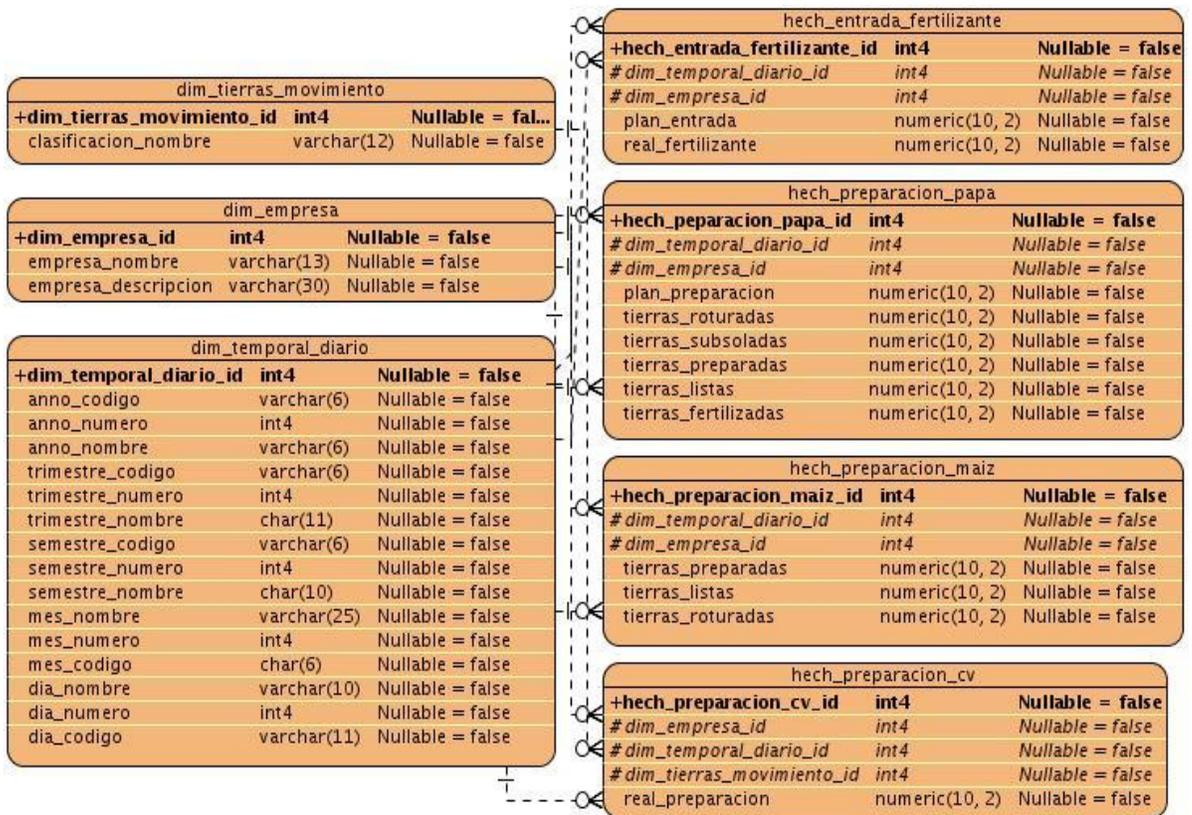


Imagen 8 : Esquema dimensional de los procesos de preparación de tierras para la siembra de papa, maíz, cultivos varios y entrada de fertilizante.

## Capítulo 2: Análisis y Diseño del mercado de datos

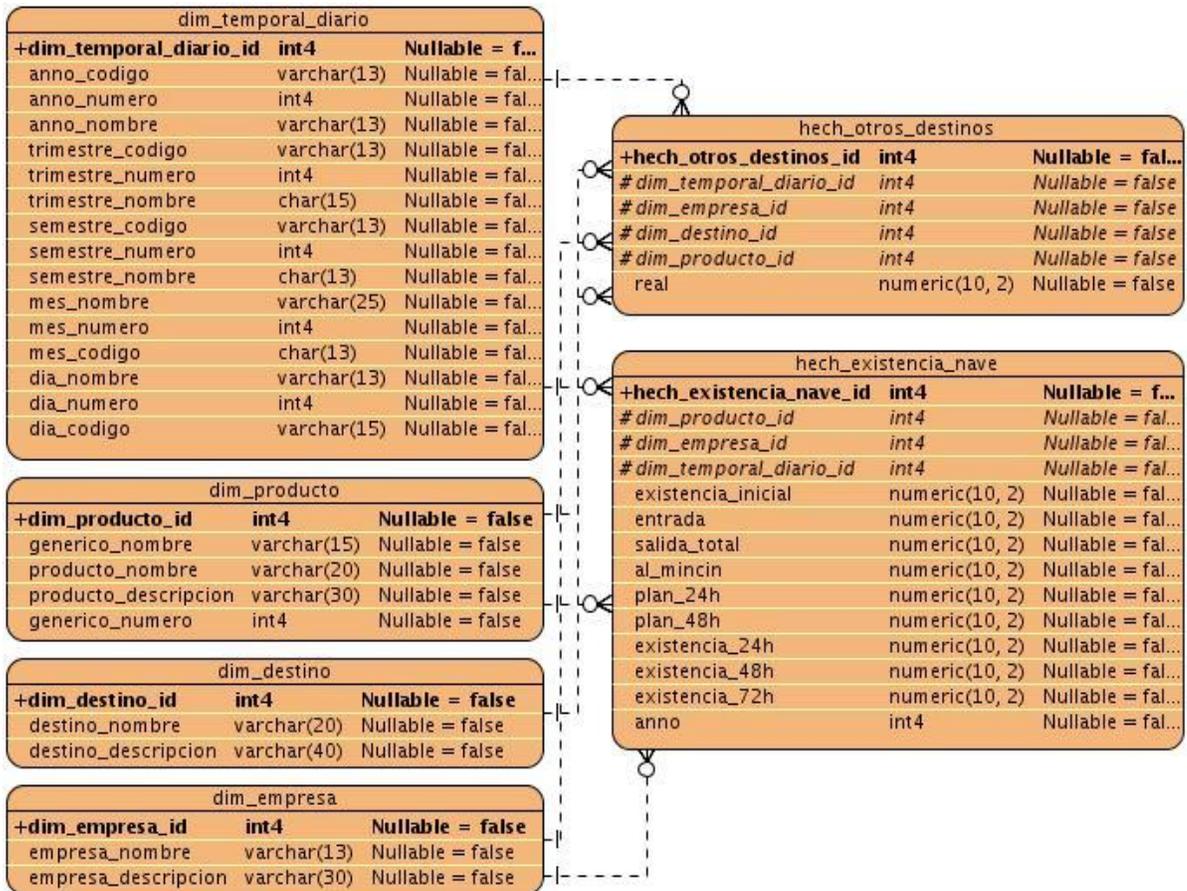
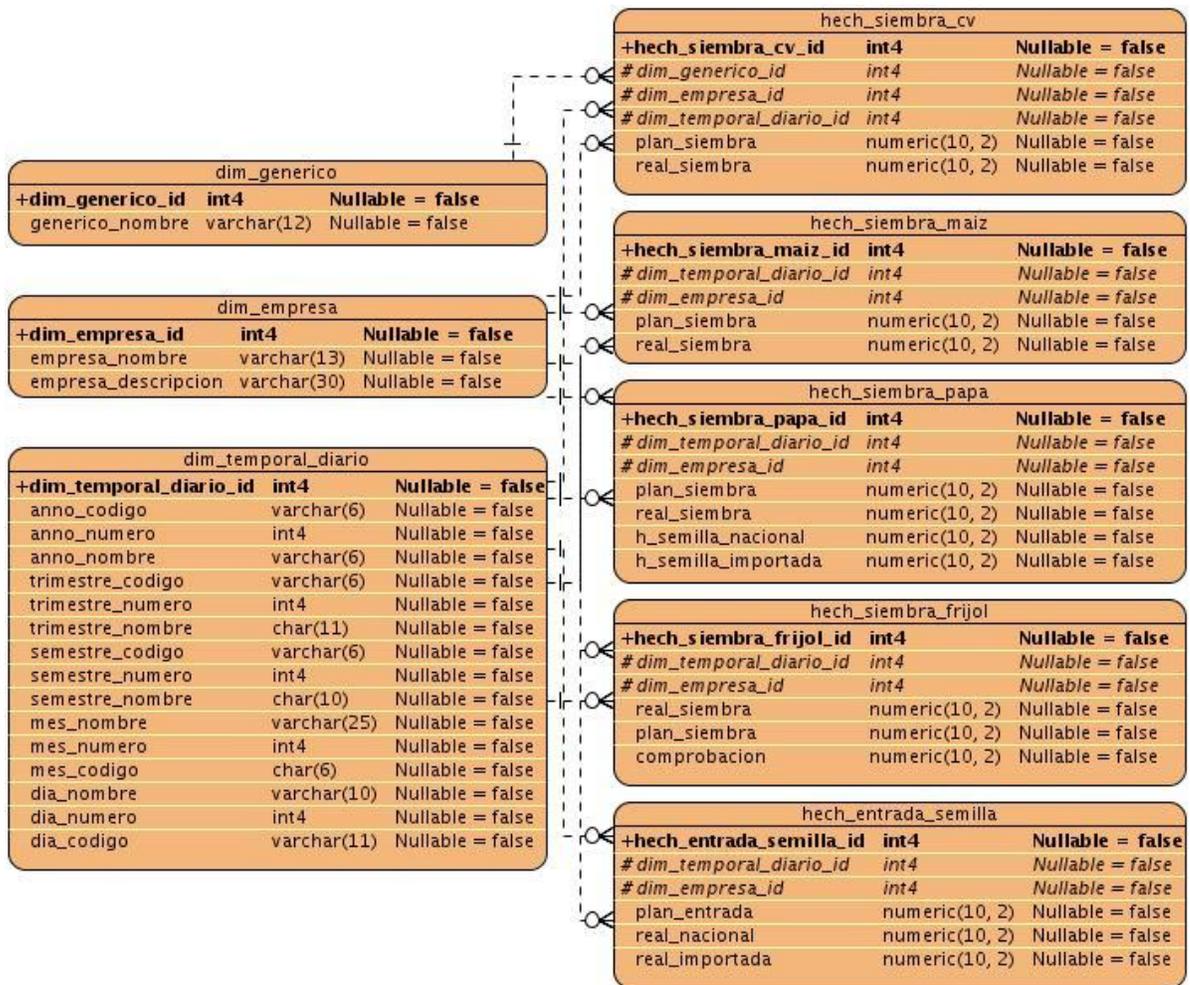


