



Facultad Regional “Mártires de Artemisa”

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

**Título:** “Mercado de Datos que contribuya a la toma de decisiones en la Dirección General de Economía de la Administración Provincial de Artemisa”.

**Autores:** Nayla Raquel Lago Silveira

Yasmany Arzuaga Cabrera

**Tutores:** Ing. Idalmys Maza Capote

Ing. Orelvi Gázquez Martínez

Lic. Omar Mar Cornelio

**Cotutor:** Lic. Lázaro Rolando Rodríguez Fernández

Artemisa, Cuba

Junio de 2012

*“La igualdad en la riqueza debe consistir en que ningún ciudadano sea tan opulento que pueda comprar a otro, ni ninguno tan pobre que se vea precisado a venderse”.*

*Jean-Jacques Rousseau*

# DECLARACIÓN DE AUTORÍA

---

Declaramos que somos los únicos autores de este trabajo y autorizamos a la Facultad Regional “Mártires de Artemisa” de la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

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

---

Nayla Raquel Lago Silveira  
Autor

---

Yasmany Arzuaga Cabrera  
Autor

---

Ing. Idalmys Maza Capote  
Tutor

---

Lic. Omar Mar Cornelio  
Tutor

---

Ing. Orelvi Gázquez Martínez  
Tutor

---

Lic. Lázaro Rolando Rodríguez Fernández  
Cotutor

## DEDICATORIA

---

A la memoria de Moisés Silveira y a Mercedes Llizo, mis abues, por haber creído siempre en mí, por darme la oportunidad de estar en este mundo, por su amor, cariño y preocupación incondicionales.

Nayla

A mis abuelas: NENA, que dios la tenga en la gloria (sé que está orgullosa de mi); y BLANCA para que se enorgullezca cada día más de mí.

A los que más sacrificios han hecho en la vida por lograr educar y guiar a un hombre de bien; a los que no les ha importado ningún obstáculo que les ha impuesto la vida para lograr tener y mantener una familia unida, decente; a los que me han dado todo el amor del mundo, a los que no se cansan de velar por mí: MIS PADRES. Por ellos soy lo que soy ahora. Los amo con mi vida.

A lo mejor que me han dado mis padres en la vida, mi HERMANA; para que siga mis pasos, para que logre aumentar los conocimientos no solo académicos, sino personales de la vida; en general: para que sea una mujer mejor en todos los aspectos de la vida, te amo con todo mi corazón.

A la mujer que me ha dado todo el amor, cariño y comprensión del mundo, mi novia.

A las personas que pensaron en algún momento de esta vida que no llegaría hasta donde estoy ahora.

A mi familia en general.

Yasmany

## AGRADECIMIENTOS

---

A mi mamá, por todos los años de preocupaciones, carreras, discusiones, alegrías, llantos y fiestas relacionados con mis estudios y mi vida. Tú has hecho de mi lo que soy, y todos mis logros son tuyos también.

A mi abuelo Moisés, que donde quiera que esté, sé que está orgulloso, por haber sido mi ejemplo, por ser la fuerza que me ayudó a seguir en los momentos que quise desfallecer.

A mi abuela Mercedes, por estar ahí todos los días de mi vida, por apoyarme, escucharme y malcriarme como sólo ella puede, por enseñarme cosas que sólo los abuelos saben.

A mi primo Carlitos, por ser mi hermano, ese que no tuve siendo niña, por crecer y vivir a mi lado cada día de nuestra infancia y compartirlo todo.

A mis hermanos Samuel y Paloma por ser mi inspiración para seguir y superarme todos los días.

A mi tía y madrina Raquelita, por toda la preocupación, la ayuda, las carreras y las malcriadeces a lo largo de mi vida.

A mis tíos Luisito, Sergito y Nestico, siempre preocupados por mí, por la escuela, la tesis y lo que se me ha ocurrido hacer en los veinticuatro años de mi vida.

A mi papá, Andrés Laffita, por dejarme ser su hija más chiquita y la más malcriada, por la familia tan linda en la que tuve la oportunidad de crecer y vivir tantos años.

A toda mi familia, por tantas preocupaciones y malcriadeces, por la ayuda y el amor que me han brindado, por estar interesados por mi futuro y mis decisiones, porque siempre me he sentido y me sentiré orgullosa de donde vengo.

A Odlanier, por estar estos cinco años ahí, este camino hubiera sido muy raro sin ti.

## AGRADECIMIENTOS

---

A Anita, Irma, Juan Carlos y Moremy, por hacerme sentir siempre parte de la familia.

A mis hermanas en la Mini, esas que escogí o me escogieron como amiga, Aylín, Aniuska, la Chiqui, Jany, Cristy y Reglita por estar en los momentos más difíciles y en los más alegres.

A mis amistades, por hacer de esta experiencia una de las mejores de mi vida, en especial a Chema, Eric, Greydis, Mily, Nagyara, Yeicy y Yaima.

A mis compañeras de cuarto, por soportar mi carácter y aguantar cada día los berrinches con las cosas de la tesis, de la escuela y demás rollos.

A mi compañero de tesis, por soportar mis majaderías y exquisiteces.

A mi tutora desde tercer año, Idalmys, por pelearnos, enseñarnos y ayudarnos para que el trabajo en el proyecto y la tesis fueran de calidad.

A Mayrín, por la paciencia y la dedicación a lo largo de este curso, cuando queríamos volvernos locos.

A aquellos que nos aconsejaron, nos dieron su opinión, nos ayudaron y orientaron: Orelvi, Margarita, Yuray, Lianne.

A los profesores que a lo largo de mi vida estudiantil me han brindado su cariño y sus conocimientos y han influido para que llegue hasta aquí, en especial a Tatiana y a Juan Félix.

A las personas que se han interesado en saber cómo iba la tesis, a los que de una forma u otra ayudaron a terminarla, a los que me han ayudado y apoyado en los cinco años de universidad y a lo largo de mi vida, a todos: **¡MUCHAS GRACIAS!**

Nayla

## AGRADECIMIENTOS

---

A mis Padres primero que todo, porque son para mí, los mejores padres del MUNDO, gracias por todos los sacrificios, gracias por decirme cada día “Puedes contar conmigo para lo que sea”, “Pa’ lante, que aquí estamos nosotros”, gracias por guiarme, por cuidar tanto de mí, por aconsejarme, por luchar conmigo, gracias por quererme tanto, gracias por darme la VIDA, gracias por enseñarme a corregir mis errores.

A mi hermanita linda, gracias por darme tanto AMOR, gracias por EXISTIR, por preocuparte tanto por mí.

A mi abuela Nena, por darme su amor y cariño (a su forma), por impulsarme, por sus consejos de la vida, por sus experiencias en la vida que buenas o malas, me enseñaron mucho. A mi abuela BLANCA por seguirme dando tanto cariño y amor incondicional en esta vida, por cuidarme, por preocuparse tanto de mí.

A mis abuelos, Juan y Néstor, a mis tíos (as) Magaly, Jorge Luis, Osmel, Yuya, Mercedes, Miriam, Tito, Nica, Juan, Edelmiro, Gustavo, Carlos, por sus consejos, por sus impulsos a seguir y lograr esta meta en la vida: ‘Graduarme de Ingeniero’, por su amor y cariño hacia mí.

A mis primos: Yoany, Yaine, Yasser, Aliesky, Yoel, Osniel, Yaimel, Héctor, muchas gracias por todo su apoyo, sus consejos, sus preocupaciones.

A mi novia, por darme todo el amor del mundo, su cariño es lo que me hace más fuerte cada día.

Al chiquitico de los repasos y dudas académicas, como si fuera mi hermano: El Perdo.

A mi compañera de tesis, muchas gracias por ser tan detallista y perfeccionista en cuanto al trabajo, son virtudes fundamentales para realizar un trabajo de tesis; gracias por coger tantas luchas.

## AGRADECIMIENTOS

---

A nuestra tutora Idalmys Maza por no dejar de guiarme, por aconsejarme, por ponerme el dedo con las tareas, por ser tan exigente con el trabajo, por ayudarnos tanto con las dudas, por estar ahí siempre para nosotros; por coger tantas luchas, por buscar tiempo de donde no había y por poner todo su empeño en tratar de que todo saliera perfecto, muchas gracias.

A Omar Mar Cornelio muchas gracias por estar ahí siempre para mí, por aconsejarme, por guiarme, gracias por seguir luchando conmigo hace 7 años.

A René Pulgarón por su preocupación, sus consejos, su impulso, sus viajes a la escuela jajajajaja.

A Leodan, Henry, Jennifer, por incitarme a seguir, gracias por su amistad (me ayudó muchísimo).

A Orelvi Gázquez muchísimas gracias por encaminarnos, por aconsejarnos.

A Mayrin Ramos gracias particularmente por haber sido mi primera tutora, por haberme aconsejado tanto, por enseñarme la importancia de las responsabilidades, gracias por guiarme y preocuparse tanto por mí.

A Yenisley, Ledesma, Andy, Jany, gracias por permitir preguntarles cualquier duda de la tesis, gracias por su amistad.

A Lianne gracias por toda su ayuda.

A todas las personas que de una forma u otra colocaron su granito de arena y me ayudaron de cualquier manera.

Yasmany



La presente investigación se enmarca en el tema de los almacenes y mercados de datos, y su empleo para mejorar los análisis de información y la toma de decisiones a partir de estos análisis. Contiene un estudio, en la cual se detallan las metodologías, herramientas, tendencias actuales y las mejores prácticas para el desarrollo de este tipo de soluciones. Se efectúa un levantamiento de información al cliente, para conocer sus necesidades y pasar al proceso de definición del diseño de la solución de forma consecuente. Esta investigación toma como referencia la metodología formulada por Ralph Kimball.

Como resultado se obtiene la estructura del modelo dimensional que comprende: las dimensiones, jerarquías, tablas de hechos, vistas materializadas y medidas necesarias para proceder con los análisis y la toma de decisiones. Se precisan las reglas del negocio utilizadas y se detalla el proceso de carga de los datos desde las distintas fuentes de la Dirección General de Economía de la Administración Provincial de Artemisa al mercado de datos. De igual manera, la solución incluye las estrategias de seguridad, respaldo y recuperación de los datos, así como el proceso de validación. Como característica especial, incluye la utilización de herramientas libres en el proceso de implementación.

**Palabras claves:** Almacenes de datos, Mercados de datos, Administración Provincial de Artemisa, Dirección General de Economía.

Introducción.....	1
Capítulo 1: Fundamentación teórica de los almacenes de datos.....	9
1.1 Almacenes de datos.....	9
1.2 Mercados de datos.....	11
1.3 Almacén de datos vs. Mercado de datos.....	12
1.4 Etapas de desarrollo de un Almacén de Datos.....	13
1.5 Estado actual de los Almacenes de Datos.....	20
1.6 Metodologías para el desarrollo de un almacén de datos.....	23
1.7 Herramientas para el desarrollo del mercado de datos.....	25
Capítulo 2: Análisis y diseño del Mercado de Datos de la Dirección General de Economía de la Administración Provincial de Artemisa.....	33
2.1 Análisis.....	33
2.2 Diseño.....	37
Capítulo 3: Implementación del Mercado de Datos de la Dirección General de Economía de la Administración Provincial de Artemisa.....	39
3.1 Estándar de Código.....	39
3.2 Implementación de los procesos de ETL.....	41
3.3 Empleo de Vistas Materializadas.....	45
3.4 Implementación de los Subsistemas de Visualización.....	46
Capítulo 4: Validación del Mercado de Datos de la Dirección General de Economía de la Administración Provincial de Artemisa.....	52
4.1 Validación y prueba.....	52
4.2 Aporte social y económico.....	54
Conclusiones.....	57
Recomendaciones.....	58
Bibliografía.....	59
Referencias Bibliográficas.....	61

## **Introducción**

La vertiginosa evolución del mundo actual, impulsada en gran medida por el avance de las ciencias de la información<sup>1</sup>, ha contribuido a elevar el desarrollo científico, tecnológico y económico en una cadena de retroalimentación ineludible.

Con la gran acumulación de información se llegó a la conclusión de que esta podría tener un fin útil, al reflejar la mayoría de las operaciones durante los llamados ciclos de negocios. Por tanto, la manipulación y el almacenamiento eficiente de grandes volúmenes de datos, se ha tornado una exigencia común. La toma de decisiones permite valorar cada proceso que origina cierto cúmulo de información y considerarlo profundamente para elegir el mejor camino a seguir, pues de la adecuada selección de alternativas depende el éxito de cualquier organización.

El análisis de la información histórica es de suma importancia para poder conocer las tendencias, características de demanda, requerimientos legales, y barreras comerciales a las que se enfrentan todas las organizaciones en los días actuales, y constituye un amplio marco en el que desarrollar la toma efectiva de decisiones. Su objetividad, identificando y evaluando las diversas alternativas en términos de la meta deseada, contribuye a la elección de la mejor opción, es decir, tomar la decisión más factible.

Los gobiernos no están exentos de esta situación, todo lo contrario, este entorno exige que eleven su capacidad para administrar información y crear conocimiento; de otra manera enfrentarán una situación de abundancia de información y déficit de conocimiento. Para afrontar este reto, se implantan sistemas de información, capaces de hacer su almacenamiento y análisis más ágil, eficiente y certero, dentro de estos sistemas se destacan los almacenes y mercados de datos.

---

<sup>1</sup> Ciencias de la información: Surgieron como campo de investigación después de la Segunda Guerra Mundial. Se encargan de investigar el procesado de información y la ingeniería de los sistemas de información. A lo largo de los años, y según su etapa de desarrollo, han estudiado la estructura, algoritmos, comportamiento e interacciones de los sistemas naturales y artificiales que guardan, procesan, acceden a y comunican información. Manifiestan el valor de la información como conocimiento y la importancia de su diseminación.

Los almacenes de datos, y dentro de ellos más específicamente los mercados, surgieron con el objetivo de hacer consultable la información que se tiene, pues organizan y orientan los datos desde la perspectiva del usuario final. Con su aparición se abrió una puerta para proporcionar una respuesta factible al manejo y análisis eficiente de la información.

El principal beneficio que aporta a una empresa, es que elimina aquellos datos que obstaculizan la labor de análisis y entrega la información que se requiere en la forma más apropiada, facilitando así el proceso de gestión. Además, su objetivo fundamental es asistir al ejecutivo en el entendimiento del pasado y contar con los elementos para la planeación del futuro a corto, mediano y largo plazo.

Los almacenes y mercados de datos comenzaron su evolución a fines de la década 1980 y principios de la década de 1990. Ha llevado casi 20 años la estandarización y madurez de esta tecnología. En la actualidad se puede demostrar que los avances alcanzados confirman que ya es una tecnología madura, estable y soluciona la problemática presentada, lo que no significa que no continúe en constante evolución.