Imagen 9: Esquema dimensional del balance de existencia en nave y otros destinos.



**Imagen 10: Esquema dimensional de los procesos de siembra de papa, frijol, maíz y entrada de semilla para siembra de papa.**

## 2.11. Arquitectura de la información

Debido al volumen de datos y la variedad de indicadores que se han identificado, la arquitectura de la información del mercado de datos para la Delegación Provincial de la Agricultura estará conformada por 2 áreas de análisis, que agrupan la información en su totalidad. Contará con 9 libros de trabajo para realizar la agrupación de los reportes generados dentro de cada área de análisis y 110 reportes que dan respuesta a las necesidades de información de los clientes. A continuación se relacionan las áreas de análisis con cada uno de los libros de trabajo que contiene.



**Imagen 11: Mapa de Navegación**

Para mayor detalle sobre este tema consultar en el artefacto Arquitectura de la Información (Plantilla Arquitectura de información v2.0), disponible en el expediente de proyecto del mercado de datos de la Delegación Provincial de la Agricultura.

### **Conclusiones del capítulo**

Durante el presente capítulo se realizó el estudio del negocio, mediante el cual se obtuvo una mayor comprensión de las actividades que se realizan en la Delegación Provincial. La interacción directa con el cliente permitió que se obtuvieran las necesidades de información y posteriormente se identificaron 110 requisitos informativos. También fueron identificados 13 requisitos funcionales, 26 requisitos no funcionales y 46 reglas de negocio. Se identificaron 7 dimensiones y 18 hechos que conforman los elementos fundamentales del modelo de datos obtenido. Se identificaron además 8 casos de uso funcionales y 9 casos de uso de información en correspondencia con las exigencias del cliente. Al concluir el capítulo quedan sentadas las bases para dar paso a la implementación del mercado de datos.

### **Capítulo 3: Implementación del mercado de datos para la Delegación Provincial de la Agricultura.**

En este capítulo se aborda todo lo referente a la implementación de la estructura física de la solución, el subsistema de integración y el subsistema de visualización. Se exponen los detalles tanto del proceso de Extracción, Transformación y Carga (ETL) como de Inteligencia de Negocio (BI), teniendo en cuenta las necesidades del cliente.

#### **3.1. Implementación de las estructuras de datos**

Las estructuras de datos son una forma de organizar un conjunto de datos con el objetivo de facilitar su manipulación. En un mercado de datos se deben crear estructuras lógicas que faciliten y optimicen el tratamiento de la información, para que el manejo de los datos se realice de manera correcta.

##### **3.1.1. Esquemas**

Los esquemas son una forma de organizar los datos. En su interior pueden contener tablas, operadores, funciones, tipos de datos, a los que el usuario puede acceder siempre y cuando tenga los permisos necesarios. Los esquemas que fueron creados se describen a continuación:

**Esquema dimensiones:** Contiene todas las tablas dimensiones generales, por ejemplo: dim\_empresa, dim\_temporal\_diario.

**Esquema mart\_agricultura:** Contiene todas las tablas hechos que almacenan la información referente al área de la Delegación Provincial de la Agricultura, así como las dimensiones que son específicas de dicha área.

**Esquema errores\_mart\_agricultura:** Contiene las tablas de error donde van a ser almacenados los datos que no puedan ser cargados en los hechos correspondientes.

**Esquema metadatos:** En este esquema se encuentran las tablas en las que se guardaran los registros de las actualizaciones, cargas de datos y cambios que se realicen sobre las estructuras.

Para obtener información más detallada, consultar el artefacto CFRA-Government-Especificación del Modelo Dimensional disponible en el expediente de proyecto del mercado de datos de la Delegación Provincial de la Agricultura.

### **3.1.2. Restricciones y secuencias**

Una restricción es una especificación o condición que obliga el cumplimiento de ciertas normas en la base de datos. Las mismas pueden ser creadas automáticamente al definir una tabla (en el caso de las llaves primarias) o ser introducidas por el programador de la base de datos, por ejemplo, que un campo tenga una longitud determinada. Las restricciones no son parte formal del modelo relacional, pero se incluyen para organizar mejor los datos.

Como las tablas y las funciones, las secuencias son objetos de las bases de datos. Se emplean para generar valores enteros únicos que se incrementan secuencialmente cada vez que la tabla en la que se encuentran es consultada. Son utilizadas generalmente para las llaves primarias de las tablas garantizando que los valores nunca se repitan.

Para la implementación de la solución se definieron 26 llaves primarias en correspondencia con la cantidad de tablas del mercado de datos y por lo tanto 26 secuencias, debido a que las llaves primarias se incrementarán de forma secuencial.

## **3.2. Roles y privilegios de acceso**

### **3.2.1. Acceso al mercado de datos**

Con el objetivo de proteger la información almacenada en el mercado de datos de accesos no autorizados quedan establecidos un conjunto de privilegios que han sido asignados a los usuarios de acuerdo al nivel de accesibilidad que necesitan.

Usuario	Permisos
Administrador	Tiene acceso total a las estructuras del mercado de datos para la administración y configuración de las mismas.
Programador ETL	Tiene asignados los permisos que se requieren solo para la realización de los procesos de extracción, transformación, carga y perfilado de datos.

**Tabla 2: Roles y privilegios de acceso al mercado de datos**

### 3.2.2. Acceso a la aplicación

Para el acceso a las vistas de análisis y reportes que se encuentran publicadas en el servidor de BI, resulta mucho más factible la creación de varios grupos de usuarios cada uno con sus permisos correspondientes acorde al papel que juegan en el negocio. De esta forma se garantiza la privacidad de la información.

Rol	Permisos
Administrador	Tiene acceso total a las funcionalidades que brinda el mercado, con capacidad de gestión sobre las mismas.
Directivo	Posee los permisos necesarios para el análisis de la información, visualización de reportes, vistas de análisis y creación de reportes Ad-hoc.

**Tabla 3: Roles y privilegios de acceso a la aplicación**

### 3.3. Implementación del subsistema de integración

ETL es el proceso que organiza el flujo de los datos y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes hacia una única estructura, en este caso el mercado de datos.