En Cuba se ha identificado muy tempranamente la utilidad y necesidad de dominar e insertar en la práctica social las Tecnologías de la Información y las Comunicaciones. Con la informatización de la sociedad y dentro de las empresas, ha crecido la capacidad de generación y almacenamiento de la información, que no puede analizarse por los métodos tradicionales existentes, mientras mayor es la capacidad para almacenar datos, mayor es la incapacidad para extraer información realmente útil de estos. Se aboga entonces por el uso de estos recursos en orden de lograr una mayor productividad y calidad.

Al calor de la Batalla de Ideas y como un paso de singular significación en el proceso de informatización de la sociedad cubana, se fundó la Universidad de las Ciencias Informáticas.

En ella se encuentra el Centro de Tecnologías y Gestión de Datos (DATEC), que tiene la misión de proveer soluciones integrales y de consultoría, relacionadas con las tecnologías de bases de datos y análisis de información. Además, organiza la producción en líneas de trabajo como por ejemplo el Desarrollo de Almacenes de Datos e Inteligencia de Negocio. Se puede afirmar entonces que la Universidad ha sido pionera en este sentido, desarrollando proyectos en los que se construyen mercados de datos, que conforman almacenes, para varias empresas del país.

Dentro de sus facultades se encuentra la Facultad Regional “Mártires de Artemisa”, que ha jugado un rol protagónico, ejecutando proyectos de gran alcance como los desarrollados para la Oficina Nacional de Estadísticas e Información (ONEI), el Ministerio de Comercio Exterior (MINCEX) y la Administración Provincial de Artemisa. En la concepción de estos dos últimos proyectos se encuentra actualmente el Departamento de Almacenes de Datos de la Facultad.

La Administración Provincial de Artemisa, creada a partir de la constitución de la nueva provincia en el año 2010, tiene como función principal la de preparar y proponer la política integral del Estado. Además, se encarga de dirigirla y ejecutarla, así como de coordinar y controlar su cumplimiento sobre la base de las estrategias de desarrollo del país. Está compuesta por 32 direcciones, entre ellas la Dirección General de Economía. Esta cuenta con 3 Direcciones, 1 Departamento Independiente y 1 Sección Independiente.

Actualmente la Dirección General de Economía tiene las siguientes insuficiencias:

- Se almacena la información en ficheros de texto y documentos impresos.
- Existen pérdidas y duplicados de información.
- La información es accesible a cualquier personal no autorizado o no definido por la administración.

Esto trae como consecuencia la falta de seguridad en los datos que se manejan y el análisis incorrecto de la información, provocando que el proceso para la toma de

decisiones sea lento y engorroso.

Esta situación problemática conduce a formular el siguiente **Problema de investigación**: ¿Cómo contribuir al análisis y seguridad de la información para agilizar la toma de decisiones en la Dirección General de Economía de la Administración Provincial de Artemisa?

Teniendo en cuenta el problema científico planteado se define como **Objeto de estudio**: Almacenes de Datos.

Se puede decir entonces que este problema se enmarca en el **Campo de acción**: Mercados de Datos para el apoyo a la toma de decisiones.

En virtud de ofrecer una solución para el problema anteriormente descrito se propone el **Objetivo general**: Desarrollar un Mercado de Datos que contribuya al análisis y seguridad de la información, para agilizar la toma de decisiones en la Dirección General de Economía de la Administración Provincial de Artemisa.

Del objetivo general expuesto se derivan los siguientes **Objetivos Específicos**:

- Investigar el estado del arte de los sistemas dedicados a la toma de decisiones.
- Fundamentar las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.
- Realizar el análisis y diseño del Mercado de Datos de la Dirección General de Economía de la Administración Provincial de Artemisa.
- Implementar el Mercado de Datos de la Dirección General de Economía de la Administración Provincial de Artemisa.
- Validar el Mercado de Datos de la Dirección General de Economía de la Administración Provincial de Artemisa a través de Casos de Prueba.

La **idea a defender** expone que, con la realización del análisis, diseño e implementación del mercado de datos para la Dirección General de Economía de la Administración Provincial de Artemisa, se logra el correcto análisis de la información y se tiene una mayor seguridad de esta, lo que beneficia al proceso de toma de decisiones.

Para dar cumplimiento a cada uno de los objetivos específicos se plantean:

## **Tareas de Investigación**

1. Fundamentación teórica sobre los almacenes y mercados de datos para la toma de decisiones.
2. Caracterización del estado actual del mercado de datos de la Dirección General de Economía de la Administración Provincial de Artemisa.
3. Determinación de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.
4. Validación del mercado de datos para la Dirección General de Economía de la Administración Provincial de Artemisa.

En función de comprender y analizar de manera adecuada el problema planteado y la solución que se propone a partir de este, se emplean:

## **Métodos de Investigación Científicos**

### **Métodos teóricos**

- ✓ **Histórico-Lógico:** A través de este método se analizó el desenvolvimiento de los mercados de datos en el tiempo y la aplicación de las mejores prácticas en su desarrollo con herramientas de código abierto. También se establecieron las posibles fuentes de datos de la Dirección General de Economía de la Administración Provincial de Artemisa para realizar el proceso de análisis.
- ✓ **Analítico-Sintético:** Permitted seleccionar, de las metodologías existentes para el desarrollo de mercados de datos, la más apropiada según las características de la Dirección General de Economía.
- ✓ **Modelado:** Mediante su empleo se pudo definir la arquitectura del sistema a

implementar para establecer las bases de desarrollo del mercado de datos.

## **Métodos empíricos**

- ✓ **Análisis documental:** Empleado para sistematizar la bibliografía y los documentos metodológicos disponibles, posibilitando el acercamiento a buen número de referentes teóricos, en su mayoría de la actual década.
- ✓ **Observación:** Realizada en las visitas a la Administración Provincial, observando los procesos que tienen lugar en la Dirección General de Economía.
- ✓ **Entrevista:** Conversación previamente definida con los Especialistas y Jefes de Direcciones, Secciones y Departamentos con el objetivo de obtener la mayor cantidad de información posible sobre la forma en la que se trabaja.
- ✓ **Encuesta:** Conjunto de preguntas elaboradas y recogidas con el que se obtuvo información sobre los procesos que tienen lugar en la Dirección General de Economía de la Administración Provincial de Artemisa y la manera en la que se llevan a cabo.

La **población** la constituyen los 51 trabajadores de la Dirección General de Economía de la Administración Provincial de Artemisa, se toma como **muestra** para la presente investigación 12 jefes de departamentos, lo cual constituye el 24%.

## **Declaración de variables**

**Variable independiente:** Mercado de Datos.

**Variable dependiente:** Proceso para la toma de decisiones.

La presente investigación se lleva a cabo por la **necesidad** que tiene la Dirección General de Economía de la Administración Provincial de Artemisa de facilitar el análisis y la seguridad de su información con el objetivo de agilizar la toma de decisiones. A nivel mundial existen almacenes y mercados de datos que se encargan de hacer más factible la gestión de los gobiernos, estos, a pesar de tener similitudes con la solución a implementar, contienen marcadas diferencias, pues su



estructuración es completamente distinta.

Además, hay que tener en cuenta que la mayoría estas soluciones son desarrolladas en sistemas económicos capitalistas, bajo licencias de software propietario. Por todo lo expresado sus experiencias nos son aplicables al entorno y la realidad del problema que se desea resolver, de esta manera se evidencia la novedad y el aporte del actual trabajo en el marco de la problemática a solucionar, adecuado a las condiciones de Cuba para implementar software.

A través de este estudio se espera como **resultado** obtener el Mercado de Datos para el apoyo a la toma de decisiones en la Dirección General de Economía de la Administración Provincial de Artemisa.

## **Estructura del Documento**

### **Capítulo 1: Fundamentación teórica de los almacenes de datos.**

En este capítulo se abordará toda la fundamentación teórica sobre los principales conceptos, metodologías y herramientas para el desarrollo del mercado de datos, que permitirá a la Dirección General de Economía apoyar el proceso de toma de decisiones.

### **Capítulo 2: Análisis y diseño del Mercado de Datos de la Dirección General de Economía de la Administración Provincial de Artemisa.**

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 de la Dirección General de Economía de la Administración Provincial de Artemisa.**

En este capítulo se realizarán todas las transformaciones necesarias como Extraer, Transformar y Cargar (ETL), proceso que deja toda la información lista para la fase siguiente, Inteligencia del Negocio (BI); donde se realizará la implementación del modelo de datos, los cubos OLAP, los reportes candidatos y la política de seguridad de los usuarios.

## **Capítulo 4: Validación del Mercado de Datos de la Dirección General de Economía de la Administración Provincial de Artemisa.**

En este capítulo se aplicarán los casos de prueba para validar el mercado de datos.

## **Capítulo 1: Fundamentación teórica de los almacenes de datos.**

En este capítulo se fundamentan teóricamente los principales conceptos, metodologías y herramientas que permitirán el desarrollo del mercado de datos a implementar, en aras de apoyar el proceso de toma de decisiones en la Dirección General de Economía de la Administración Provincial de Artemisa.

Se aborda el proceso de evolución de los almacenes y mercados de datos tanto en el mundo, como en el caso particular de Cuba valorando sus ventajas y desventajas. Todo esto conducirá a evaluar el modo en que se puede alcanzar, con esta propuesta, las metas y objetivos proyectados con el desarrollo de la presente investigación.

### **1.1 Almacenes de datos.**

Sobre los almacenes de datos existen varias definiciones ofrecidas por grandes especialistas y profesionales del tema, por ejemplo:

Bill Inmon<sup>2</sup>, considerado el padre del almacén de datos lo define como *“un conjunto de datos integrados, orientados a una materia, que varía con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones de la administración.”* [1]

Ralph Kimball<sup>3</sup>, otro gran especialista sobre estas colecciones de datos, reconocido mundialmente y escritor del famoso libro *“The Data Warehouse Toolkit (El juego de herramientas del almacén de datos)”* afirma que es *“una copia de los datos de la transacción estructurados específicamente para la pregunta y el análisis.”* [2]

Susan Osterfeldt, quien también incursionó en este ámbito, ofrece su opinión propia

---

<sup>2</sup> William Harvey Inmon (1945): experto reconocido mundialmente. Creador de la Corporate Information Factory. (Fábrica de Información Corporativa).

<sup>3</sup> Ralph Kimball: Doctor en Filosofía, ha sido uno de los mayores visionarios en la industria del Almacén de Datos desde 1982, actualmente, reconocido conferencista, consultante y profesor.

acerca de lo que es un almacén de datos, diciendo: “*Yo considero al almacén de datos como algo que provee dos beneficios empresariales reales como: integración y acceso a datos. Un almacén de datos elimina una gran cantidad de datos inútiles y no deseados, como también el procesamiento desde el ambiente operacional clásico.*” [3]

Luego de realizar la sistematización a la bibliografía accesible se considera que los elementos abordados en cada una de las definiciones, aunque difieren en algunos aspectos, giran todas sobre el mismo eje central y contribuyen al sustento teórico de la presente investigación pues conceptualizan al almacén de datos en correspondencia con los estándares que se tienen en cuenta para su desarrollo.

Los autores de esta investigación asumen el criterio de Ralph Kimball, que un **Almacén de datos** es una colección de datos interesantes para la empresa, organizados, integrados y orientados al negocio. Utilizándolo se puede brindar al usuario una visión histórica a la hora de la toma de decisiones.

### **Características de un almacén de datos.**

- Integrado: como los datos almacenados provienen de fuentes diferentes deben integrarse en una estructura consistente que elimine las inconsistencias existentes en los datos.
- Temático: los datos se organizan por los principales temas de la organización, esto es lo que lo diferencia de los sistemas operacionales que se basan en los procesos cotidianos de la organización.
- Histórico: el almacén contiene grandes volúmenes de información histórica sobre las actividades de la organización y va variando en el tiempo, recibiendo periódicamente nuevos datos.
- No volátil: la información que existe en un almacén de datos es permanente, los datos no cambian, esto garantiza que las consultas que se realicen sobre el almacén de datos siempre producirá el mismo resultado. [4]

## **Ventajas de un almacén de datos.**

- Integrar datos históricos sobre la actividad de la organización (o negocio) en un único repositorio.
- Analizar los datos del negocio desde la perspectiva de su evolución en el tiempo.
- Prever tendencias de evolución del negocio.
- Identificar nuevas oportunidades de negocio y tomar decisiones estratégicas.
- Reducir los costes materiales y humanos en la toma de decisiones. [4]

## **Desventajas de un almacén de datos.**

- Riesgo de fracaso en la construcción del sistema, al subestimar los costes de captura y preparación de los datos.
- Riesgo de fracaso en la construcción del sistema por cambios continuos en los requisitos de los usuarios.
- Problemas con la privacidad de los datos. [4]

## **1.2 Mercados de datos.**

Un **Mercado de datos** está orientado a temas específicos o aplicaciones específicas y contiene datos de sólo una línea del negocio. El ámbito de la información que contiene es más pequeño que en un almacén, los datos se obtienen de un menor número de fuentes y su tiempo de desarrollo es menor. [5]

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

- Se centran en los requisitos de los usuarios asociados a un departamento o área de negocio concretos.
- No contienen datos operacionales detallados.
- Son más sencillos a la hora de utilizarlos y comprender sus datos, debido a que la cantidad de información que contienen es mucho menor que en los Almacenes de Datos. [6]

## **Ventajas de un mercado de datos.**

- Mayor rapidez de consulta.
- Consultas SQL<sup>4</sup> y/o MDX<sup>5</sup> sencillas.
- Validación directa de la información.
- Facilidad para la historización de los datos.
- Permite manejar información confidencial.
- Refleja rápidamente sus beneficios y cualidades.
- Reduce la demanda del depósito de datos. [7]

## **Desventajas de un mercado de datos.**

- Se pierde capacidad de procesamiento debido al crecimiento de los datos.
- Los usuarios necesitan acceder a varios mercados de datos.
- Compleja administración del mercado de datos.
- Dificultad para construir debido al corto plazo de desarrollo. [7]

## **1.3 Almacén de datos vs. Mercado de datos.**

En 1998 Bill Inmon declaró “El elemento más importante que enfrentan los directores de tecnologías de la información en este año, es si construir primero los Almacenes de Datos”. [8] Esta afirmación todavía está vigente.

Existen interrogantes que resultan imprescindibles evaluar antes de decidir realizar la construcción de un Almacén de Datos: [8]

- ✓ ¿La aproximación se realizará de arriba hacia abajo (top-down) o de abajo hacia arriba (bottomup)?
- ✓ ¿Empresarial o departamental?
- ✓ ¿Cuál primero, el almacén de datos o el mercado de datos?
- ✓ ¿Construir un piloto o directamente el almacén de datos completo?

---

4 SQL: Structured Query Language(Lenguaje de Consulta Estructurado): Lenguaje que permite hacer consultas a la información almacenada en la base de datos.

5 Las expresiones multidimensionales (MDX es el acrónimo de Multi-Dimensional Expressions) es un lenguaje de consulta para bases de datos multidimensionales sobre cubos.

✓ ¿Mercados de Datos dependientes o independientes?

Las respuestas de estas preguntas llevan a una planificación profunda de lo que realmente se va a desarrollar. Cada respuesta es crucial para escoger la decisión correcta sobre si utilizar un Almacén de Datos o utilizar un Mercado de Datos.

En su libro “Data Warehouse Fundamentals (Fundamentos de los Almacenes de Datos)” Poulraj Ponniah determina un conjunto de elementos que dan claridad sobre las diferencias existentes entre ambos conceptos.

Estas diferencias se muestran en la tabla siguiente: [8]

<b>Almacén de Datos</b>	<b>Mercado de Datos</b>
Corporativo o red empresarial. Es la unión de todos los mercados de datos. Los datos son recibidos desde el área de procesamiento. Consultas sobre la presentación de recursos. Estructura para vista corporativa de los datos.	Departamental. Un simple proceso del negocio. Unión en forma de estrella (hechos y dimensiones). Tecnología óptima para el acceso a los datos y el análisis. Estructura para adaptarse a la vista de los datos departamentales.

**Tabla 1: Comparación entre Almacén de Datos y Mercado de Datos.**

Tomando como base lo expresado por Ponniah, los Mercados de Datos, como estructura, son un Almacén de Datos, pero reducido a un departamento específico sirviendo como fuente de análisis del tema que concierne a dicho departamento. La unión de todos los mercados de datos de la organización es lo que conformaría la vista global de los datos, es decir, el Almacén de Datos.

### **1.4 Etapas de desarrollo de un Almacén de Datos.**

Para el proceso de desarrollo del almacén de datos se cuenta con tres etapas

fundamentales:

- ✓ Análisis y diseño.
- ✓ Extracción, transformación y carga (ETL).
- ✓ Inteligencia de negocio (BI).

## **Análisis y diseño.**

Toda fase inicial de un proyecto de gran alcance como son los almacenes de datos, debe contar con bases sólidas que le permitan avanzar hacia las otras fases de desarrollo con un mayor nivel de seguridad y organización. Se deben identificar los posibles riesgos.

Debe contar además, con un plan de acción efectivo que minimice los riesgos de fracaso y tener conocimiento de un grupo de elementos y herramientas que intervendrán en el desarrollo de dicho proyecto. Esto, unido al suficiente dominio del funcionamiento del proyecto y tener claridad de cuáles son los objetivos que se pretenden alcanzar, son las bases para comenzar el desarrollo de un Almacén de Datos.

Para un Almacén de Datos es prioritario reconocer las necesidades analíticas que tiene una organización y establecer los objetivos que se persiguen al desarrollar uno. Este es el primer paso a seguir, en el análisis y la comprensión del nuevo entorno, el entorno analítico. [4]

Es en esta etapa cuando es loable realizar un estudio de factibilidad para analizar lo viable del proyecto, se reflejará el ciclo completo de Almacén de Datos con sus costos reales que serviría de referencia para determinar cuáles serían los gastos en una empresa mediana que emprenda un proyecto de esta envergadura.

En cuanto al diseño del almacén, una arquitectura de este tipo establece el marco de trabajo, estándares y procedimientos a seguir para la construcción de un Almacén de Datos a nivel empresarial. El objetivo de las actividades de la arquitectura es simple, integrar al almacén de datos las necesidades de información empresarial. [4] Es Ralph Kimball quien propone la arquitectura mejor estructurada



y de mejor comprensión para el diseño de un almacén.

Los resultados más importantes que se pueden obtener de la aplicación de esta arquitectura son:

- El modelo de datos fuente.
- El modelo de datos conceptual del Almacén de Datos.
- Arquitectura tecnológica del Almacén de Datos. [4]

En cuanto al modelo de datos del Almacén de Datos se pueden encontrar dos formas a seguir:

Modelado relacional: se basa en el uso de Diagramas Entidad Relación (ERD).

**Diagramas Entidad Relación (ERD)**: Lenguaje para el modelado de los datos dentro de un sistema de información. Es muy útil para la creación de esquemas conceptuales de bases de datos. No son recomendables para el diseño de un Almacén de Datos por no garantizar una óptima recuperación de la información almacenada. Las consultas muy complejas que abarcan una gran cantidad de registros y tablas hacen de estos modelos, algo muy difícil de entender para los usuarios.

Modelado dimensional: se basa en el uso de Esquemas Estrellas o de Esquemas de Copo de Nieve.

**Esquema de Estrella**: Es la arquitectura de almacén de datos más simple. En este diseño, la tabla de hechos está rodeada por dimensiones y juntos forman una estructura que permite implementar mecanismos básicos utilizables con herramientas de consultas OLAP. Implementa un diseño lógico relacional de manera que las tablas de hechos representan la tercera forma normal (3FN) y las dimensiones representan la segunda forma normal (2FN). [9]

**Esquemas de Copo de Nieve**: Es una variedad más compleja del esquema estrella. Lo que distingue a la arquitectura en copo de nieve del esquema de

estrella, es que las tablas de dimensiones en este modelo representan relaciones normalizadas en tercera forma normal (3NF) y forman parte de un modelo relacional de base de datos. Se usa cuando las tablas de dimensiones son muy extensas y complejas haciéndose difícil representar los datos en un esquema estrella. [10]

### Modelo multidimensional

Para el diseño de un Almacén de Datos, varios autores han seguido diferentes conceptos para abordar el modelado de un almacén, entre los que se encuentran, esquema en estrella, esquema en copo de nieve, esquema en constelación, entre otros. Independientemente de la metodología seguida, todos ellos utilizan un modelo multidimensional de datos.

El modelado multidimensional es: modelos de datos como conjuntos de medidas descritas por dimensiones.

- Adecuado para resumir y organizar datos (por ejemplo hojas de cálculo).
- Enfocado para trabajar sobre datos de tipo numérico.
- Más fácil de visualizar y entender que el modelado E/R. [11]

### Características del modelo multidimensional

La estructura básica de un almacén de datos para el modelo multidimensional está definida por dos elementos: esquemas y tablas.

Como cualquier base de datos relacional, el Almacén de Datos se compone de tablas; de dos tipos de tablas, específicamente en el modelo multidimensional.

Tablas de Hechos: tabla central en un esquema dimensional y en ella se guardan las medidas numéricas del negocio, por ejemplo: número de libros vendidos.

Tablas dimensionales: son las tablas que se alimentan de las tablas de hechos, por lo que contienen el detalle de los valores que se encuentran asociados a estas. [12]

La colección de tablas dentro del almacén se conoce como esquema y por lo general caen dentro de dos categorías básicas: esquema en estrellas y esquemas

en copo de nieve.

## Componentes del modelo multidimensional

El modelo multidimensional está compuesto por hechos, cubos, medidas, dimensiones, jerarquías, niveles, y atributos, donde todos sus datos contienen cierto nivel de detalle denominado granularidad.

**Hecho:** Representa un evento o actividad específica, tiene dimensiones y medidas. Un ítem de negocio, una transacción o un evento que tiene significancia para el negocio. Corresponden a una colección de ítems de datos y datos de contexto. Son aquellos datos que residen en una tabla de hechos y que son utilizados para crear indicadores, a través de sumalizaciones preestablecidas. Un hecho debe estar relacionado, al menos, con una dimensión: “El tiempo”. [13]

**Cubo:** principal objeto del proceso analítico en línea (OLAP<sup>6</sup>, Online Analytic Processing), una tecnología que proporciona rápido acceso a los datos de un almacén de datos, conjunto de datos que normalmente se construyen a partir de un subconjunto de un almacén de datos y se organiza y resume en una estructura multidimensional definida por un conjunto de dimensiones y medidas. [14]

**Medida:** Es un atributo numérico de un hecho que representa el performance o comportamiento del negocio relativo a la dimensión. Las medidas representan los valores que son analizados. Deben ser numéricas, porque estos valores son las bases de las cuales el usuario puede realizar cálculos. Cruzan todas las dimensiones en todos los niveles. Si la medida es no numérica debemos codificarla a un valor numérico y cuando tengamos que exponerla, decodificarla para mostrarla con el valor original. [13]

**Dimensión:** Característica de un hecho que permite su análisis posterior, en el proceso de toma de decisiones. Determina el contexto del hecho (quién participó,

---

<sup>6</sup> Online Analytic Processing/Proceso Analítico en Línea. Es un proceso analítico de datos online que permite al usuario seleccionar y extraer la información desde diferentes puntos de vista. Se traduce en un conjunto de técnicas y estándares definidos por organismos de estandarización. Pueden ser definidos algunos tipos de OLAP dependiendo de las técnicas que se utilicen a la hora de obtener los datos y la forma en la que están estructurados, etcétera. Dentro de estos tipos están las herramientas que permitirán realizar análisis de los datos.

cuándo y dónde pasó, y su tipo). Es una entidad de negocios respecto a la cual se deben calcular las métricas (clientes, productos, tiempo). Las dimensiones tienden a ser discretas y jerárquicas <país, región, departamento, provincia, distrito>. Es una colección de miembros, unidades o individuos del mismo tipo que permiten categorizar un hecho. [13]

**Jerarquía:** Una jerarquía representa una relación lógica entre los datos de una dimensión. Se presentan al interior de una dimensión. Pueden existir varios niveles (dos o más). Relación “1-n” o “padre-hijo” entre atributos consecutivos de un nivel superior y uno inferior. Es una clasificación u ordenación de abstracciones. [13]

**Nivel:** es un elemento de la jerarquía de dimensiones. Los niveles describen la jerarquía desde el nivel de datos más altos (más resumido) hasta el más bajo (más detallado). [14]

**Atributo:** provee información adicional acerca de los datos. [14]

**Granularidad:** La granularidad es el nivel de detalle en que se almacena la información. A mayor nivel de detalle, mayor posibilidad analítica, pues los datos podrán ser resumidos o sumariados. Los datos con granularidad fina (nivel de detalle) podrán ser resumidos hasta obtener una granularidad media o gruesa. No sucede lo mismo en sentido contrario. [13]

## **Extracción, transformación y carga (ETL).**

En aquellos escenarios donde exista un almacén de datos, este necesita ser alimentado de datos ya existentes en otras fuentes, y para ello aparecen los procesos de ETL.

- Extracción: este es el primer paso para la obtención de la información de las diferentes fuentes hacia el almacén de datos. Como los datos provienen de fuentes distintas, generalmente se encuentran en formatos distintos y organizados a la manera de cada organización, la extracción deja los datos en un formato listo para la transformación.
- Transformación: una vez que la información es extraída, el proceso de transformación se encarga de preparar los datos de la manera adecuada

para integrarlos en el almacén de datos. Para ello este proceso se compone de algunas actividades como: limpieza de datos, integración de formato, integración de datos.

- **Carga:** una vez que los datos han sido extraídos de las diferentes fuentes y transformados, se procede a cargar el almacén de datos, después de la carga inicial se procede a mantener el almacén, actualizándolo periódicamente. [14]

## **Inteligencia de negocio (BI).**

La gran mayoría de las organizaciones poseen abundante información, pero carecen de métodos y herramientas que permitan su uso eficiente. La inteligencia de negocios es la clave para la solución de dicho problema, pues posee vías para mejorar el proceso de toma de decisiones, convirtiéndolo en una ventaja competitiva.

La Inteligencia de Negocios o Business Intelligence (BI) se puede definir como el proceso de analizar los bienes o datos acumulados en la empresa y extraer una cierta inteligencia o conocimiento de ellos.

El proceso de Inteligencia de Negocio apoya a los usuarios que toman decisiones con la información correcta o que ellos necesitan, en el momento correcto y lugar correcto, lo que le facilita tomar mejores decisiones de negocios. Esto se puede llevar a cabo gracias a la visualización de los reportes que se realizan a la información, como resultado final del proceso. Ejemplos de estos son:

**Vistas de Análisis:** Reportes iterativos y dinámicos. En las vistas de análisis se realiza la navegación dimensional sobre los cubos desde la plataforma de inteligencia de negocio, visualizándose los resultados de las consultas y mostrándose al usuario en formato web por el cliente Jpivot.

**Reportes ad-hoc:** Reportes que se crean en el momento sin una planificación de antemano mediante el envío de una solicitud o consulta, para obtener información específica. El resultado muestra la información más actual en el formato que se

especifique. Son informes especiales que se crean cuando alguien pregunta por ellos, no en un punto específico.

Una de las herramientas que permite mejor visualización de los datos a los usuarios son las OLAP a través de reportes y gráficos. Estas herramientas se clasifican en tres grupos fundamentales:

**Herramientas MOLAP:** Multidimensional On-line Analytical Process/Proceso Analítico Multidimensional en Línea. Son herramientas que acceden a datos que no están almacenados en registros de tablas, sino que almacenan los datos en arreglos de varias dimensiones, llamados cubos. Estos cubos utilizan índices para optimizar el acceso a los datos.

**Herramientas ROLAP:** Relational On-line Analytical Process/Proceso Analítico Relacional en Línea. Son herramientas OLAP que crean vistas multidimensionales extrayendo los datos de bases de datos SQL ordinarias, es decir, relacionales. Estas herramientas simulan los datos multidimensionales usando sofisticadas técnicas de indexación, cachés, metadatos, etcétera.

**Herramientas HOLAP:** Hybrid On-line Analytical Process/Proceso Analítico Híbrido en Línea. Permiten un análisis híbrido de la información, es decir que une lo mejor de los dos tipos anteriores. El análisis HOLAP ayuda a reducir costes de hardware pues se necesita menos disco que en las bases de datos relacionales. Además, la respuesta de las consultas sobre las bases de datos multidimensionales son más rápidas que sobre las relacionales. Como aspecto negativo, los datos multidimensionales deben ser cargados antes de ser consultados y refrescados cuando se actualizan los datos de la organización.

## 1.5 Estado actual de los Almacenes de Datos.

En América Latina existen empresas como Visa, Arcor, Wallmart, Procter & Gamble, Whirpol, Tv Azteca, Baxter, GNP, Warner Lambert y Sabre, que han venido

incorporando el uso de los almacenes de datos para la toma de decisiones a nivel gerencial. También se puede mencionar a American Stores (Estados Unidos), Canadian Tyre (Canadá), Owens Corning Glass (Estados Unidos), Karsten Ping Golf Clubs, que han obtenido grandes avances en este sentido.