#### Extracción de los datos

Uno de los primeros pasos para dar comienzo al proceso de ETL es el estudio de las fuentes de datos. En el caso de los datos históricos, la información se obtiene a través de los ficheros xls<sup>9</sup>, que son la herramienta de trabajo de la Delegación en la actualidad. Antes de comenzar la extracción se realizaron tareas de estandarización, con el objetivo de lograr una estructura uniforme de las fuentes de datos que permitiera la adecuada obtención de la información. Luego se procede a la extracción de los datos de las distintas fuentes.

### **Transformación y Limpieza**

Una vez realizado el proceso de extracción, se realizó la limpieza de los datos provenientes de las diferentes fuentes, ya que los mismos pueden tener errores o estar incompletos. El proceso de transformación de los datos se realizó de acuerdo a las reglas del negocio que se definieron, como es el caso de realizar tratamientos de valores nulos sustituyéndolos por valores preestablecidos. Con esto se busca obtener datos completos, consistentes, y lo más precisos posibles.

### **Carga**

En este paso final del proceso, los datos fueron cargados al mercado, poniendo en práctica todas las restricciones que se definieron. Después de realizar la carga de los datos históricos, estos se encuentran ya organizados y actualizados para que el cliente pueda usarlos y manipularlos satisfactoriamente. En lo adelante, para realizar las cargas incrementales la fuente será la base de datos del Sistema Informativo del Gobierno (SIGOB), con la que el cliente trabajará diariamente introduciendo información que más tarde pasará a formar parte del mercado de datos.

#### **3.3.1. Implementación de las transformaciones**

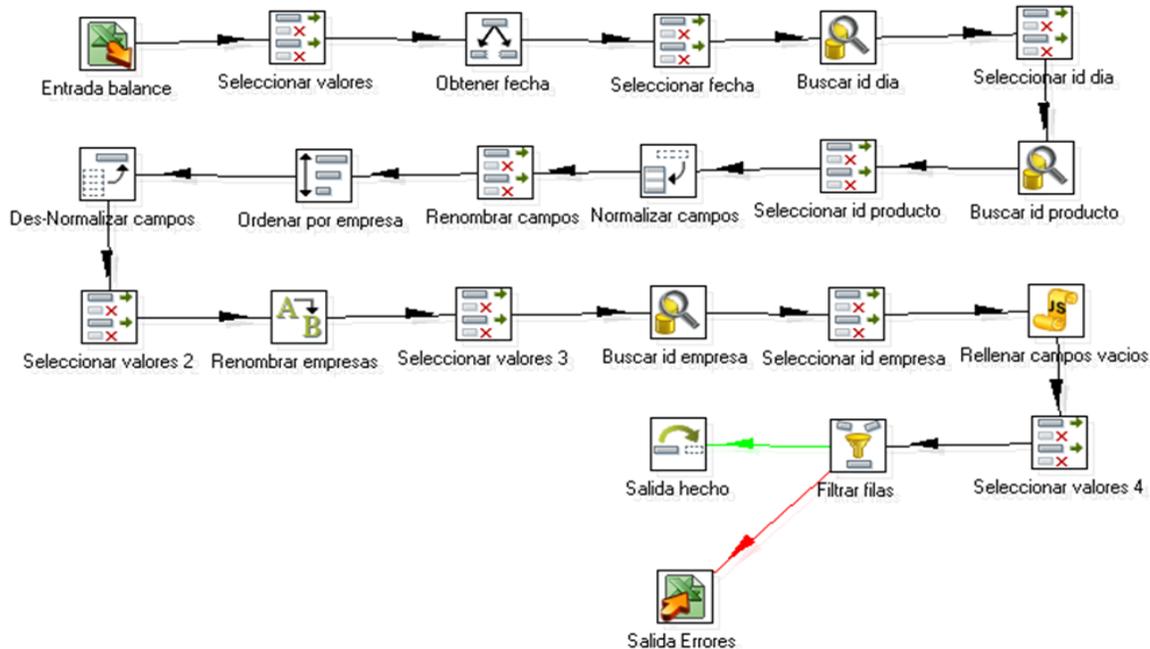
Para la realización del proceso de ETL correspondiente a la carga de los datos históricos, se implementaron 22 transformaciones en total, de las cuales 21 se utilizan para la carga de los hechos y 1 para la carga de las dimensiones.

---

<sup>9</sup> Es un acrónimo de *Microsoft Excel Spreadsheet* (Hoja de cálculo de Microsoft Excel) y pertenece a la categoría extensión de archivo

## Capítulo 3: Implementación del mercado de datos

A continuación se muestra la imagen de la transformación que carga los datos del hecho balance de existencia en nave:



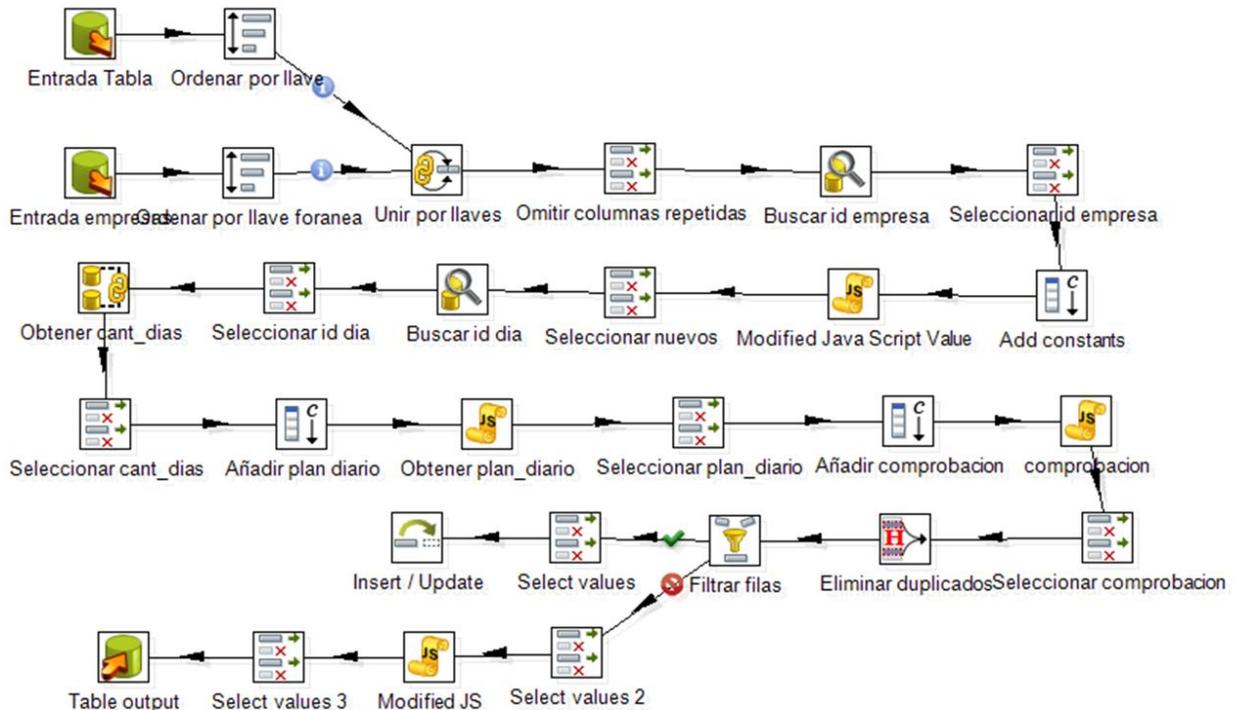
**Imagen 12: Transformación para la carga del hecho existencia en nave**

En esta transformación se cargan los indicadores de la existencia en nave de cultivos varios a partir de un fichero xls hacia la tabla hech\_existencia\_nave que se encuentra en la base de datos. Mediante las búsquedas realizadas se obtienen los campos id de las dimensiones con las que se relaciona dicha tabla de hechos. Se realizan validaciones para que no se inserten datos nulos, y los pasos de normalización y desnormalización de campos permiten obtener la estructura necesaria para cargar correctamente en el mercado. Es de vital importancia destacar que los datos erróneos son salvados en ficheros xls para facilitar su manipulación en la corrección del error.

Paralelo al desarrollo de las transformaciones para la carga de los datos históricos se efectuó la implementación de las transformaciones para las cargas incrementales. Se implementaron 46 transformaciones para la carga de las tablas

## Capítulo 3: Implementación del mercado de datos

de hechos y actualización de las dimensiones. La imagen que se muestra a continuación corresponde a la carga del hecho siembra de frijol.