Ejemplos de estos mismos avances en Europa los constituyen las siguientes instituciones: Carrefour (España), Smith Books (Gran Bretaña), BonPreu (España), SAR Group (España), Great Universal (Gran Bretaña), Corte Inglés (Francia), Cortefiel (Francia), Eroski (Francia), Supermercado Casino (Francia), Migros Genossenschaftsbund (Suiza), Otto Versand (Alemania), Helene Curtis (España).

Además, grandes trasnacionales como, Coca Cola, Walt Disney, Nike, Maybelline, Adidas, 3M, Bosh Siemens, se han incorporado a la utilización de los almacenes de datos para la realización de estudios de mercado y de inteligencia de negocio.

Entre los gobiernos que utilizan almacenes de datos para agilizar sus gestiones se encuentran los de:

- Estados Unidos.
- México.
- España.

Estos y otra cifra considerable de gobiernos, actualmente usan las tecnologías de la información y el conocimiento en sus procesos internos y en la entrega de los productos y servicios del Estado, tanto a los ciudadanos como a la industria. Muchas de las tecnologías involucradas y sus implementaciones son las mismas o similares a aquellas correspondientes al sector privado, mientras que otras son específicas o únicas en relación a las necesidades del gobierno.

Algunos ejemplos de gobiernos que utilizan portales digitales y mercados de datos, para hacer más efectiva su gestión son:

- DIME - Dirección de Informática del Ministerio de Economía de Tucumán.
- Gobierno Fácil - Portal de Gobierno Digital de Costa Rica.

- Gobierno en Línea - Portal de Gobierno Digital de Colombia.

Guatemala, por las características de sus recursos naturales, tiene un enorme potencial para el desarrollo sostenible de su gente a partir del agua. Este país carecía de un Sistema Nacional Integrado de Información y Conocimiento del Agua, pero el Gobierno Central decidió implementar un Mercado de Datos como propuesta integradora para responder a la problemática.

También la Junta de Gobierno de Castilla y León en España implementó un Mercado de Datos, denominado SIE, que proporciona información en cantidad, tiempo y calidad adecuada para la toma de decisiones, facilitando la aplicación de técnicas de explotación y análisis de datos para la obtención de conclusiones acerca de tendencias y relaciones entre grandes volúmenes de información.

En Cuba ha venido asentándose una cultura tecnológica sobre el tema de los almacenes de datos. Aunque todavía faltan muchos aspectos por mejorar, ya se han visto algunos ejemplos que han dado pasos firmes dentro de esta rama. El más elocuente es el Almacén de Datos Comercial de la Corporación CIMEX. La que se dedica fundamentalmente a la exportación e importación de mercancías.

Forman parte de ella un conjunto de empresas que se encuentran enfocadas en diversos negocios, aquí se puede citar la red de Comercio Minorista y la Dirección de Logística, esta última dedicada al Comercio Mayorista. El mismo centra su atención en la actividad del comercio, principalmente en el manejo de inventario, permitiendo una gestión de compra-venta eficiente, con una finalidad de disminuir los costos, sin afectar al cliente, permitiendo prestaciones eficaces y con la calidad requerida, aumentando las ganancias o utilidades de las empresas.

En el XIII Concurso Nacional de Computación y en la Feria de Informática del 2002 se presentó un Almacén de Datos para Cubacel desarrollado sobre plataforma Oracle con grandes resultados obtenidos a partir de su implantación. Existen otras entidades como Unión CUPET y Copextel que actualmente cuentan con sus



respectivos almacenes.

Los almacenes de datos en Cuba se han introducido como una manera eficaz de ayudar al proceso de toma de decisiones en sectores importantes para el desarrollo. Algunos ejemplos que lo reflejan claramente son:

- Gestión Comercial de la Corporación CIMEX.
- Gestión Contable de la Empresa de Proyectos de Arquitectura e Ingeniería.
- Gestión en el Banco Nacional de Cuba.

En la bibliografía consultada durante la investigación no se encontró evidencia de que se haya empleado ningún almacén o mercado de datos para facilitar la gestión del gobierno en Cuba, a nivel de país, ni en ninguna de las provincias o municipios.

## **1.6 Metodologías para el desarrollo de un almacén de datos.**

El desarrollo de un software no es una tarea fácil, es por eso que surge la necesidad de usar una metodología. Las metodologías son un conjunto de procedimientos, técnicas y un soporte documental que imponen un proceso disciplinado sobre el desarrollo del software con el objetivo de hacerlo más predecible y eficiente.

Para la elaboración del mercado de datos se debe identificar una metodología que guíe el proceso de desarrollo a este tipo de solución. Para hacer uso de la misma en soluciones de almacenes de datos e inteligencia de negocios, solo se hará referencia a los criterios de Kimball e Inmon por marcar claramente su tendencia en cuanto al tema de almacenes de datos.

Aunque existen otras que se basan en estas propuestas, no siguen obligatoriamente una específica, sino que realizan una selección de lo mejor de cada una; algunas de ellas son: SQLBI, DM2, CRISP-DM, Hefesto y Metodología para el diseño conceptual de almacenes de datos.

## **Inmon.**

Esta metodología debe el nombre a su creador William H. Inmon. La visión de Inmon se basa en un enfoque descendente (top-down), propone construir primero el almacén de datos y a partir de este, los mercados de datos, plantea la creación de un repositorio de datos corporativo como fuente de información consolidada, persistente, histórica y de calidad.

Al ser construido descendentemente los mercados de datos se nutren del Almacén de Datos Corporativo, convirtiéndose en un complejo empresarial de bases de datos relacionales.

Inmon afirma que la creación de una base de datos relacional con una leve normalización necesita ser la base para los mercados de datos, por lo que no se crean los mercados de datos directamente desde los sistemas de procesamiento de transacciones en línea a través de un área de ensayo. En lugar de ello, se crean a partir de la arquitectura relacional de los datos corporativos.

## **Kimball.**

Ralph Kimball es reconocido a nivel mundial en el diseño de almacenes de datos y como creador del enfoque multidimensional. Se ha dedicado desde hace unos años al desarrollo de su metodología para que este concepto sea bien aplicado en las organizaciones y se asegure la calidad en el desarrollo de estos proyectos.

La propuesta de Kimball se basa en dividir el mundo de inteligencia de negocio entre los hechos y las dimensiones, esta es eficaz y conduce a una solución completa en un corto período de tiempo. Entre sus características principales está que su arquitectura es ascendente (bottom-up), plantea que se debe crear por cada departamento, un conjunto de Mercados de Datos independientes, orientados a los temas que estén relacionados con él. [15]

## **Selección de la metodología.**

Existen diversas metodologías para la construcción de un almacén de datos y un mercado de datos, pero para la Dirección General de Economía de la

Administración Provincial de Artemisa se requiere una aplicable a la situación. En este caso se seleccionó la metodología de Kimball por los siguientes elementos:

- Identifica la tabla de hechos y la tabla de dimensiones, lo cual agiliza el proceso de desarrollo y con ello la toma de decisiones.
- Propone la construcción de mercados de datos departamentales y después el almacén de datos, lo que tiene como ventaja, que la creación y la puesta en marcha de los mercados de datos se produce en un lapso de tiempo corto y después se valora si se construye o no el almacén de datos.
- Existe amplia documentación de la misma, así cualquier duda que exista puede ser atendida rápidamente.
- Es una metodología madura y reconocida por los usuarios dedicados al tema, además de tener bien definidas sus etapas, actividades, roles y artefactos. [16]

Para complemento de la misma y por las características de trabajo de la universidad, se tomó lo planteado por Leopoldo Zenaido Zepeda en su tesis de doctorado: incluir los casos de uso para guiar el proceso de desarrollo, y así lograr estar más alineados a las tendencias y normas de la universidad.

## **1.7 Herramientas para el desarrollo del mercado de datos.**

### **Visual Paradigm for UML.**

Visual Paradigm, es una herramienta UML profesional que soporta el ciclo de vida completo del desarrollo de software. El software de modelado UML ayuda a una más rápida construcción de aplicaciones de calidad, mejores y a un menor coste. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación.

Soporta un conjunto de lenguajes, tanto en la generación de código, como de ingeniería inversa sobre Java, C++, CORBA IDL, PHP, XML Schema, Ada y Python. Además, apoya la generación de código C#, VB. NET, código script de Flash,

Delphi, Perl, Ruby, entre otros. Ingeniería Inversa también apoya clase Java, .NET, .dll y .exe, JDBC.

Características:

- Producto de calidad.
- Soporta aplicaciones web.
- Las imágenes y reportes generados no son de muy buena calidad.
- Varios idiomas.
- Generación de código para Java y exportación como HTML.
- Fácil de instalar y actualizar.
- Compatibilidad entre ediciones.
- Diseño centrado en casos de uso y enfocado al negocio, que generan un software de mayor calidad.

Como herramienta de modelado se utilizará Visual Paradigm en su versión 6.4.

### **SQL Power Architect.**

Permite realizar perfiles de datos en bases de datos de origen y generar automáticamente los metadatos de ETL de forma rápida y muy sencilla. Con esta herramienta, el diseñador podrá abrir múltiples conexiones concurrentes a bases de datos, crear y explorar perfiles de datos fuente, arrastrar y soltar esquemas de datos, tablas y columnas dentro del modelo de datos, y confeccionar la base de datos resultante con su plantilla ETL asociada.

Hasta el diseño más complicado, la base de datos más grande o el modelo más inabarcable, podremos gestionarlo y manejarlo con Power Architect. [17]

Algunas de sus poderosas características son:

- Conexión con múltiples bases de datos al mismo tiempo.
- Compara modelos de bases de datos e identifica discrepancias.
- Dibuja y borra tablas y columnas en su visualizador.
- Genera Código SQL a partir del esquema en el visualizador.
- Genera Informes visuales.

- Realiza Ingeniería Inversa (recuperar modelos de una base de datos existente).
- Guarda el código y la estructura en el proyecto para poder trabajar remotamente. [18]

Como herramienta de diseño se trabajará con la versión 0.9.15 de SQL Power Architect.

## **PgAdmin.**

Es una herramienta de código abierto para la administración de bases de datos PostgreSQL y derivados, incluye:

- Interfaz administrativa gráfica.
- Herramienta de consulta SQL.
- Editor de código procedural.
- Agente de planificación SQL/shell/batch.
- Administración de Slony-I.

PgAdmin 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 hace simple la administración.

Está disponible en más de una docena de lenguajes y para varios sistemas operativos, incluyendo Microsoft Windows, Linux, FreeBSD, Mac OSX y Solaris. Resiste versiones de servidores 7.3 y superiores. Versiones anteriores a 7.3 deben usar el PgAdmin II. [19]

Como herramienta para la administración de bases de datos PostgreSQL se usará la versión 1.12.0 de PgAdmin.

## **PostgreSQL.**

Es un Sistema de Gestor de Base de Datos, que no pertenece a ninguna compañía, sino que es dirigido por una comunidad de desarrolladores, que se hace llamar “PostgreSQL Global Development Group” y organizaciones comerciales que

están a cargo de su perfeccionamiento. Tiene más de 15 años de desarrollo activo y se ha ganado la reputación de ser confiable y mantener la integridad de los datos. Existen compañías que aseguran haberlo empleado durante varios años y con altas tasas de actividad, sin experimentar problema alguno. El PostgreSQL se ejecuta en la mayoría de los sistemas operativos más utilizados en el mundo, incluido Linux, versiones de UNIX y, por supuesto, Microsoft Windows XP o superior.

Debido a sus características técnicas sobresalientes, el PostgreSQL se ha ganado la admiración y el respeto de sus usuarios, así como el reconocimiento de la industria, ha sido ganador del Linux New Media Award for Best Database System y tres veces ganador del The Linux Journal Editors' Choice Award for best DBMS (Corporate Executive Board, 2008). [20]

Las características que presenta PostgreSQL son:

- La cantidad máxima de bases de datos que permite es ilimitada.
- El tamaño máximo de las tablas es de 32 Tb.
- El tamaño máximo de registro es de 1.6 Tb.
- El máximo de tamaño del campo es de 1 Gb.
- El máximo de registros por tablas es ilimitado.
- El máximo de campos por tabla es 250 a 1600 en dependencia de los tipos de datos usados.
- El máximo de índices por tablas es ilimitado.

El código fuente de PostgreSQL está disponible bajo los más liberales términos de licencia de código abierto: la licencia BSD (Berkeley Software Distribution), por tanto pueden hacerse todas las modificaciones, mejoras o cambios que se estimen convenientes. [21]

Como Gestor de Base de Datos se empleará Postgre SQL, versión 9.0.

### **Pentaho Data Integration.**

Es una aplicación de software de código abierto cuyo objetivo es proporcionar a las

empresas capacidades como: generar reportes, efectuar análisis de los datos, hacer minería de datos, flujos de trabajo y en términos generales, aplicar Inteligencia de Negocios (Business Intelligence). [22]

Como herramienta para la limpieza, transformación y carga de datos se utilizará Pentaho Data Integration en su versión 4.1.0.

### **Pentaho Schema Workbench.**

Es una herramienta para la creación de cubos OLAP, un entorno visual para el desarrollo y prueba de dichos cubos. Con esta aplicación, se puede configurar una conexión JDBC como el modelo físico, para luego elaborar el esquema lógico de manera simple y efectiva. Para ello el entorno ofrece un editor de esquemas con la fuente de datos subyacente para su validación.

Permite la ejecución de consultas MDX contra el esquema y la base de datos y la navegación por la base de datos subyacente. Esta interfaz de diseño permite crear y probar esquemas visualmente. Estos archivos son los modelos de esquemas XML de metadatos que se crean en una estructura específica. Estos modelos XML se pueden considerar las estructuras de forma de cubo que utilizan hechos existentes y tablas de dimensiones. [23]

Como herramienta para el diseño de cubos OLAP se trabajará con la versión 3.2.1 de Pentaho Schema Workbench.

### **Pentaho BI-Server.**

Es una solución para ampliar el uso de la inteligencia de negocio. Se considera fácil solución de software que integra el poder de análisis e integración de datos para proporcionar puntos de vista que mejoran la toma de decisiones. Incluye, basados en roles, interfaces de auto-servicio para todo tipo de usuarios dentro de un marco bien definido y un punto centralizado de administración. Esto ayuda a las organizaciones a simplificar y acelerar el despliegue de inteligencia empresarial.

Beneficios:

- Información necesaria para decidir más rápido, y mejor. BI Server puede

aprovechar las inversiones anteriores, su integración con agilidad en los entornos existentes permite proporcionar la información necesaria para conducir el negocio hacia adelante lo más rápidamente posible. Esto, con el beneficio adicional de liberar recursos de las tediosas tareas para centrarse en proyectos más estratégicos, asegura que es una fuerza impulsora detrás del éxito de la organización.

- Aumenta la consistencia de los datos, permite más eficiencia a las decisiones impulsadas por datos y a la administración. A través de la gestión de datos centralizada, seguridad y metadatos compartidos, ofrece una representación y control coherente de la información, proporcionando resultados fiables que pueden fácilmente remontarse a la fuente. Esto alivia la tarea repetitiva de validar que los informes o resultados son correctos. Además, la integración con Microsoft Office permite retomar el control de especial uso de Excel sin obstaculizar el deseo de los usuarios de este entorno.

Como herramienta para la creación de reportes de datos se usará la versión 3.6.0 de BI-Server.

### **Pentaho Report Designer.**

Es un editor basado en Eclipse con prestaciones profesionales y capacidad de personalización de informes a las necesidades de los negocios destinado a desarrolladores. Esta herramienta está estructurada de forma que los desarrolladores puedan acceder a sus prestaciones de forma rápida. Incluye un editor de consultas para facilitar la confección de los datos que serán utilizados en un informe. [24]

Como herramienta adicional para la creación de reportes se empleará Pentaho Report Designer, versión 3.7.0.

### **Otros aspectos de la selección a tener en cuenta.**

La Administración Provincial de Artemisa, como la mayoría de las empresas cubanas, se encuentra en proceso de migración hacia la independencia



tecnológica. Esta política, implantada en la Facultad Regional “Mártires de Artemisa”, ha sido uno de los factores claves para la selección de las herramientas con las que se trabajará. Cada una de ellas cumple de manera eficiente con las funcionalidades que se demandan con gran calidad y también con la política de uso de software libre.

La herramienta seleccionada como Sistema Gestor de Base de Datos es el PostgreSQL. En este sentido la versión que se utilizará es lo suficientemente estable y segura.

Para el modelo de datos o diseño del almacén, la herramienta escogida es el Power Architect porque podrá abrir múltiples conexiones concurrentes a bases de datos, crear y explorar perfiles de datos fuente, arrastrar y soltar esquemas de datos, tablas y columnas, dentro del modelo de datos.

Con la herramienta Pentaho Data Integration se puede de manera muy simple tomar datos de una fuente (archivos locales y remotos, bases de datos, repositorios), aplicar un procesamiento a dichos datos (filtros, condiciones, cálculos, consultas), y almacenar los resultados en un destino (archivos, base de datos, repositorio). Posee alta compatibilidad de la herramienta ETL con el gestor PostgreSQL en los llamados Paquetes, donde se ofrecen servicios de importación, exportación, transporte, y transformación de datos; rendimiento eficiente y gran velocidad de respuesta de las consultas.

El servidor PostgreSQL posee un potente motor de recuperación de los datos y permite la optimización de las complejas consultas enviadas o preparadas desde el Pentaho Data Integration, así como capacidad favorable del Pentaho Data Integration para extraer y cargar datos utilizando el PostgreSQL.

Con estas herramientas se tiene la posibilidad no sólo de almacenar los datos al más atómico de los detalles, sino que también se puedan guardar solamente los agregados necesarios.

La herramienta utilizada para facilitar el trabajo dentro del Grupo de Inteligencia de Negocios es el BI-Server, manteniendo otras herramientas y lenguajes ya definidos, como Pentaho Schema Workbench (para la elaboración de los cubos OLAP) y MDX (lenguaje de consulta para cubos OLAP).

Lo planteado en este capítulo servirá de base para desarrollar un mercado de datos que apoye a la toma de decisiones en la Dirección General de Economía de la Administración Provincial de Artemisa, para ello se realizó una caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos y se seleccionaron las más ajustables al entorno y características de la entidad.

### **Capítulo 2: Análisis y diseño del Mercado de Datos de la Dirección General de Economía de la Administración Provincial de Artemisa.**

Este capítulo abordará aspectos relacionados con los requerimientos del sistema, en él se describe el diseño del diagrama de casos de uso del sistema, el modelo de datos y los subsistemas de integración y visualización. De manera general, aborda el resultado del análisis y el diseño del Mercado de Datos para la Dirección General de Economía de la Administración Provincial de Artemisa.

Se expondrá la manera en que se propone la construcción del mercado de datos, que no es una meta sencilla, pues se precisa obtener un sistema bien estructurado, robusto y capaz de responder a las necesidades de los usuarios finales, en aras de alcanzar las metas y objetivos proyectados con el desarrollo de la presente investigación.

#### **2.1 Análisis.**

En el proceso de análisis se identifican los requerimientos indispensables para lograr satisfacer las necesidades del cliente y realizar un producto con calidad. Este proceso se lleva a cabo a partir del diagnóstico obtenido a través de las entrevistas y encuestas efectuadas en las visitas a la Administración Provincial de Artemisa para levantar información (*Ver anexos 1 y 2*). En este proceso de análisis se registran los requisitos funcionales, los no funcionales y también los informativos determinados para el proyecto. Por otra parte, se definen los casos de uso del sistema que permiten determinar cómo se va a desarrollar el mercado de datos, agrupando cada uno de los requisitos.

#### **Requisitos.**

Los requisitos son aquellas necesidades que se definen y registran con el objetivo de permitir, a través del ciclo de desarrollo, la implementación con calidad de un buen producto, en aras de que el cliente quede satisfecho con su elaboración.

### Requisitos Informativos.

Los requisitos informativos son aquellos que responden a las necesidades de información del cliente, es decir, que deben mostrar la información que se solicita de la manera en que el cliente desee o requiera. En el mercado de datos para la Dirección General de Economía de la Administración Provincial de Artemisa se identificaron 50 Requisitos Informativos agrupados en 31 Casos de Uso. Un ejemplo de requisito informativo para este mercado es: Obtener el Presupuesto del año actual, el Presupuesto del mes actual, el Real del mes actual, la Diferencia entre el Real del mes actual y el Presupuesto del mes actual, el Por ciento Real del Mes y el Por ciento Real del Año, por indicadores y mes.

### Requisitos Funcionales.

Los requisitos funcionales son lo que se necesita que el sistema haga, es decir, definen el comportamiento interno del software y lo establecen según se requiera. En la presente investigación se delimitan 6 Requisitos Funcionales agrupados en 5 Casos de Uso que validan fundamentalmente la autenticación de usuarios y los procesos de extracción, transformación y carga de los datos. Ellos son:

1. Asegurar la extracción de los datos.
2. Asegurar la limpieza de los datos.
3. Asegurar la carga de los datos desde distintas fuentes.
4. Autenticar usuario.
5. Realizar los reportes de los indicadores de la Dirección General de Economía.
6. Realizar vistas de análisis de los indicadores de la Dirección General de Economía.

### Requisitos No Funcionales.

Los requisitos no funcionales son características o capacidades que el sistema debe tener o cumplir. Especifican criterios que pueden usarse para juzgar la operación de un sistema. Para la concepción de este mercado de datos se han

determinado 34 Requisitos no funcionales, que incluyen los de:

- Usabilidad
  - ➔ Dominar conocimientos básicos sobre Base de datos.
  - ➔ Tener conocimiento en análisis y tratamiento de la información.
- Fiabilidad
  - ➔ La integridad de los datos que se manejan es fundamental para el éxito de la solución de almacenamiento y análisis de datos.
- Disponibilidad
  - ➔ El sistema debe estar disponible 100% entre las 8:00 am y las 5:00 pm de lunes a viernes.
  - ➔ La carga de los datos se realizara después de las 7:00 pm.
- Eficiencia
  - ➔ El sistema debe permitir la abundancia de usuarios conectados simultáneamente sin que se afecte el tiempo de respuesta.
  - ➔ El tiempo de respuesta debe ser en tiempo real.
- Restricciones de diseño
  - ➔ Para la programación en el Mercado de Datos se utilizará PL/pgSQL como lenguaje dentro del Sistema Gestor de base de datos y para la implementación de los procesos de integración de datos el lenguaje JavaScript.
- Requisitos para la documentación de usuarios en línea y ayuda del sistema.
  - ➔ Se dispondrá de una Guía de ayuda sobre componentes del sistema, además de una gran información sobre herramientas ETL, presentadas en el sitio de la comunidad de Pentaho.
- Componentes Comprados
  - ➔ Las herramientas utilizadas están licenciada bajo GPL, son Open Source y totalmente gratis, ampliamente reutilizables y extensibles.
- Interfaz
  - ➔ Los reportes estadísticos deben contar con una interfaz simple que facilite la interacción usuario-aplicación.

- Interfaces de usuario
  - ➔ Las interfaces de salida no serán cargadas con información innecesaria.
- Interfaces Hardware
  - ➔ En el proceso de integración es necesario la utilización de una memoria mínima de 2 GB para el proceso de transformación.
- Interfaces Software
  - ➔ Debe existir un navegador asociado al sistema operativo que se escoja para lograr que las interfaces web de las tablas de salida puedan visualizarse.
- Interfaces de Comunicación
  - ➔ El sistema necesita estar conectado directamente a un dispositivo de red.
- Requisitos de Licencia
  - ➔ Las herramientas a utilizar se encuentran bajo licencia GPL.

### **Temas de análisis.**

La definición de tema de análisis es uno de los pasos más significativos en el proceso de análisis de un Mercado de Datos, pues enfoca el desarrollo de la solución en las áreas temáticas presentes en el negocio, facilitando así el cumplimiento de las metas trazadas. Fueron definidos 24 temas de análisis:

- Análisis de los indicadores de Eficiencia del Territorio.
- Análisis de los indicadores de Ejecución de Divisa.
- Análisis de los indicadores de Preparación para la Defensa.
- Análisis de los indicadores de las Inversiones.
- Análisis de los indicadores de las Importaciones.
- Análisis de los indicadores del Consumo Energético.
- Análisis de los indicadores de Disponibilidad.
- Análisis de los indicadores de Cuenta Propia.
- Análisis de los indicadores de Accidentes de Trabajo.
- Análisis de los indicadores de Salario Medio/Productividad.
- Análisis de los indicadores del Presupuesto de la Asistencia Social.

- Análisis de los indicadores del Presupuesto de la Seguridad Social.
- Análisis de los indicadores del Balance de la Seguridad Social.
- Análisis de los indicadores de las Improcedencias de la Seguridad Social.
- Análisis de los indicadores de las Estadísticas de la Seguridad Social.
- Análisis de los indicadores de los Pagos de la Seguridad Social.
- Análisis de los indicadores de Ingresos y Gastos.
- Análisis los indicadores del Programa Especial.
- Análisis de los indicadores del Presupuesto Central del Estado.
- Análisis de los indicadores de las Cuentas.
- Análisis de los indicadores de Multas.
- Análisis de los indicadores de Inmuebles.
- Análisis de los indicadores del consumo estatal.
- Análisis de los indicadores de eficiencia en las empresas.

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

Los diagramas de casos de uso del sistema representan gráficamente a los procesos y su interacción con los actores; el programador de ETL y los especialistas de cada departamento que interactúan con los casos de uso funcionales y los informativos. *Ver anexo 3: Diagrama de Casos de Uso del Sistema.*

### **2.2 Diseño.**

El diseño de un mercado de datos abarca la concepción de cada una de las dimensiones, tablas de hechos y medidas, así como las relaciones que tienen unas con otras según los procesos de negocio a tener en cuenta para su representación. Incluye asimismo, la manera en que se van a integrar y agrupar los elementos del mercado para la solución final.

### **Modelo de Datos.**

Para desarrollar el modelo dimensional de datos se debe tener en cuenta el análisis de la información a través de los reportes que se necesiten visualizar. Esto significa que, al modelar un mercado de datos, se debe tener como meta fundamental la información deseada en los reportes.

#### Dimensiones.

Las dimensiones suministran el contexto en el que se obtienen las medidas de un hecho, se utilizan para seleccionar y agrupar los datos en un nivel de detalle deseado. *Ver anexo 4: Listado de dimensiones.*

#### Tablas de hechos.

Las tablas de hechos contienen medidas numéricas, las cuales describen los resultados (rendimiento) del negocio y las claves de las dimensiones asociadas a ellas. *Ver anexo 5: Listado de tablas de hecho.*

#### Medidas.

Las medidas son los valores de datos numéricos que se analizan, son las variables de salida en el diseño, de ahí que representen lo contable que se necesita conocer. *Ver anexo 6: Listado de medidas por tabla de hecho.*

El modelo de datos diseñado para la solución constituye una manera de representar las relaciones entre los hechos y dimensiones. *Ver anexo 7: Modelo de datos para el Mercado.*

Lo expuesto a través de este capítulo servirá de base para pasar a implementar el mercado de datos que propone la presente investigación. Habiendo desarrollado el análisis y diseño de forma consecuente, caracterizando los elementos del modelo de datos dimensional, sus relaciones, formas de agrupación y visualización según los procesos que tienen lugar en la Dirección General de Economía de la Administración Provincial de Artemisa, se llevará a cabo la mencionada implementación del mercado.



### **Capítulo 3: Implementación del Mercado de Datos de la Dirección General de Economía de la Administración Provincial de Artemisa.**

A lo largo del presente capítulo se expondrá cómo se lleva a cabo la implementación del Mercado de Datos. Ilustrando los procesos de realización todas las transformaciones necesarias como Extraer, Transformar y Cargar (ETL), que deja toda la información lista para la fase siguiente, Inteligencia del Negocio (BI); donde se realizará la implementación del modelo de datos, los cubos OLAP, los reportes candidatos y la política de seguridad de los usuarios.

#### **3.1 Estándar de Código.**

La utilización del estándar de código, es muy importante en esta fase, ya que está presente en el diseño de las tablas físicas. A continuación se muestra el estándar que se definió para la Base de datos y para ETL:

##### **Base de Datos:**

- ✓ Base de Datos para el AD: [nombre]\_dwh.
- ✓ Base de Datos para el ODS: [nombre]\_ods.
- ✓ Catálogo Personalizado: [nombre]\_meta.
- ✓ Base de datos de Intercambio: [nombre]\_middle.

##### **Esquemas:**

- ✓ Objetos de Configuración: config\_[nombre esquema].
- ✓ Dimensiones Compartidas: dimensiones\_mincex.
- ✓ Esquemas de datos: mart\_[temática].