**Imagen 13: Transformación para la carga del hecho siembra de frijol**

La imagen #13 muestra la transformación realizada para la carga del hecho siembra de frijol. Inicialmente se extraen los datos de la fuente (en este caso la base de datos del SIGOB). Posteriormente se realiza una unión por clave, dado que para obtener la información de un mismo hecho se requiere obtener datos de diferentes tablas y es necesario unir los datos que son comunes para situarlos en un mismo componente. Se añade una constante denominada comprobación y se le asigna valor a través de cálculos definidos en un componente Java Script. Luego se realizan las búsquedas en la base de datos para encontrar las coincidencias entre la fuente y las dimensiones, se eliminan los duplicados, se valida que los identificadores resultantes de las búsquedas no provengan nulos y finalmente se inserta la información en la tabla destino. Los errores que ocurran en la transformación serán cargados en una tabla de error que se encuentra en el

esquema errores\_mart\_agricultura, con el objetivo de garantizar su persistencia y fácil acceso.

### 3.3.2. Implementación de los trabajos

Cuando la conexión al mercado de datos se encuentra funcionando correctamente, se procede a la carga del mismo. Los trabajos posibilitan la agrupación de las transformaciones que tengan un objetivo en común o una misma periodicidad de carga para ser ejecutados en conjunto de forma secuencial.

Para que los trabajos se realicen satisfactoriamente se debe definir cuáles son las dimensiones estáticas y cuáles no, ya que en un trabajo sólo se incluyen aquellas dimensiones en las que los valores que tenían anteriormente puedan sufrir cambios.

Se implementaron en total 17 trabajos, de los cuales 3 pertenecen a la carga de los datos históricos y los restantes a las cargas incrementales.

La imagen #14 muestra el trabajo que agrupa las cargas de la información histórica de los 18 hechos del mercado de datos, así como la carga de las dimensiones específicas.



**Imagen 14: Trabajo poblar mercado de datos**

En la imagen #15 se muestra el trabajo que agrupa las cargas incrementales del mercado de datos. En dependencia de las transformaciones que se hallan en el interior de cada trabajo, se ha programado la carga de forma tal que la información no se almacene duplicada y se realice la actualización de las tablas de metadatos correspondientes para llevar a cabo correctamente el control de cambios.



**Imagen 15: Trabajo que agrupa las cargas incrementales**

### 3.4. Implementación del subsistema de visualización

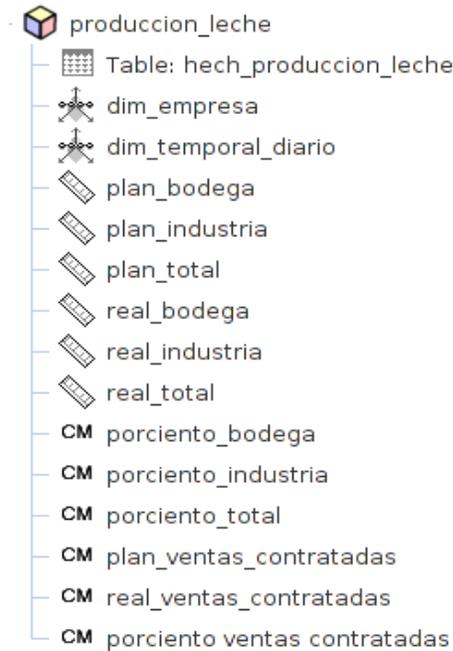
Al finalizar los procesos de ETL, todos los datos históricos han sido cargados en las estructuras del mercado y se encuentran disponibles para ser consultados. Se procede entonces a la implementación de los subsistemas de visualización los cuales permitirán el análisis y disposición de los datos acumulados para transformarlos en información que permita generar el conocimiento necesario para agilizar el proceso de toma de decisiones.

#### 3.4.1. Implementación de los cubos de datos

Para lograr una adecuada presentación de los datos al usuario, uno de los elementos fundamentales a implementar son los cubos multidimensionales, los cuales pueden poseer un número indefinido de dimensiones por lo que son llamados hipercubos. Permiten organizar la información estructuralmente y proveen un rápido acceso a los datos almacenados independientemente de la cantidad de datos a visualizar.

En el sistema se definieron 18 cubos principales y otros 10 que organizan la información almacenada mediante las vistas materializadas, representando los dos

últimos años analizados en el negocio. Se hizo uso de la herramienta Pentaho Workbench para el modelado de su estructura. A continuación se mostrará uno de estos cubos.



**Imagen 16: Diseño del cubo multidimensional del proceso producción de leche**

Los cubos se definieron uno por cada tabla hecho, debido a que estas tablas son las fuentes de información principal de la organización estructural de los datos, además se observan las tablas dimensiones relacionadas con este hecho así como las medidas y los campos calculables que posee cada uno.

### 3.4.2. Vistas de análisis y reportes

Las vistas de análisis y los reportes son la información que el cliente desea que se muestre como finalidad del producto. Los mismos fueron identificados luego de realizar un análisis de los ficheros Excel donde se recoge toda la información referente de la Delegación Provincial de la Agricultura. La arquitectura de información del mercado de datos está definida por 9 libros de trabajo, en los cuales van a estar las vistas y los reportes asociados a cada uno de ellos.

La solución cuenta con 43 vistas de análisis que se visualizan mediante la

## Capítulo 3: Implementación del mercado de datos

herramienta Pentaho BI-Server. Dicha herramienta facilita el análisis de la información mediante tablas de datos y gráficas, permitiendo realizar cambios a los reportes como: modificar, eliminar, cambiar el tipo de gráfica, imprimir, guardar en un archivo con formato pdf o xls y crear un nuevo reporte.

A continuación se describirá un ejemplo de los reportes candidatos que fueron identificados:

- Obtener el porcentaje de cumplimiento con respecto a la producción total, el porcentaje de cumplimiento con respecto a la industria, el porcentaje de cumplimiento con respecto a la bodega y el porcentaje de cumplimiento con respecto a las ventas contratadas de la producción de leche diaria/mensual/semestral/trimestral/anual por empresa.

En esta vista de análisis se muestran los porcentos de cumplimiento de las empresas en la producción de leche con respecto a los planes de producción total, de la industria, de la bodega y de las ventas contratadas en series de tiempo.

Empresa	Tiempo	Medidas			
		• porcentaje total	• porcentaje industria	• porcentaje bodega	• porcentaje ventas contratadas
▢ Todas las empresas	▢ Ver todo	41.70%	28.76%	38.91%	33.33%
Los Naranjos	▢ Ver todo	39.36%	27.32%	36.20%	28.25%
	▢ 2011	39.36%	27.32%	36.20%	28.25%
	▢ 2do Semestre	39.36%	27.32%	36.20%	28.25%
	▢ 4to Trimestre	39.36%	27.32%	36.20%	28.25%
	▢ octubre	39.36%	27.32%	36.20%	28.25%
	1	93.40%	76.03%	85.16%	76.99%
	2	92.60%	72.63%	108.61%	76.41%
	3	93.96%	77.91%	78.24%	77.95%
	4	93.80%	77.91%	78.24%	77.95%
	5	93.96%	72.63%	108.61%	76.41%
	6	93.85%	74.94%	93.35%	76.87%
7	93.96%	72.63%	108.61%	76.41%	
8	94.30%	8.35%	59.48%	13.72%	

**Imagen 17: Vista de análisis % de cumplimiento del plan de producción de leche.**

## Capítulo 3: Implementación del mercado de datos

La solución cuenta con 44 reportes, entre los que se encuentran los cortes a nivel de provincia y los partes de cumplimiento e incumplimiento a nivel de empresas, incluyendo gráficas de barras que evidencian el comportamiento de los diferentes indicadores. Para el diseño de las plantillas y la implementación de los reportes se utilizó la herramienta Pentaho Report Designer. A continuación se enuncia y describe un ejemplo que corresponde al siguiente pedido de información:

- Obtener plan de caña, real de caña, % de cumplimiento con respecto al plan de de caña, cantidad de tierras listas para caña, plan de cultivo total, real de cultivo total, % de cumplimiento con respecto al plan de cultivo total, cantidad de tierras listas para cultivo total de la preparación de tierras para la ganadería diaria en la provincia.