##### **Tablas:**

- ✓ Configuración: cf\_[nombre tabla].
- ✓ Dimensiones: dim\_[nombre].
- ✓ Hechos: hech\_[nombre]
- ✓ Vistas Materializadas: vm\_[nombre]\_#a

**Secuencias:**

- ✓ Secuencias dimensiones: seqDim\_[nombre\_campo]
- ✓ Secuencias configuración: seqConf\_[nombre\_campo]
- ✓ Secuencias hechos: seqHech\_[nombre\_campo]

**Funciones:**

- ✓ ETL: fc\_etl\_[nombre tabla]
- ✓ Sincronización: fc\_scn\_[nombre tabla]
- ✓ Trigger: fc\_tgr\_[nombre tabla]
- ✓ Particionamiento: fc\_prt\_[nombre tabla]
- ✓ Replicación: fc\_rp\_[nombre tabla]

**Vistas:** v\_[nombre]

**Índices:** idx\_[nombre tabla]-[nombre campo]

**Constraints:**

1. Claves Primarias: pk\_[tabla]\_id
2. Clave Foránea: fk\_[tablaFuente]\_id
3. Unicidad: unq\_[nombre]

**Trigger:**

1. Control de Cambio: tgr\_cdc\_[nombre]
2. Sincronización: tgr\_snc\_[nombre]
3. Réplica :tgr\_rp\_[nombre]

**Columnas de Tabla:**

Dimensión: [nombre dimensión]\_id

[nombre de nivel]\_codigo

[nombre de nivel]\_nombre

[nombre de nivel]\_descripcion

[nombre de nivel]\_[nombre atributo]

Hechos: [nombre hecho]\_id

## Transformaciones ETL:








Transformación: fc\_etl\_[nombre de tabla]












Trabajo: [nombre del hecho al que está asociado]

### 3.2 Implementación de los procesos de ETL.

El proceso de ETL es parte del ciclo de vida de implementación de un Mercado de Datos. Este se considera de importancia vital, debido a que es la forma en la que los datos se guardan. Los datos deberán ser extraídos de las fuentes, transformados, limpiados y cargados desde archivos Excel y/o una Base de Datos hacia el Mercado de Datos, por lo que se considera fundamental conocer cómo se realizarán cada una de estas actividades. La transformación de los datos se hará de acuerdo a las reglas del negocio definidas. Transformaciones como los cambios de formato, se encargan de asegurar la unicidad y estandarización de cada tipo de datos. Un ejemplo lo constituye la fecha, que deberá ser tratada como un solo tipo de dato y siempre de manera similar.

#### Principales componentes empleados en la implementación de los procesos de ETL.

1.  Entrada de tabla
2.  Entrada de Excel
3.  Entrada de Access
4.  Selecciona campos y renombrar campos
5.  Insertar y Actualizar Datos
6.  Salida tabla
7.  Salida a Access

8.  Salida a Excel
9.  Buscar datos
10.  Código Java
11.  Partir campo
12.  Insertar columna
13.  Calculadora
14.  Fila única
15.  Ordenar fila columna]
16.  Filtrar fila
17.  Agrupar por
18.  Eliminar

### Implementación de las transformaciones.

A continuación se expone un ejemplo de transformaciones realizadas:

#### ETL de la dimensión producto (dim\_producto)



*Figura 1: Proceso de carga de la dimensión producto.*

Este proceso tiene como objetivo extraer los datos del Excel que contiene la información de la dimensión producto y cargarla en la tabla dim\_producto del modelo de datos propuesto para implementar el Mercado de Datos.

**ETL de la dimensión de tiempo semestral (dim\_tiempo\_semestral)**

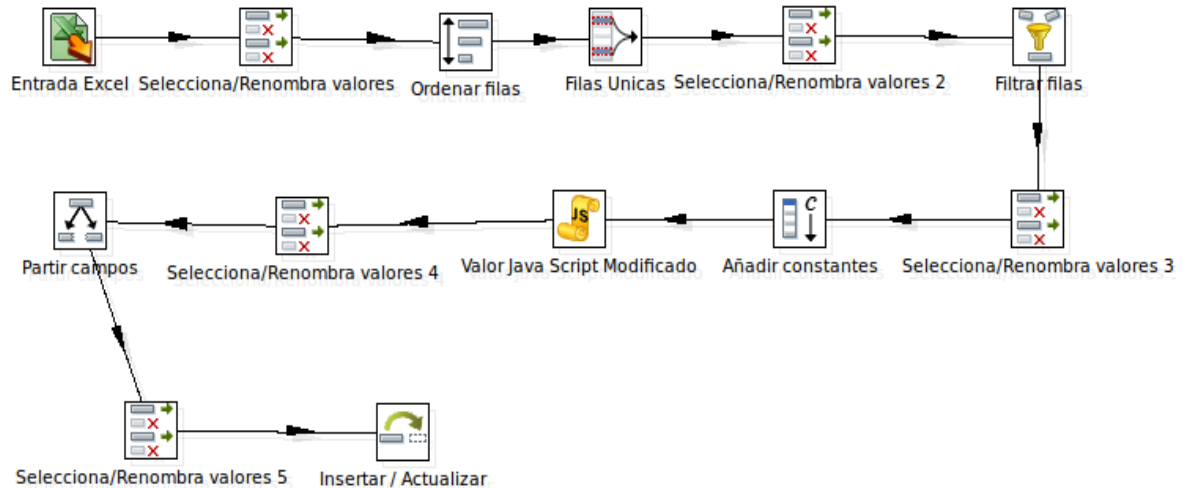
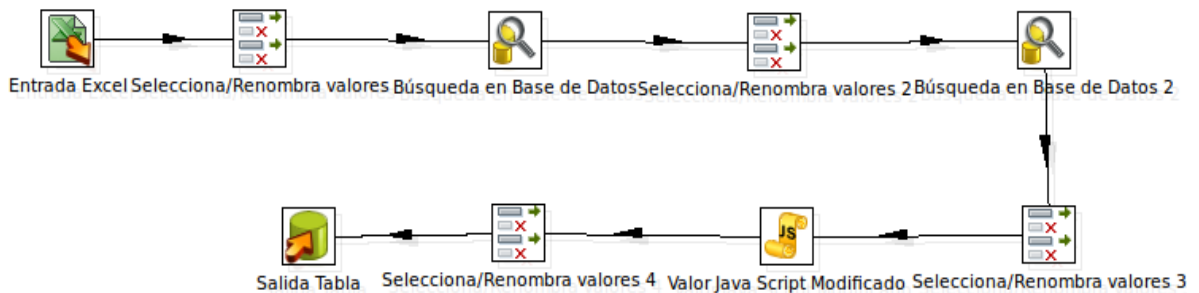


Figura 2: Proceso de carga de la dimensión de tiempo semestral.

Este proceso tiene como objetivo extraer los datos del Excel que contiene la información de la fecha de emisión del modelo de importaciones por productos para el Grupo Empresarial y cargarla en la tabla dim\_tiempo\_semestral, pues se emite mensualmente.

**ETL del hecho importaciones por productos para el Grupo Empresarial**



**(hech\_importaciones\_productos\_ge)**

Figura 3: Proceso de carga del hecho importaciones por productos para el Grupo Empresarial.

Este proceso tiene como objetivo extraer, transformar y cargar los datos fuentes a la tabla de hecho hech\_importaciones\_productos\_ge. Los datos para cargar esta tabla de hecho provienen del fichero xls. El sistema va comprobando cada atributo de la tabla fuente con la tabla destino y se buscan en las tablas de la base de datos relacionadas con el hecho los atributos que garanticen que existe la información necesaria para hacer la carga.

### Implementación de los trabajos (Jobs).

Los trabajos se encargan de ejecutar un grupo de transformaciones en un orden lógico definido, primero las dimensiones y después los hechos.

#### Trabajo utilizado para la ejecución de la transformación del hecho importaciones por productos para el Grupo Empresarial y de transformaciones de las dimensiones relacionadas con él.



Figura 4: Trabajo para cargar las dimensiones y el hecho.



Figura 5: Trabajo ejecutado correctamente.

### **3.3 Empleo de Vistas Materializadas.**

Las vistas materializadas son tablas de la base de datos muy similares (prácticamente iguales) a las tablas de hechos. La diferencia entre una tabla de hecho y una vista materializada, consiste principalmente, en que una tabla de hecho almacena toda la información histórica de la que se dispone y una vista materializada sólo la información histórica para un período determinado, lo que permite ser más consecuente a la hora de ofrecer datos útiles pues es más relevante conocer el comportamiento de tendencias del último año o los últimos dos años, que el de veinte.

Además las vistas materializadas se utilizan para mejorar el rendimiento de la base de datos y agilizar el acceso a los reportes de información. Estas vistas se programan en el gestor de bases de datos PostgreSQL, primeramente creando una copia del hecho al que se le va a desarrollar la vista y luego definiendo las funciones disparadoras para las vistas y los hechos, que contienen el código, encargado de actualizar la vista cuando se modifiquen, eliminen o inserten datos en la tabla de hecho correspondiente y mantener la relación entre ellos.

El mercado de datos cuenta con 45 tablas de hecho y para cada una de ellas se definió, teniendo en cuenta el estándar de código, una vista materializada que muestra la información de los dos últimos años.

Los datos funcionales que se muestran en los reportes de información se pueden optimizar obteniéndolos de las vistas materializadas para ahorrar tiempo, tanto a la hora del sistema brindar una respuesta como a la de los usuarios finales encontrar determinada información.

```

$begin
if (TG_OP='INSERT') then
INSERT INTO mart_economia.vm_hech_cuentas_mcpal_2a(
vm_pk_hech_cuentas_mcpal_2a_id,      fk_dim_cuenta_id,      fk_dim_conceptos_cmcpal_id,
fk_dim_tiempo_mensual_id,          real_mes,              acumulado,              ano)
VALUES (NEW.pk_hech_cuentas_mcpal_id, NEW.fk_dim_cuenta_id, NEW.fk_dim_conceptos_cmcpal_id,
NEW.fk_dim_tiempo_mensual_id, NEW.real_mes,NEW.acumulado,NEW.ano);
end if;

if (TG_OP='UPDATE') then
delete from mart_economia.vm_hech_cuentas_mcpal_2a;
INSERT INTO mart_economia.vm_hech_cuentas_mcpal_2a(
vm_pk_hech_cuentas_mcpal_2a_id,      fk_dim_cuenta_id,      fk_dim_conceptos_cmcpal_id,
fk_dim_tiempo_mensual_id,          real_mes,              acumulado,              ano)

SELECT pk_hech_cuentas_mcpal_id, fk_dim_cuenta_id, fk_dim_conceptos_cmcpal_id, fk_dim_tiempo_mensual_id,
real_mes, acumulado, ano FROM mart_economia.hech_cuentas_mcpal
where ano >= (SELECT max(ano) AS max FROM mart_economia.hech_cuentas_mcpal)-1;
end if;

if (TG_OP='DELETE') then
delete from mart_economia.vm_hech_cuentas_mcpal_2a
where vm_pk_hech_cuentas_mcpal_2a_id = OLD.pk_hech_cuentas_mcpal_id;
end if;

return null;
end;$BODY$
LANGUAGE plpgsql VOLATILE
COST 100;
ALTER FUNCTION mart_economia.fc_tgr_hech_cuentas_mcpal() OWNER TO postgres;

```

*Figura 6: Código de la función disparadora para el hecho cuentas municipales.*

### 3.4 Implementación de los Subsistemas de Visualización.

Visualizar la información de un mercado de datos, tiene como objetivo fundamental extraer, de las fuentes disponibles, el conocimiento necesario para beneficiar a la toma de decisiones. Los procesos de Inteligencia de Negocio son los encargados de mostrar el resultado final del ciclo de desarrollo de un mercado o almacén de datos a través de la confección de cubos multidimensionales y reportes de información que pueden ser de varios tipos.

#### Diseño de cubos.

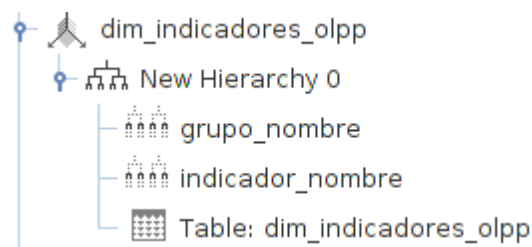
A través de la herramienta Pentaho Schema Workbench se crean los cubos multidimensionales que responden a cada tabla de hecho y las dimensiones asociadas a ellas.

Por cada tabla de hecho se define un cubo que carga la información, en el caso del

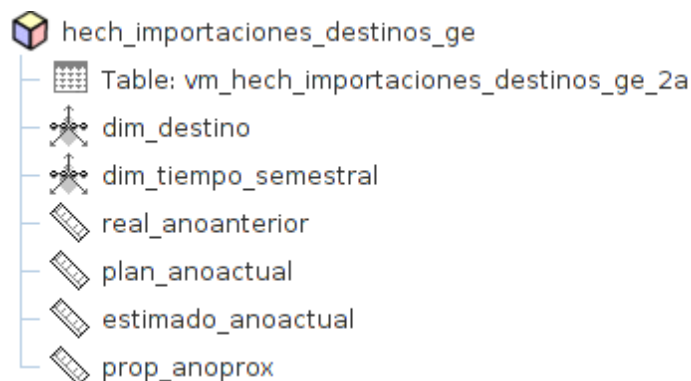


presente mercado se definió la carga de los cubos a partir de las vistas materializadas en aras de optimizar los reportes de información. Los cubos y dimensiones con sus medidas, jerarquías y niveles respectivamente, así como la conexión con la base de datos que contiene el mercado, son salvados en un fichero de tipo .xml generado por la herramienta.

Se modelaron en total 45 cubos multidimensionales junto a las 43 dimensiones asociadas para la implementación de los Subsistemas de Visualización.



*Figura 7: Modelado de la dimensión indicadores para el Órgano Legislativo del Poder Popular en la herramienta Pentaho Schema Workbench.*



*Figura 8: Diseño del cubo correspondiente al hecho para las importaciones por Destinos en el Grupo Empresarial.*

## Arquitectura de información.

Una vez culminado el diseño de los cubos y el modelado de las dimensiones, se define la estructura organizativa del mercado de datos, que está compuesta por

cinco Áreas de Análisis, en respuesta a las direcciones que integran la Dirección General de Economía, 23 libros de trabajos en correspondencia con los temas de análisis definidos y 50 reportes que se derivan de los requisitos informativos.

- [-] MD\_Direccion\_Economia
  - [-] AA Analisis de los Indicadores de Economia y Planificacion
    - [-] LT Analisis de los indicadores de Eficiencia del Territorio
    - [-] LT Analisis de los indicadores de Ejecucion de Divisa
    - [-] LT Analisis de los indicadores de Preparacion para la Defensa
    - [-] LT Analisis de los indicadores de las Importaciones
    - [-] LT Analisis de los indicadores de las Inversiones
    - [-] LT Analisis de los indicadores del Consumo Energetico
  - [-] AA Analisis de los Indicadores de Energia
    - [-] LT Analisis de los indicadores del consumo estatal
  - [-] AA Analisis de los Indicadores de Finanzas y Precios
    - [-] LT Analisis de los indicadores de Ingresos y Gastos
    - [-] LT Analisis de los indicadores de Inmuebles
    - [-] LT Analisis de los indicadores de Multas
    - [-] LT Analisis de los indicadores de las Cuentas
    - [-] LT Analisis de los indicadores del Presupuesto Central del Estado
    - [-] LT Analisis los indicadores del Programa Especial
  - [-] AA Analisis de los Indicadores de Perfeccionamiento Empresarial
    - [-] LT Analisis de los indicadores de eficiencia en las empresas
  - [+] AA Analisis de los Indicadores de Trabajo y Seguridad Social

*Figura 9: Arquitectura de información del Mercado de Datos.*

La definición de Áreas de Análisis y Libros de Trabajo en la estructura de navegación del mercado permite agrupar los reportes de información según la afinidad de sus temas, facilitando así la comprensión de los usuarios.