Cultivo total				Cultivo de caña			
Tierras en movimiento			De ella: listas	Tierras en movimiento			De ella: listas
Plan	Real	%		Plan	Real	%	
3,23	42,44	1313,93	31,00	0,39	5,30	1358,97	17,90



**Imagen 18: Reporte parte diario provincial de la preparación de tierras para ganadería**

La imagen muestra la estructura del reporte que responde al corte de información provincial de la preparación de tierras para la ganadería. Se visualizan los indicadores asociados al proceso como son el plan, real, % de cumplimiento y cantidad de tierras listas, clasificados por el tipo de cultivo al que pertenecen

(cultivo total y cultivo de caña en el presente caso). Graficar los niveles medidos para cada uno de los indicadores otorga la facilidad de conocer a simple vista la diferencia que existe entre los tipos de cultivos, y de forma más específica los niveles de cumplimiento obtenidos en la provincia.

### **3.4.3. Reportes Ad-hoc**

Hasta el momento las vistas de análisis y reportes creados satisfacen las necesidades de información de los usuarios, pero también es necesario tener en cuenta que en una situación determinada puede surgir un nuevo pedido de información de carácter inmediato que no se encuentre incluido en los reportes predefinidos. El cliente siempre necesita variar un reporte o contar con la posibilidad de comenzar desde cero, para poder elegir los indicadores que desea medir o simplemente agregarle un nuevo filtro. Para dar soporte a los nuevos requerimientos que puedan surgir se implementan herramientas de consulta Ad-hoc. Estas se crean con el objetivo de facilitar al usuario final la creación de reportes capaces de responder al instante preguntas del negocio, sin la necesidad de verse obligado a realizar consultas complejas sobre el mercado de datos.

Mediante las facilidades que brindan los reportes Ad-hoc se brinda tanto a los usuarios experimentados como a los que poseen menos conocimientos, la capacidad de acceder de forma oportuna y sencilla a la información que requiere en una situación imprevista. Luego si el pedido se hace habitual es aconsejable incluirlo a los reportes predefinidos.

Para que el usuario pueda diseñar su propio reporte a partir de la información que contiene el mercado de datos, se implementaron los metadatos utilizando la herramienta Pentaho Metadata Editor.

### **3.5. Técnicas de optimización**

La aplicación de técnicas de optimización para el mercado de datos, tiene como propósito fundamental mejorar el rendimiento de las consultas al verse obligado a procesar grandes cantidades de datos. Para el cliente es fundamental obtener la

información necesaria en el momento preciso, no antes o después, dado que la accesibilidad a los datos referentes al negocio es un punto clave para la toma de decisiones. En el mercado de datos se aplicaron técnicas de indexado y la creación de vistas materializadas.

### 3.5.1. Índices

Son estructuras de datos utilizadas para mejorar la velocidad de las operaciones, permitiendo un rápido acceso a los registros de una tabla en una base de datos. Al aumentar drásticamente la velocidad de acceso, se suelen usar sobre aquellos campos sobre los cuales se hacen frecuentes búsquedas. En los mercados de datos se almacenan gran cantidad de datos, lo que hace que el número de registros de las tablas que lo componen alcancen valores extremadamente altos. En la presente propuesta de solución se hace uso de los índices para evitar la realización de lecturas secuenciales en las búsquedas lo que permitiría disminuir el tiempo de ejecución de las consultas.

PostgreSQL implementa cuatro tipos diferentes de índices (Árboles B, Árboles R, Hash y GiST) cada uno pensado idealmente para cubrir ciertos tipos de consultas. En este caso se va hacer uso del método Árboles B para la creación de los índices del mercado. En el artefacto Especificaciones del diseño físico que se encuentra disponible en el expediente del proyecto del mercado de datos para la Delegación Provincial de la Agricultura se describen los detalles del plan de indexado a aplicar. A continuación se muestra un ejemplo del código SQL empleado para la creación de los índices.

```
-- Index: mart_agricultura.Idx_generico_numero
-- DROP INDEX mart_agricultura.Idx_generico_numero;

CREATE INDEX Idx_generico_numero
ON mart_agricultura.dim_producto
USING btree
(generico_numero);
```

**Imagen 19: Código para la creación de índices**

### 3.5.2. Vistas materializadas

A partir del momento en que el mercado de datos es desplegado y comienza su uso en el entorno de trabajo, se encuentra constantemente sujeto a nuevas cargas. Por consiguiente va aumentando rápidamente su tamaño y al utilizarlo aumenta la dificultad de responder las consultas de los usuarios de forma rápida y eficiente. Las vistas materializadas son utilizadas con el propósito de optimizar consultas complejas que tengan altos tiempos de respuesta. Para el mercado de datos se definió un conjunto de vistas, cuidadosamente escogidas a partir de un análisis de las consultas más frecuentes ejecutadas en el mercado de datos.

Con cierto grado de periodicidad el cliente necesita obtener los datos de los procesos que se modelaron en el mercado con una perspectiva histórica de 1 ó 2 años de antigüedad. Los partes de producción de carne y leche son los reportes más frecuentemente utilizados del área de análisis ganadería y del área de análisis cultivos varios los partes de cumplimiento de la demanda y existencia en nave. Para lograr la optimización en cuanto al acceso y visualización de los datos se crearon vistas materializadas que se actualizan con información referente a los procesos antes mencionados del año actual y los dos últimos años.

El SGBD utilizado no incluye entre sus funcionalidades la creación de vistas materializadas, por lo que es necesario simularlas por medios de funciones disparadoras (*trigger*). Para implementar las vistas materializadas se crea una tabla exactamente igual al hecho que se desea resumir, y como parte de la estructura de la nueva tabla una función que permita actualizar sus valores (Ver Anexo 4) y un disparador que ejecute la función (Ver Anexo 3). Dentro de la estructura del hecho del mercado se crea de igual forma una función, solo que en este caso se encarga de insertar o eliminar de la vista los datos que cumplan o no con una determinada condición (Ver Anexo 2), además de implementar un disparador que permita ejecutar dicha función (Ver Anexo 1).

### Conclusiones del capítulo

## Capítulo 3: Implementación del mercado de datos

En este capítulo quedan expuestos los detalles de la implementación de las tablas físicas del mercado de datos en PostgreSQL, ajustándose al modelo físico diseñado previamente y a las convenciones de nombrado establecidas. De igual forma quedan definidos los esquemas, secuencias y restricciones. Se obtuvo con éxito la información necesaria para proceder a la realización del proceso ETL, por medio del cual se transformó y cargó hacia el mercado una muestra representativa de cada categoría de la información que se manipula en la Delegación. Luego de haber cargado ejemplos de datos en cada tabla, se realizó la implementación del subsistema de visualización, quedando así conformadas las vistas de análisis y los reportes que darán respuesta a las necesidades de información del cliente. Para la optimización de las consultas se realizó el indexado mediante Árboles B y las vistas materializadas.

## Capítulo 4: Validación de las funcionales del mercado de datos para la Delegación Provincial de la Agricultura.

Una vez que se ha dado por concluida la fase de implementación, se debe dar paso a una de las etapas más importantes en el ciclo de desarrollo de un software: las pruebas. Su aplicación revelará no conformidades que servirán de guía para la corrección de las deficiencias encontradas, obteniéndose como resultado una aplicación con mayor calidad, eficiencia y confiabilidad. De esta forma se garantiza que se haya cumplido con las especificaciones iniciales que quedaron definidas para el mercado de datos.

### 4.1. Proceso de prueba

La prueba es el proceso de ejecución de un programa con el fin de encontrar deficiencias. Una prueba tiene éxito si descubre errores que no han sido detectados hasta entonces.

Las pruebas deben centrarse en dos objetivos:

1. Probar si el software no hace lo que debe hacer.
2. Probar si el software hace lo que no debe hacer. (Pressman, 2007)

Existen diferentes tipos de pruebas, cada uno aplicable en un entorno diferente de acuerdo a los objetivos que persiga y en dependencia de las personas involucradas en su realización:

Niveles de Pruebas			
Pruebas	Objetivo	Participantes	Ambiente
Unitaria	Detectar errores en los datos, lógica, algoritmos.	Programadores	Desarrollo
Integración	Detectar errores de interfaces y	Programadores	Desarrollo

	relaciones entre componentes.		
<b>Funcional</b>	Detectar errores en la implementación de requerimientos.	Probadores, Analistas	Desarrollo
<b>Sistema</b>	Detectar fallas en el cubrimiento de los requerimientos.	Probadores, Analistas	Desarrollo
<b>Aceptación</b>	Detectar fallas en la implementación del sistema.	Probadores, Analistas, Clientes	Productivo

**Tabla 4: Detalles de los niveles de pruebas**

## 4.2. Pruebas aplicadas al mercado de datos para la Delegación Provincial de la Agricultura

### 4.2.1. Diseño y aplicación de los casos de prueba funcionales

Los casos de prueba se componen de un conjunto de condiciones o variables bajo las cuales el analista determinará si los requisitos del mercado de datos son parcialmente o completamente satisfactorios. Son un tipo de prueba funcional que ayuda a verificar que el producto desarrollado realice las funciones para las que ha sido creado en base a los requerimientos que ha solicitado el usuario final.