### **Implementación de las Vistas de Análisis.**

Las Vistas de Análisis son un tipo de reportes de información iterativos y dinámicos donde se realiza la navegación dimensional sobre los cubos a partir del Pentaho BI Server, visualizándose en formato web.

Después de establecida la arquitectura de información se procede a publicar los cubos contenidos en el archivo .xml a través de Pentaho Schema Workbench. Cuando se ha llevado a cabo la publicación se accede al Pentaho BI Server para generar las Vistas de Análisis a través de la pantalla de inicio de esta herramienta.

Municipios	Tiempo	Medidas			
		● Expedientes con Certificado Contable	● Expedientes Inscritos en el Registro Administrativo de Patrimonio	● Expedientes Inscritos en el Registro Administrativo de la Vivienda	● Total de Expedientes Iniciados
Alquízar	Marzo 2011	6	18	8	26
Artemisa	Marzo 2011	7	11	9	20
Bahía Honda	Marzo 2011	8	9	10	19
Bauta	Marzo 2011	3	9	10	19
Caimito	Marzo 2011	2	2	3	5
Candelaria	Marzo 2011	8	14	0	14
Guanajay	Marzo 2011	1	0	1	1
Güira de Melena	Marzo 2011	5	0	0	
Maríel	Marzo 2011	0	2	3	5
San Antonio de los Baños	Marzo 2011	4	2	3	5
San Cristóbal	Marzo 2011	9	12	10	22

*Figura 11: Vista de Análisis del Mercado de Datos.*

Una vez generada la vista se guarda en el libro de trabajo al que corresponde con el formato VA\_[nombre\_vista].

### **Implementación de los Reportes.**

Los reportes dan respuesta a las necesidades de información del cliente a través de consultas a la base de datos donde se encuentra el mercado. Se implementan con la herramienta Pentaho Report Designer.



**Administración Provincial de Artemisa**

**Dirección General de Economía**

Departamento Independiente de Energía

**RESUMEN PROVINCIAL**

**Informe sobre el Consumo del Sector Estatal**

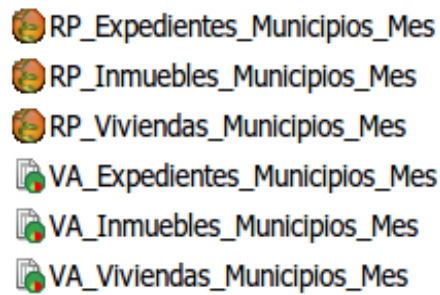
**Mes:** Julio 2011

<b>Sectores</b>	<b>Plan del Mes</b>	<b>Consumo del Mes</b>
Barra de la Provincia	24877,8	25093
Consumo Sector Estatal	11497,1	11714
Centros Seleccionados	6538,28	5967
Resto del Estatal	18339,5	4150

*Figura 12: Reporte del Mercado de Datos.*

Los reportes pueden ser muy útiles a la hora de brindar información en formato adecuado para la impresión o presentación de informes y resúmenes.

Cuando son diseñados se publican a través del Pentaho Report Designer hacia la herramienta Pentaho BI Server y se ubican en el libro de trabajo que les corresponde, denominándolos RP\_[nombre\_reporte].



*Figura 13: Conjunto de vistas de análisis y reportes del Mercado de Datos agrupados en un libro de trabajo.*

Lo expuesto a través de este capítulo evidencia cómo se ha llevado a cabo la implementación del mercado de datos que propone la presente investigación. Habiéndola efectuado de forma consecuente, a partir del desarrollo de los subsistemas de ETL y de visualización según los procesos que tienen lugar en la Dirección General de Economía de la Administración Provincial de Artemisa, quedó definido un total de 50 reportes y 50 vistas de análisis para satisfacer los pedidos de información requeridos por parte del cliente y dejando el mercado listo para probar su funcionalidad.

### **Capítulo 4: Validación del Mercado de Datos de la Dirección General de Economía de la Administración Provincial de Artemisa.**

En el presente capítulo se aplicarán los casos de prueba con el objetivo de validar el mercado de datos a través de la verificación de coincidencia entre los resultados esperados y los obtenidos, todo orientado a satisfacer las necesidades de información de los usuarios finales, es decir, el personal de la Dirección General de Economía de la Administración Provincial de Artemisa.

#### **4.1 Validación y prueba.**

Las pruebas permiten medir la calidad del software. En aras de determinar la calidad de un producto de software deben llevarse a cabo actividades que permitan comprobar el nivel de cumplimiento de las especificaciones iniciales del sistema, y es entonces que las pruebas de software desempeñan un papel primordial. Ellas se encargan de verificar el desarrollo que va alcanzando el producto durante todas sus etapas, identificando posibles fallos de diversas índoles como son: de implementación, calidad, usabilidad y seguridad, entre otros. Estas pruebas fueron aprobadas y comprobadas por el grupo de Calidad de Software del Centro de Desarrollo de la Facultad Regional “Mártires de Artemisa” Ver Anexo 8.

#### **Diseño de Casos de Prueba.**

El diseño de los casos de prueba pertenece a las pruebas funcionales y se realiza para comprobar que la documentación del producto se corresponde con lo establecido en el mercado, o sea este se ejecuta sobre ciertas condiciones para descubrir errores antes de la entrega del software al cliente. En el caso del mercado de datos para la Dirección General de Economía de la Administración Provincial de Artemisa se diseñaron 31 casos de prueba, uno correspondiente a cada caso de uso.

<b>Escenario</b>	<b>EC 1.1:</b> Balance de la Seguridad Social.		
<b>Descripción</b>	Permite visualizar tipos de reportes con las variables presentes en ellos.		
<b>Perfiles de Análisis</b>	Comprobaciones	Tiempo Mensual	
<b>Indicadores a medir</b>	Parcial	Debe	Haber
<b>Respuesta del sistema</b>	Se muestran los reportes referentes al Balance de la Seguridad Social con los valores correspondientes a los perfiles de análisis.		
<b>Flujo central</b>	<p>Se abre la aplicación.</p> <p>Se autentifica.</p> <p>Se entra al sistema.</p> <p>En el menú lateral del navegador se selecciona el área de análisis Trabajo y Seguridad Social.</p> <p>Se selecciona el libro de trabajo Análisis de los indicadores del Balance de la Seguridad Social y se selecciona el reporte deseado.</p> <p>En el área de trabajo se visualiza la tabla correspondiente al reporte.</p>		
<b>Resultado de la Prueba</b>	Satisfactorio.		

**Tabla 2: Diseño del caso de prueba “Visualizar los indicadores específicos del Balance de la Seguridad Social”.**

**Estrategias de pruebas para el Plan de Pruebas.**

En la solución se establece un plan de pruebas para verificar la calidad del proceso

de implementación, en el que se recogen precondiciones y medidas que facilitan la identificación de resultados verídicos y eficientes. A continuación se muestra un ejemplo de las pruebas establecidas para la capa de visualización:

<b>Descripción</b>	Se verifica el acceso y privilegios que posee cada usuario o rol de la Dirección de Colaboración Económica, definido con respecto al sistema.
<b>Precondiciones</b>	Acceso restringido a la información, solo se puede ver la información necesaria del mercado para la Dirección de Colaboración Económica para el trabajo de cada usuario o rol. Cada usuario o rol debe tener acceso pleno a la información que maneja, no pueden existir información inaccesible para ellos si en el negocio fue definido su acceso.
<b>Medidas</b>	Cumple o no cumple con los privilegios de acceso.

**Tabla 3: Prueba de acceso a la información por roles o niveles de usuario.**

#### **4.2 Aporte social y económico.**

El costo de desarrollar y mantener activo un mercado de datos bajo licencias de software propietario en el mundo, se extiende desde miles hasta millones de dólares actualmente. El caso de empleo de licencias de software libre reduce ese costo de manera significativa, y particularmente, en Cuba, las empresas dedicadas a implementar software generalmente no exceden en sus precios los 500 pesos convertibles ó 12 000 pesos en moneda nacional.

Para el caso de los mercados de datos, su costo estimado se encuentra bastante cercano a las cifras más elevadas debido a varias razones como la complejidad y el tiempo de desarrollo, la cantidad de personal y las tecnologías a emplear en la construcción y el despliegue de la solución.



La presente investigación brinda a los especialistas de la Dirección General de Economía de la Administración Provincial de Artemisa una herramienta capaz de contribuir al correcto análisis y la seguridad de la información, proporcionando un nivel de análisis basado en experiencias y una proyección de la información para agilizar el proceso de toma de decisiones, y de esta manera, eliminar las pérdidas de datos contables que son fundamentales en la economía de la provincia y, por consiguiente, del país. Permite reducir tiempos de espera por parte de los usuarios ofreciendo un incremento en la rapidez de las consultas y brindando un acceso integrado, consistente, fiable y rápido a los datos. Esto impacta positivamente al lograr mayor productividad por decisiones correctas en menos tiempo e implica un desempeño positivo en los procesos que se llevan a cabo en esta Dirección, dotándola de información estratégica para su mejorar sus funciones de trabajo.

Además, constituye un significativo paso de avance en el ahorro para la Administración Provincial de Artemisa por insumos como el material de oficina y otros recursos que eran destinados al análisis de la información a través de otras vías y asimismo por el costo de implementación e implantación de un sistema de este tipo por cualquiera de las empresas que desarrollan software en Cuba.

Se obtuvo como resultado el Mercado de Datos para la Dirección General de Economía que permite centralizar la información de los sistemas fuentes y la gestionarla consecuentemente para el apoyo a la toma de decisiones. Las principales funcionalidades con que cuenta el mercado de datos son:

- Creación de reportes ad- hoc.
- Creación de Vistas de Análisis.
- Creación de reportes de datos.
- Graficación de la información.
- Autenticación de usuarios según sus niveles de acceso.

En el transcurso del presente capítulo se han expuesto los elementos que permiten obtener los resultados de la etapa de validación y prueba, explicando la estrategia

seguida para lograr llevar a término la misma. Además de describir las distintas pruebas efectuadas para controlar el cumplimiento de los objetivos definidos. De los 31 casos de prueba aplicado, correspondientes a los casos de uso del sistema todos arrojaron resultados satisfactorios.

## **Conclusiones**

Para finalizar la presente investigación se concluye que, con su elaboración, se han cumplido de forma propicia los objetivos trazados al inicio a través de las tareas definidas para guiar el proceso de desarrollo del sistema.

Existe fundamentación teórica que avala el uso de almacenes de datos para propiciar el correcto análisis de la información y el apoyo a la toma de decisiones.

Se consultó y sistematizó la bibliografía accesible relacionada con las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.

Se realizó el análisis y diseño, identificando las dimensiones y tablas de hechos correspondientes a los procesos apreciados en el negocio.

Se efectuó la implementación mediante los subsistemas de integración y visualización, obteniendo como resultado los reportes que dan solución a los requisitos definidos.

A través de los casos de prueba se evidenció la factibilidad de la solución implementada, así como su calidad y eficiencia dadas por la respuesta satisfactoria a las necesidades del cliente y/o los usuarios finales.

El mercado de datos que se propone da solución al problema planteando pues se logra el correcto análisis de la información y se tiene una mayor seguridad de esta, lo que beneficia al proceso de toma de decisiones en la Dirección General de Economía de la Administración Provincial de Artemisa.

### Recomendaciones

Con el propósito de enriquecer la propuesta realizada en este trabajo, se sugiere:

- Mantener actualizado el Mercado de Datos ante los posibles cambios que ocurran para/con los datos almacenados.
- Continuar con el proceso de desarrollo en próximas versiones para lograr que el sistema sea más robusto y flexible, y satisfaga nuevas necesidades que puedan surgir por parte de los clientes.
- Desplegar la solución propuesta en esta investigación y capacitar adecuadamente al personal que debe interactuar con ella.
- Tener en cuenta los niveles de accesibilidad definidos para los roles/usuarios descritos y establecer los permisos de acceso acorde a estos.
- Realizar, una vez desplegada la solución, el proceso de extracción, transformación y carga al mercado, dándole continuidad y culminación.

## Bibliografía

1. Adamson, Christopher. 2006. Mastering Data Warehouse Aggregates. EUA: Wiley Publishing Inc, 2006.
2. Alarcón José Manuel, H. (2006) "Administración de SGBD PostgreSQL." Volume, DOI.
3. Devlin, Barry. 1997. Data Warehouse from Architecture to Implementaton, 1997.
4. England, Ken y Powell, Gavy. 2007. Performance, Optimization and Tunning handbook. eua: EISEVIER Inc, 2007.
5. García, Lucina: Apuntes sobre Data Warehousing. Italian Cuban Seminar on Informatics Methods and Applications. Universidad de La Habana, Cuba, Diciembre, 2000.
6. Hernández, A. Coello. El paradigma cuantitativo de la investigación científica. Ciudad de La Habana. Editorial universitaria, 2002.
7. Huamantumba, Rayner. 2007. Manual para diseño y desarrollo de Data Mart. 2007.
8. Hurtado, y otros. 2003. Base de Datos y Data Warehouse: Herramientas Estratégicas para la eficacia comercial. Granada: s.n., 2003.
9. Imhoff, Claudia, Galemno, Nicholas y Geiger, Jonathan G. 2003. Mastering Data Warehouse Desing, Relational and Dimentional Techniques. EUA: Wiley Publishing Inc, 2003.
10. Inmon, William. Building the Data Warehouse. 3ra. John Wiley & Sons, Inc, 2002. 432 p. 0471081302.
11. Inventario del agua y Almacenes de Datos. Disponible en: <http://www.aguainfo.com/?tag=almacenes-de-datos>.
12. Kimball, Ralph. The Data Warehouse Toolkit. 1a. New York, John Wiley & Sons, Inc, 1996. 388 p. 0471153370.
13. Kimball, Ralph y Ross, Margy. 2002. The Data Warehouse Toolkit. EUA: Wiley Publishing Inc, 2002.