Para verificar los requerimientos del sistema se llevó a cabo el diseño de los casos de prueba, definiendo la realización de un caso de prueba para cada caso de uso, lo que garantiza cubrir la verificación de cada uno de los requisitos. Se diseñaron 9 casos de prueba basados en casos de uso, los detalles de los resultados, las entradas y las condiciones evaluadas pueden ser consultadas a través de los artefactos “Casos de Prueba basados en casos de uso” que se encuentran disponibles en el expediente de proyecto del mercado de datos de la Delegación Provincial de la Agricultura. Los resultados que arrojaron dichas pruebas fueron satisfactorios.

### 4.2.2. Pruebas de unidad

Las pruebas de unidad tienen como objetivo verificar la funcionalidad y estructura de cada componente individualmente una vez que ha sido implementado. Se realizaron pruebas unitarias a todas las transformaciones que cargan los datos para las tablas de hecho y dimensiones del almacén.

En cada caso se verificó que todos los campos fueron cargados con los valores provenientes de la base de datos SIGOB, sin que se perdiera información. Se comprobó que no existieran datos duplicados ni valores nulos y que los campos calculables que se ejecutan en los procesos de ETL fueran resueltos de acuerdo a las reglas del negocio. Al concluir la realización de las pruebas se obtuvieron resultados satisfactorios.

La documentación referente al tema se encuentra en el artefacto Registro de pruebas unitarias (CFRA-Government-RegPruebaUnitaria), disponible en el expediente de proyecto del mercado de datos de la Delegación Provincial de la Agricultura.

### 4.3. Resultados obtenidos

Luego de haber concluido la implementación y prueba de la solución, se ha logrado la integración, limpieza y disponibilidad de los datos históricos de la Delegación Provincial de la Agricultura de Artemisa, así como la rápida visualización de la información con motivo de agilizar el análisis.

El resultado final es un mercado de datos que sirva de apoyo en el proceso de toma de decisiones basándose en información del negocio, que cuenta con las siguientes funcionalidades:

Visualización de la información mediante: Vistas de análisis, reportes de datos, reportes Ad- hoc y gráficos.

Seguridad de la información mediante: Autenticación de usuarios en la aplicación.

### 4.4. Aporte social y económico

Mediante el desarrollo de un mercado de datos para la Delegación Provincial de la Agricultura se brinda una herramienta para el análisis de la información con motivo de la toma de decisiones, que genera valor agregado a partir de los datos de los procesos del negocio, dotando a la entidad de un activo muy valioso: El conocimiento.

La utilización del mercado de datos repercute directamente en la economía de la provincia debido a que posibilita el acceso inmediato al conocimiento acumulado de la organización, permitiendo a los ejecutivos tomar decisiones más rápidas y mejor informadas. Mejores decisiones convergen hacia mejores estrategias, lo que resulta en el aumento de los niveles productivos y se gana tanto en calidad como en cantidad. Al aumentar la cantidad, se reducen los costos de los productos y así se logra que se encuentren más accesibles a la mayoría de la población.

Se eliminan las pérdidas de tiempo esperando por información que finalmente es incorrecta o no encontrada, así se optimizan los procesos productivos. El trayecto de la información desde el origen hasta su destino se simplifica y se hace más corto, evitándose de esta forma la pérdida y corrupción de los datos estadísticos que llevarían consigo grandes afectaciones económicas.

### Conclusiones del capítulo

Durante el transcurso del presente capítulo se exponen los detalles de las pruebas del mercado de datos aplicando pruebas unitarias y funcionales. Durante esta fase se diseñaron y aplicaron los casos de pruebas, las deficiencias encontradas fueron corregidas de forma exitosa arrojando como resultado un producto con mayor calidad y más completo en funcionalidad.

### **Conclusiones generales**

Al finalizar el proceso de desarrollo del mercado de datos se puede afirmar que se le ha dado cumplimiento de forma satisfactoria al objetivo general del presente trabajo de diploma, obteniéndose con creces los resultados esperados. De igual forma durante el ciclo completo se cumplieron estrictamente cada una de las tareas de la investigación planteadas.

1. Se realizó la fundamentación teórica, mediante la cual quedaron establecidas las bases teóricas que sustentan la solución y la metodología a utilizar, así como las herramientas idóneas.
2. Se caracterizó el proceso de análisis de la información en la Delegación Provincial de la Agricultura de Artemisa.
3. Se desarrolló el mercado de datos de la Delegación Provincial de la Agricultura de Artemisa.
4. Se validó funcionalmente el mercado de datos de la Delegación Provincial de la Agricultura de Artemisa, alcanzando resultados satisfactorios.

Al concluir la investigación se obtuvo una herramienta funcional de inteligencia de negocios que servirá de apoyo a la toma de decisiones en la Delegación Provincial de la Agricultura de Artemisa.

### **Recomendaciones**

Poblar el mercado de datos con toda la información estadística de la Delegación Provincial de la Agricultura para poder hacer uso de las funcionalidades que brinda.

Realizar las pruebas de aceptación con el propósito de identificar si existe alguna deficiencia dada a raíz del trabajo con el mercado de datos en un entorno real de producción.

### Referencias bibliográficas

**Comunidad de Pentaho. 2010.** community.pentaho.org. [Online] 2010. [Cited: Enero 27, 2012.] [community.pentaho.com/projects/bi\\_platform/Pentaho BI Suite.pdf](http://community.pentaho.com/projects/bi_platform/Pentaho%20BI%20Suite.pdf).

**Apache Software Foundation. 2012.** Apache Tomcat. [Online] 2012. [Cited: febrero 17, 2012.] [http://tomcat.apache.org/...](http://tomcat.apache.org/)

**ArPUG. 2009.** ArPUG. [Online] 2009. [Cited: Enero 29, 2012.] <http://www.arpug.com.ar/trac/wiki/PgAdmin>.

**Bernabeu, Ricardo Dario. 2009.** *DATA WAREHOUSING: Investigación y Sistematización de Conceptos - HEFESTO: Metodología propia para la Construcción de un Data Warehouse.* [PDF] Córdoba, Argentina : s.n., Abril 21, 2009. Vol. 1.1.

**Curto, Josep. 2007.** Information Management Data Warehousing .Data Warehouse y Datamart. [Online] 2007. [Cited: Marzo 12, 2012.] <http://informationmanagement.wordpress.com/2007/10/07/data-warehousing-data-warehouse-y-datamart/>.

**González Hernández, Yanisbel. 2011.** *PROPUESTA DE METODOLOGIA PARA EL DASARROLLO DE ALMACENES DE DATOS EN DATEC.* La Habana : s.n., 2011.

**Gravitar . 2012.** Gravitar . [Online] 2012. [Cited: Enero 28, 2012.] [http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/...](http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/)

**Guijarro Riera, Angélica María and Amalia Alexandra, Guijarro Riera. 2010.** *Diseño e Implementación de un Sistema de Control del Proceso de Crédito y Cobranza por medio de Indicadores de Gestión para una Sociedad Financiera ubicada en la ciudad de Guayaquil.* Guayaquil : s.n., 2010.

**Herrera, C. 2007.** Adictos al trabajo. [Online] octubre 30, 2007. <http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=datawarehouse4..>

**Hobbs, Lilian and otros. 2005.** *Oracle Data Base 10g Data Warehousing.* EUA : ELSEVIER Digital Press, 2005.

**Humantumba, Rayner. 2007.** *DataMart paso a paso.* [www.ruedatecnologica.com] 2007.

**Inmon, William. 2005.** Building the Data Warehouse. EUA : Wiley Publishing Inc, 2005.

**Kimball, Ralph and Ross, Margv. 2002.** The Data Warehouse Toolkit. EUA : Wiley Publishing Inc, 2002.

**Kimball, Ralph. 2009.** *The Data Warehouse Toolkit*. 2009.

**msdn. 2011.** msdn. [Online] 2011. [Cited: febrero 6, 2012.] <http://msdn.microsoft.com/es-es/library/ms244714%28v=vs.80%29.aspx...>

**Navarro, Alberto Límia, et al. METODOLOGÍA PARA EL DESARROLLO DE SOLUCIONES DE ALMACENES DE DATOS E INTELIGENCIA DE NEGOCIO EN CENTALAD.** Universidad de las Ciencias Informáticas. La Habana : s.n.

**Ortiz, Julio.E. 2009.** *Diseño e implementación de un Mercado de Datos para la Oficina Nacional de Estadísticas*. 2009.

**Ponniah, Paulraj. 2001.** *Data Warehousing Fundamentals*. EUA : Wiley Publishing Inc, 2001.

**Pressman, R.S. 2007.** *Ingeniería del Software. Un enfoque práctico*. 6ta Edición. 2007. Vol. I.

**Productive. 2007.** Business Intelligence. [Online] 2007. [Cited: Enero 29, 2012.] [http://www.productive.com.ar/productos\\_bi.php...](http://www.productive.com.ar/productos_bi.php...)

**Sinnexus. 2011.** Sinnexus. [Online] 2011. [Cited: Marzo 2012, 14.] <http://www.sinnexus.com>.

**Torres, L. 2007.** BI.Terminología Básica 1a Parte. [Online] 2007. <http://www.gavitar.biz/index.php/bi/bi-terminologia-1/>.

**Visual Paradigm. 2012.** Visual Paradigm . [Online] 2012. <http://www.visualparadigm.com>.

**Zapata, Rafael Matamoros. 2009.** *Implantación en una empresa de un sistema Business Intelligence SaaS / On Demand a través de la plataforma LITEBI*. Universidad Politécnica de Valencia. 2009. Proyecto fin de carrera.

**Bibliografía consultada**

**Adamson, Cristopher. 2006. Mastering Data Warehouse Aggregates. EUA : Wiley Publishing Inc, 2006.**

**Bernabeu, Ricardo Dario. 2009. DATA WAREHOUSING: Investigación y Sistematización de Conceptos - HEFESTO: Metodología propia para la Construcción de un Data Warehouse. [PDF] Córdoba, Argentina : s.n., 21 de Abril de 2009. Vol. 1.1.**

**Clemente, Gerardo. 2008. Un Sistema para el Mantenimiento de Almacenes de Datos. 2008.**

**Curto, Josep. 2007. Information Management Data Warehousing .Data Warehouse y Datamart. [En línea] 2007. [Citado el: 12 de Marzo de 2012.]**  
<http://informationmanagement.wordpress.com/2007/10/07/data-warehousing-data-warehouse-y-datamart/>.

**Díaz, S.O. 2009. Inteligencia de Negocios. 2009.**

**González Hernández, Yanisbel. 2011. PROPUESTA DE METODOLOGIA PARA EL DASARROLLO DE ALMACENES DE DATOS EN DATEC. La Habana : s.n., 2011.**

**Guijarro Riera, Angélica María y Amalia Alexandra, Guijarro Riera. 2010. Diseño e Implementación de un Sistema de Control del Proceso de Crédito y Cobranza por medio de Indicadores de Gestión para una Sociedad Financiera ubicada en la ciudad de Guayaquil. Guayaquil : s.n., 2010.**

**Hechavarría, Yunió Ricardo. 2010.. Análisis, Diseño e Implementación del mercado de datos indicadores relacionados con la ciencia e innovación tecnológica para la Oficina Nacional de Estadísticas. 2010.**

**Hernández Sampieri, Roberto. 2006. Metodología de la Investigación. 4ta edición. s.l. : McGraw-Hill Interamericana, 2006.**

**Hernández Sampieri, Roberto, Fernández Collado, Carlos y Baptista Lucio, Pilar. 2006. Metodología de la Investigación. 4ta. s.l. : Mc Graw-Hill Interamericana, 2006.**

**Herrera, C. 2007. Adictos al trabajo. [En línea] 30 de octubre de 2007.**  
<http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=datawarehouse4..>

- Hobbs, Lilian y otros. 2005.** *Oracle Data Base 10g Data Warehousing*. EUA : ELSEVIER Digital Press, 2005.
- Humantumba, Rayner. 2007.** *DataMart paso a paso*. [www.ruedatecnologica.com] 2007.
- Inmon, William. 2005.** *Building the Data Warehouse*. EUA : Wiley Publishing Inc, 2005.
- Jorge, Yanisleidis. 2011.** *Sistema de Información de Gobierno. Mercado de datos para las series históricas del Sector externo*. 2011.
- Kimball, Ralph. 2009.** *The Data Warehouse Toolkit*. 2009.
- Kimball, Ralph y Ross, Margv. 2002.** *The Data Warehouse Toolkit*. EUA : Wiley Publishing Inc, 2002.
- Langefors, Börje. 1973.** *Teoría de los sistemas de información*. 1973.
- Las Líneas de evolución de las bases de datos.* **Piattini Velthuis, Mario. 2000.** 2000.
- Lopez, Fabian. 2011.** *Sistema de Información de Gobierno. Mercado de datos para las áreas de Cultura y Deporte*. 2011.
- Lora, Heidi Carmona. 2007.** *Modelo y Diseño de un DWH usando vistas materializadas*. 2007.
- Navarro, Alberto Límia, y otros. METODOLOGÍA PARA EL DESARROLLO DE SOLUCIONES DE ALMACENES DE DATOS E INTELIGENCIA DE NEGOCIO EN CENTALAD.** Universidad de las Ciencias Informáticas. La Habana : s.n.
- Ortiz, Julio.E. 2009.** *Diseño e implementación de un Mercado de Datos para la Oficina Nacional de Estadísticas*. 2009.
- Perez, Santiago. 2006.** *Apoyo para la toma de decisiones*. 2006.  
<http://www.edutecne.utn.edu.ar/sistemas-informacion/Data%20Mining-DataWarehouse.pdf>.
- Pompa, Aylenis. 2011.** *Sistema de Información de Gobierno. Mercado de datos Educación*. 2011.
- Ponniah, Paulraj. 2001.** *Data Warehousing Fundamentals*. EUA : Wiley Publishing Inc, 2001.
- Pressman, R.S. 2007.** *Ingeniería del Software. Un enfoque práctico*. 6ta Edición. 2007. Vol. I.

- Rivadera, R.Gustavo.** *La metodología de Kimball para el diseño de almacenes de datos(Data warehouses).*
- Rodríguez, Laritza. 2011.** *Sistema de Información de Gobierno. Mercado de datos Tecnologías de la información.* 2011.
- Rodríguez, Marianela. 2010.** *Mercado de datos Indicadores relacionados con el Comercio Agropecuario para la Oficina Nacional de Estadísticas.* 2010.
- San Juan, Yelena I y Recio, Yudith. 2011.** *Mercado de Datos Agricultura, Ganadería y Silvicultura para el Sistema de Información del Gobierno .* 2011.
- Torres, L. 2007.** BI.Terminología Básica 1a Parte. [En línea] 2007.  
<http://www.gavitar.biz/index.php/bi/bi-terminologia-1/>.
- Vega, Liliam, Rojas, Luis y Placeres, Cecilia. 2008.** *LA INTELIGENCIA DE NEGOCIO.SU IMPLEMENTACIÓN MEDIANTE LA PLATAFORMA PENTAHO.* Habana : s.n., 2008.
- Velasco, R.H. 2009.** Almacenes de datos. [En línea] 2009.  
<http://www.rhernando.net/modules/tutorials/doc/dw.html>.
- Yepes López, José. 1991.** *El desarrollo de los sistemas de información y documentación.* 1991.
- Zapata, Rafael Matamoros. 2009.** *Implantación en una empresa de un sistema Business Intelligence SaaS / On Demand a través de la plataforma LITEBI.* Universidad Politécnica de Valencia. 2009. Proyecto fin de carrera.
- Zepeda Sánchez, Leopoldo Zenaido. 2008.** *Metodología para el Diseño Conceptual de Almacenes de Datos.* Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia. Valencia : s.n., 2008.

### Glosario de términos

**HTTP:** Siglas de *Hyper Text Transfer Protocol* (protocolo de transferencia de hipertexto). Se utiliza para la comunicación web, permitiendo que los usuarios finales accedan a la aplicación e interactúen con la misma.

**HTML:** Siglas de *HyperText Markup Language* (lenguaje de marcado de hipertexto). Es el lenguaje de marcado predominante para la elaboración de páginas web, usado para describir la estructura y el contenido.

**Servlets:** Son programas que se ejecutan en el servidor, realizando la función de una capa intermedia entre una petición proveniente de un navegador web u otro cliente HTTP y las aplicaciones del servidor.