14. Lockhart, Thomas. 2000. Tutorial de PostgreSQL. 2000.
15. López Espinosa, José Antonio. Acerca de las ciencias de la información. Disponible en: [http://bvs.sld.cu/revistas/aci/vol1\\_1\\_93/aci04193.htm](http://bvs.sld.cu/revistas/aci/vol1_1_93/aci04193.htm)
16. Osterfeldt, Susan. Minería de Datos. Disponible en: <http://www.bibliociencias.cu/gsd/collect/libros/index/assoc/HASH0106/b6fac6b9.dir/doc.pdf>
17. Pedroso Izquierdo, Evelyn. Breve historia del desarrollo de la Ciencia de la Información. Disponible en: [http://bvs.sld.cu/revistas/aci/vol12\\_2\\_04/aci07204.htm](http://bvs.sld.cu/revistas/aci/vol12_2_04/aci07204.htm).
18. Peñaloza, Lucía Victoria Hernández. 2008. Tesis para logra el título de Magíster: Diseño y Construcción de un Data Mart para la mantención de Indicadores de Sostenibilidad de la Industria del Salmón. Chile: s.n., 2008.
19. Ponniah, Paulraj. Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals. New York, Chichester, Weinheim, Brisbane, Singapore y Toronto, John Wiley & Sons, Inc., 2001. 518 p. 0471412546.
20. SIE (Almacén de Datos Multidimensional). Disponible en: [http://www.jcyl.es/web/jcyl/Estadistica/es/Plantilla100/1131978002439/\\_/\\_/](http://www.jcyl.es/web/jcyl/Estadistica/es/Plantilla100/1131978002439/_/_/).

## Referencias Bibliográficas

1. Inmon, W. H. 2002. Building the Data Warehouse. New York : John Wiley & Sons, Inc., 2002.
2. Kimball, R. y Margy, R. 2002. The Data Warehouse Toolkit. 2002.
3. Romero, A. L. (2006). Bodegas de Datos: Consideraciones de Implementación, Seguridad y Control. Obtenido de <http://www.acis.org.co/fileadmin/Conferencias/ConfAlfonsoLuqueNov30-06.pdf>
4. García, Gerardo Clemente. Riunet. Riunet. [Online] Junio 2008. Disponible en:  
<http://riunet.upv.es/manakin/bitstream/handle/10251/2505/tesisUPV2842.pdf>.
5. Hobbs, Lilian, y otros. 2005. Oracle Database 10g Data Warehousing. EUA: ELSEVIER Digital Press, 2005.
6. Cabrera, María Evelia Casales. Facultad de Ciencias. Facultad de Ciencias. [Online] Enero 2009. <http://hp.fciencias.unam.mx/~alg/bd/dwh.pdf>.
7. Sinexus. 2007. Sinexus. [En línea] 2007. [http://www.sinnexus.com/business\\_intelligence/olap\\_avanzado.aspx](http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx).
8. Ponniah, Paulraj. 2001. Data Warehousing Fundamentals. EUA : Wiley Publishing Inc, 2001.
9. ETL-Tools.Info. (2006). "Esquema de constelación de hechos ", from [http://etl-tools.info/es/bi/almacendedatos\\_esquema-constelacion.htm](http://etl-tools.info/es/bi/almacendedatos_esquema-constelacion.htm).
10. ETL-Tools.Info. (2006). "Esquema en copos de nieve." from [http://etl-tools.info/es/bi/almacendedatos\\_esquema-copo-de-nieve.htm](http://etl-tools.info/es/bi/almacendedatos_esquema-copo-de-nieve.htm).
11. Berzal, Fernando. [Online] <http://elvex.ugr.es/idbis/db/docs/intro/Modelomultidimensional.pdf>.
12. Wolff, Carmen Gloria. Modelamiento Multidimensional. Modelamiento Multidimensional. [Online] Agosto 28, 2002. <http://www.inf.udec.cl/~revista/ediciones/edicion4/modmulti.PDF>.

## REFERENCIAS BIBLIOGRÁFICAS

---

13. Díaz, Mg. Samuel Oporto. 2009. [www.wiphala.net](http://www.wiphala.net). [En línea] 22 de agosto de 2009. [www.wiphala.net/courses/dwh/ICSI243/2009-II/class/class\\_11\\_bi.ppt](http://www.wiphala.net/courses/dwh/ICSI243/2009-II/class/class_11_bi.ppt)
14. Guevara Lenis, Jorge Eduardo. Valencia Arcos, Janeth del Carmen. [Online] Junio 2007. <http://bibdigital.epn.edu.ec/bitstream/15000/445/1/CD-0827.pdf>.
15. Curto, Josep. 2007. Information Management Data Warehousing. Data Warehouse y Datamart. [En línea] 2007 <http://informationmanagement.wordpress.com/2007/10/07/data-warehousing-data-warehouse-y-datamart/>
16. Límia Navarro, Alberto. Delfino, Anisley. Hernández López, Asnioby. Medina Mustelier, Doris. Cárdenas Tandrón, Iván M. Ortiz, Julio Ernesto. Socas, Marisleidys. Calvo, Osniel. Peña Vazquez, Yanet. Gonzales Hernández, Yanisbel. Otros compañeros del Centro UCI-Villa Clara. Metodología de Desarrollo de Soluciones BI y Warehousing (BI & W). Ciudad de La Habana, Universidad de las Ciencias Informáticas: s.n.
17. Power Architect. Disponible en: <http://www.sqlpower.ca/page/architect>
18. Power Architect Características. Disponible en: <http://www.dataprix.com/empresa/productos/sql-power-software>
19. La plataforma Pentaho Open Source Business Intelligence. [En línea] 2008. Disponible en: <http://pentaho.almacen-datos.com/>
20. ArPug. (2009). ArPug - Grupo de usuarios PostgreSQL de Argentina. Retrieved 2 11, 2010, from ArPug - Grupo de usuarios PostgreSQL de Argentina. Disponible en: <http://www.arpug.com.ar>
21. Technologies, B. P. (2010). Visual Paradigm. Retrieved 2 11, 2010, from Visual Paradigm. Disponible en: <http://www.visual-paradigm.com>
22. Pentaho Data Integration. Disponible en: <http://www.inqbation.com/pentaho-data-integration/>



## REFERENCIAS BIBLIOGRÁFICAS

---

23. Schema Worbench. Disponible en:  
<http://mondrian.pentaho.com/documentation/workbench.php>
24. Pentaho Report Design. [En línea]. 2011. Disponible en:  
<http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>

*Anexo 1: Entrevista efectuada a los trabajadores de la Dirección General de Economía de la Administración Provincial de Artemisa.*

*La presente Entrevista se efectúa a los Jefes de Departamento de la Dirección General de Economía como parte del Levantamiento de Información para el desarrollo del Sistema Informativo de la Administración Provincial de Artemisa y del Trabajo de Diploma “Mercado de Datos que contribuya a la toma de decisiones en la Dirección General de Economía de la Administración Provincial de Artemisa” para lo cual necesitamos que conteste las preguntas siguientes. ¡Gracias por su cooperación!*

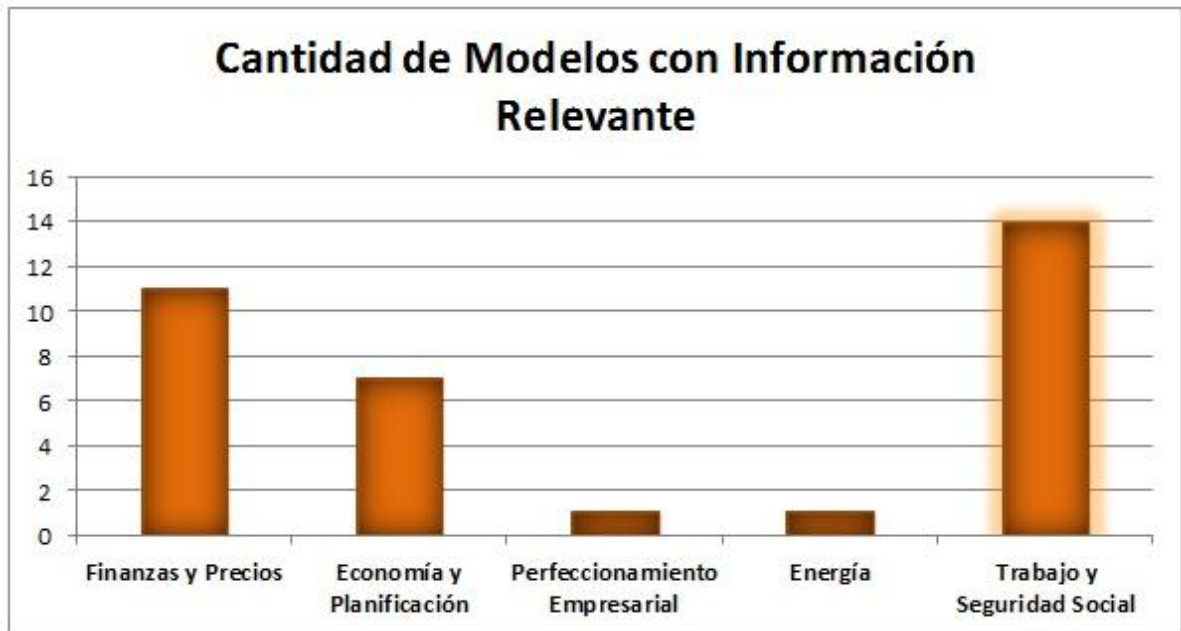
1. *¿Con cuántos modelos de información relevante trabaja su departamento?*
2. *¿Con qué frecuencia recibe y/o analiza la información relevante con la que trabaja?*
3. *¿A través de qué vías recibe la información?*
4. *¿En qué formato la reciben y/o la gestionan?*
5. *¿Cuáles son los principales problemas que tienen con la información?*

*Análisis de los resultados*

Se entrevistó a los 12 Jefes de Departamento de la Dirección General de Economía de la Administración Provincial de Artemisa y a partir de las respuestas que brindaron se elaboran las tablas siguientes que resumen los resultados obtenidos:

1. Cantidad de modelos de información relevante con los que trabaja cada departamento.

Departamento Balances Materiales	4
Departamento Control y Análisis del Plan	3
Departamento Asistencia Social	1
Departamento Seguridad Social	4
Departamento Empleo y Capacidad	7
Departamento Trabajo y Salario	2
Departamento Multa	1
Departamento Patrimonio	3
Departamento Presupuesto	3
Departamento Tesorería	4
Departamento Independiente Energía	1
Sección Independiente Perfeccionamiento Empresarial	1



**Gráfico 1: Modelos de información relevante que trabaja cada departamento.**

2. Frecuencia con la que se recibe y/o analiza la información relevante.

Departamento Balances Materiales	Mensual
	Semestral
Departamento Control y Análisis del Plan	Mensual
	Trimestral
	Anual
Departamento Asistencia Social	Mensual
	Semestral
Departamento Seguridad Social	Mensual
Departamento Empleo y Capacidad	Quincenal
	Mensual
Departamento Trabajo y Salario	Mensual
	Trimestral
Departamento Multa	Mensual
Departamento Patrimonio	Mensual
Departamento Presupuesto	Semanal
	Mensual
Departamento Tesorería	Mensual
Departamento Independiente Energía	Mensual
Sección Independiente Perfeccionamiento Empresarial	Trimestral



Gráfico 2: Frecuencia con la que se recibe/ analiza la información relevante.

## 3. Vías por las que se recibe la información.

Departamento Balances Materiales	Correo electrónico
Departamento Control y Análisis del Plan	Manual
	Correo electrónico
Departamento Asistencia Social	Correo electrónico
Departamento Seguridad Social	Correo electrónico
	Soportes extraíbles
Departamento Empleo y Capacidad	Correo electrónico
	Soportes extraíbles
Departamento Trabajo y Salario	Correo electrónico
	Manual
Departamento Multa	Vía telefónica
	Correo electrónico
Departamento Patrimonio	Manual
	Correo electrónico
Departamento Presupuesto	Vía telefónica
	Correo electrónico
Departamento Tesorería	Correo electrónico
	Soportes extraíbles
Departamento Independiente Energía	Correo electrónico
Sección Independiente Perfeccionamiento Empresarial	Manual
	Soportes extraíbles

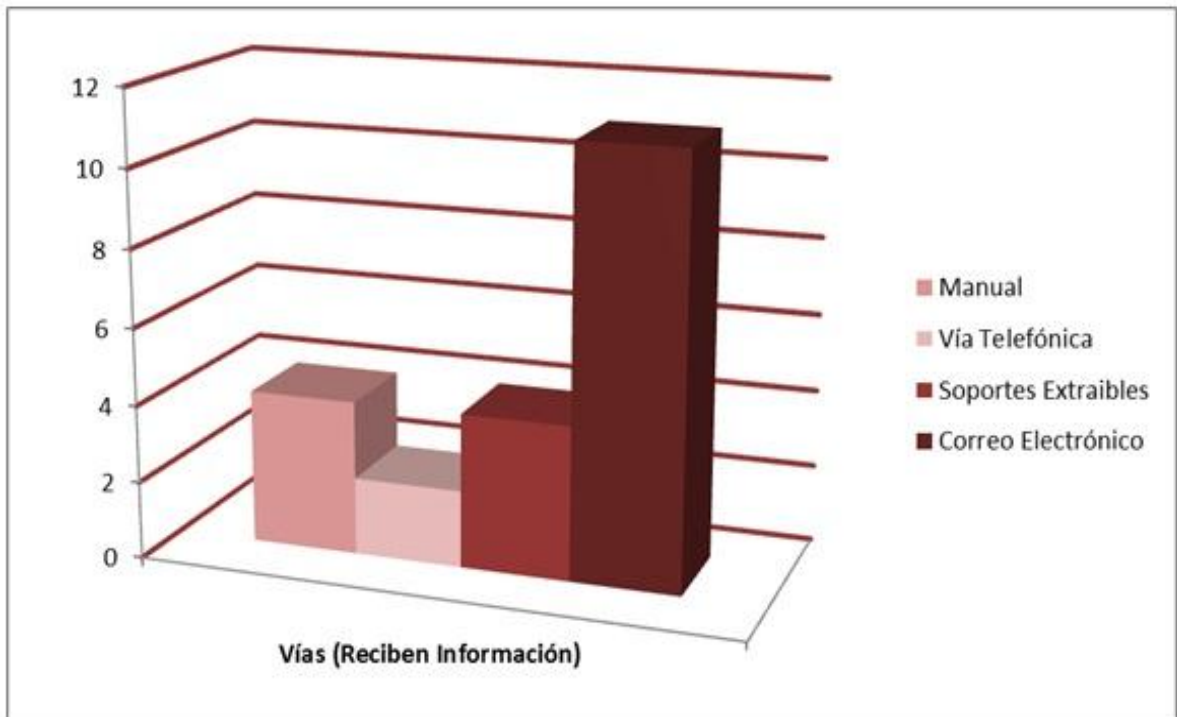


Gráfico 3: Vías por las que se recibe la información.