**Servlets/JSP:** Tecnología similar a los *servlets* que ofrece una conveniente forma de agregar contenido dinámico a un archivo HTML por utilizar código escrito en Java dentro del archivo utilizando etiquetas especiales que son procesadas por el servidor web antes de enviarlos al cliente.

**ETL:** Es el proceso que organiza el flujo de los datos entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un almacén de datos, reformatearlos, limpiarlos y cargarlos en otra base de datos.

**Sumarizaciones:** También llamada agregación, es una forma de mostrar los datos de una manera más resumida, permitiendo precisamente calcular valores agregados, que no son los datos directos registrados, sino datos derivados de ellos.

**Cubo:** Son subconjuntos de datos de un almacén de datos, organizado y sumariado dentro de una estructura multidimensional.

**Precompilado:** En almacenamiento de datos, se refiere a los valores que son calculados con anterioridad y almacenados físicamente para ser accedidos cuando se requiera.

**Particionamiento:** Reparto o división de un todo en varias partes.

**Base de datos (BD):** Serie de datos organizados y relacionados entre sí, los cuales son recolectados y explotados por los sistemas de información de una empresa o negocio en particular.

**Homogeneidad:** Utilizado para referirse a un sistema informático, quiere expresar que todos sus componentes comparten una misma arquitectura.

**SGBD:** Software que dirige y controla todas las gestiones que se realizan en las bases de datos. Permite almacenar y posteriormente acceder a los datos de forma rápida y estructurada.

**Robustez:** Capacidad de los sistemas software de reaccionar adecuadamente ante condiciones excepcionales.

**Usabilidad:** Se refiere a la capacidad de un software de ser comprendido, aprendido, usado y ser atractivo para el usuario, en condiciones específicas de uso.

**Portabilidad:** Cualidad de un programa, de un sistema operativo o de un lenguaje que se puede ejecutar en diversos tipos de ordenador.

**Sistemas de información:** Conjunto de elementos interrelacionados con el propósito de prestar atención a las demandas de información de una organización, para elevar el nivel de conocimientos que permitan un mejor apoyo a la toma de decisiones.

**Integridad:** A groso modo, la integridad es el mantener con exactitud la información tal cual fue generada, sin ser manipulada o alterada por personas o procesos no autorizados.

**Consistencia:** Propiedad que asegura que una transacción no romperá con la integridad de una base de datos, pues respeta todas las reglas y directrices de ésta.

**Flexibilidad:** Relativa a la variedad de posibilidades con las que el usuario y el sistema pueden intercambiar información. También abarca la posibilidad de diálogo, la multiplicidad de vías para realizar la tarea, similitud con tareas anteriores y la

optimización entre el usuario y el sistema.

**SQL:** Siglas de *Structured Query Language* (lenguaje estructurado de consulta). Es un lenguaje de alto nivel, no procedural, normalizado, que permite la consulta y actualización de los datos de bases de datos relacionales.

**Código abierto:** Término con el que se conoce al software distribuido y desarrollado libremente. Tiene un punto de vista orientado a los beneficios prácticos de compartir el código.

**XML:** Siglas de *Extensible Markup Language* (lenguaje de marcas generalizado). Se trata de un lenguaje utilizado para estructurar la información en cualquier documento que contenga texto.

**MDX:** Es el lenguaje a través del cual podemos explotar la información que reside en los motores OLAP y satisfacer las consultas analíticas.

**Cardinalidad:** Es el número de entidades por la cual otra entidad puede asociarse mediante una relación. Puede ser de uno a uno, uno a muchos, muchos a uno y muchos a muchos.

## Anexos

### Anexo 1. Código de disparador que ejecuta la función para actualizar la vista materializada.

```
-- Trigger: tgr_cdc_actualizar_vm_demanda_1a on mart_agricultura.hech_demanda
-- DROP TRIGGER tgr_cdc_actualizar_vm_demanda_1a ON mart_agricultura.hech_demanda;

CREATE TRIGGER tgr_cdc_actualizar_vm_demanda_1a
  AFTER INSERT OR UPDATE OR DELETE
  ON mart_agricultura.hech_demanda
  FOR EACH ROW
  EXECUTE PROCEDURE mart_agricultura.fc_tgr_vm_demanda_1a(E'\\\\\\x5c783563373833353633333'
```

### Anexo 2. Código de función que actualiza la vista materializada

```
begin
if (TG_OP='INSERT') then
INSERT INTO mart_agricultura.vm_hech_demanda_1a(
  vm_hech_demanda_1a_id, dim_destino_id, dim_empresa_id, dim_producto_id,
  dim_temporal_diario_id, "real", plan, anno)

  VALUES (NEW.hech_demanda_id, NEW.dim_destino_id, NEW.dim_empresa_id,
  NEW.dim_producto_id, NEW.dim_temporal_diario_id, NEW."real", NEW.plan, NEW.anno);
end if;

if (TG_OP='UPDATE') then
DELETE from mart_agricultura.vm_hech_demanda_1a;

INSERT INTO mart_agricultura.vm_hech_demanda_1a(
  vm_hech_demanda_1a_id, dim_destino_id, dim_empresa_id, dim_producto_id,
  dim_temporal_diario_id, "real", plan, anno)

SELECT hech_demanda_id, dim_destino_id, dim_empresa_id, dim_producto_id,
dim_temporal_diario_id, "real", plan, anno FROM mart_agricultura.hech_demanda
WHERE anno = (SELECT max(anno) AS max FROM mart_agricultura.hech_demanda);

end if;

if (TG_OP='DELETE') then
delete from mart_agricultura.vm_hech_demanda_1a where
  vm_hech_demanda_1a_id = OLD.hech_demanda_id;
end if;

return null;

end;
```

### Anexo 3. Código de disparador que ejecuta la función que elimina los registros de la vista materializada.

```
-- Trigger: tgr_snc_eliminar_vm_hech_demanda_1a on mart_agricultura.vm_hech_demanda_1a
-- DROP TRIGGER tgr_snc_eliminar_vm_hech_demanda_1a ON mart_agricultura.vm_hech_demanda_1a;

CREATE TRIGGER tgr_snc_eliminar_vm_hech_demanda_1a
AFTER INSERT OR UPDATE
ON mart_agricultura.vm_hech_demanda_1a
FOR EACH ROW
EXECUTE PROCEDURE mart_agricultura.fc_scn_vm_demanda_1a(E'\\\\\\x5c7835633738303000');
```

### Anexo 4. Código de función que elimina los registros de la vista materializada que no pertenezcan al último año.

```
declare
begin

if (TG_OP='INSERT' OR TG_OP='UPDATE') then
delete from mart_agricultura.vm_hech_demanda_1a
where anno < (SELECT max(anno)
FROM mart_agricultura.vm_hech_demanda_1a);

end if;

return null;|

end;
```

### Anexo 5. Entrevista efectuada a los trabajadores de la Delegación Provincial de la Agricultura de Artemisa.

1. ¿Cuáles son los objetivos de su organización? ¿Qué se está tratando de resolver? ¿Cuáles son las prioridades de la entidad que deben resolverse?
2. ¿Existen categorías de la información dentro del negocio? ¿Es posible dividir las actividades que se realizan en el negocio por sector, rama, departamento, dirección según el tipo de información que se maneja?
3. ¿Cuál es la información o los tipos de reportes que actualmente se puede obtener en el negocio? ¿Cuáles son las claves del negocio en riesgo actuales? ¿Qué previenes al conocer los objetivos del negocio? ¿Cuál es el impacto de este

conocimiento en la organización? ¿Existen aspectos externos que puedan afectar el curso de la información?

4. ¿Con qué frecuencia se obtiene información?
5. ¿Qué cantidad de información en general se maneja?
6. ¿Cómo usted sabe que está haciendo bien su trabajo? ¿Cómo usted identifica los problemas o cómo los soluciona?
7. ¿Cuáles son las métricas para medir el éxito? ¿Cómo obtienes los indicadores (métricas) para monitorear los procesos?
8. ¿Qué criterios se tienen en cuenta para analizar los indicadores?
9. ¿Qué análisis le gustaría realizar sobre esos indicadores?
10. ¿Existen cortes de información o peticiones más comunes? ¿Qué tipo de análisis a la medida suele realizar con más frecuencia?
11. ¿Cómo le gustaría que se le presentaran los datos en la pantalla? (gráficos (tipos), tablas, indicadores en la parte superior).