



Facultad Regional de Artemisa “Mártires de Artemisa”

**TRABAJO DE DIPLOMA PARA OPTAR POR EL TÍTULO DE INGENIERO  
EN CIENCIAS INFORMÁTICA**

**Título:** Mercado de Datos que contribuye a la toma de decisiones en la  
Dirección de Inspección de la Administración Provincial de Artemisa.

**Autor:** Alejandro Heredia Molina

**Tutora:** Ing. Mayrin Ramos Mestre

**Co-Tutor:** Raúl Edis Riverón Núñez

Artemisa junio 2012



***"Las ciencias tienen las raíces amargas, pero muy dulces los frutos."***

**DECLARACIÓN DE AUTORÍA**

Declaro ser autor de la presente tesis y reconozco a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

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

\_\_\_\_\_  
**Autor:** Alejandro Heredia Molina

\_\_\_\_\_  
**Tutor:** Ing. Mayrin Ramos Maestre

\_\_\_\_\_  
**Co-Tutor:** Raúl Edis Riverón Núñez

*A toda mi Familia en Especial a mi Madre*

### *Agradecimientos*

*A mi madre y mi padre todo el amor y cariño con que me han enseñado, su apoyo y dedicación.*

*A mi familia que siempre apoyándome y ayudándome.*

*A Magda, Norberto, Oriol y Moty por haber confiado en mí y por su cariño incondicional.*

*A Yirina y su familia por haberme acogido y darme su apoyo y cariño.*

*A los Akatsuky, Peter y al Noe por dejarme formar parte de ellos y estar siempre ahí cuando los necesitaba.*

*A todas mis amigas por el tiempo dedicado durante los cinco años.*

*A Saray, Yisel, Rodaisy, Leidy Laura, Lianne por siempre estar ahí para ayudarme y apoyarme.*

*A mi tutora Mayrín Ramos Maestre por su gran paciencia y por siempre tenerme algún caramelo de dulzura para sobrellevarme.*

*A mi co-tutor Riveron por su paciencia y sus consejos que siempre estuvieron ahí para ayudarme.*

*A todos los profesores durante los cinco años por su paciencia y dedicación.*

*En general a todos el colectivo de la mini.*

### **Resumen**

El presente trabajo de diploma abarca los temas de Almacenes de Datos (AD), Mercados de Datos (MD) y su utilización en la toma de decisiones como necesidad fundamental en el actual mundo de negocios. Se utiliza la metodología de proceso de desarrollo en la línea soluciones de AD e inteligencia de negocio, basada en la metodología de Kimball, con adaptaciones de los recursos humanos en otros proyectos desarrollados.

Se realiza un análisis y diseño partiendo de los requisitos que se establecieron en el levantamiento de información.

Para la implementación del mercado de datos se utilizaron herramientas libres, por lo que los procesos de Extracción, transformación y carga (ETL) y el de inteligencia de negocio (BI) se desarrollaron satisfactoriamente.

Se le realizan pruebas a la solución propuesta las mismas evidencian que el producto tiene la calidad debida y que puede ser correctamente implantado en la empresa correspondiente.

El MD desarrollado surge para almacenar los datos de la actual Administración Provincial de Artemisa debido a toda la problemática que presenta la misma en cuanto al manejo de sus datos, permitiendo un fácil manejo de la información, contribuyendo de esta manera a la integralidad, confiabilidad y disponibilidad de la misma para el proceso de tomas de decisiones.

Palabras Claves: almacenes de datos, mercado de datos, toma de decisiones.

**Contenido**

Contenido .....	6
Introducción.....	1
Capítulo 1: Fundamentación teórica.....	7
Introducción.....	7
1.1 Estado del arte .....	7
1.2 Sistemas de almacenes de datos.....	8
1.3 Propósitos de los MD.....	11
1.4 Características Generales.....	12
1.5 Componentes del MD.....	14
1.6 Modelamiento.....	15
1.7 Modos de almacenamiento de datos.....	19
1.8 Metodologías para el desarrollo del MD .....	23
1.9 Herramientas a utilizar.....	26
Conclusiones del Capítulo.....	29
Capítulo 2: Análisis y diseño del MD para la Dirección de Inspección.....	30
Introducción.....	30
2.1 Reglas del Negocio .....	30
2.2 Especificación de requerimientos.....	31
2.3 Requisitos funcionales.....	33
2.4 Requisitos no funcionales.....	33
2.5 Casos de uso del sistema.....	35
2.6 Subsistemas de integración y visualización.....	41
Conclusiones del Capítulo.....	42
Capítulo 3: Implementación del MD de la Dirección de Inspección.....	43
Introducción.....	43
3.1 Implementación de la base de datos.....	43
3.2 Implementación de los subsistemas de integración .....	45
3.3 Implementación de los trabajos.....	47
3.4 Implementación del subsistema de visualización de datos.....	48
Conclusiones del Capítulo.....	50
Capítulo 4. Validación del MD de la Dirección de Inspección del Consejo de la Administración Provincial de Artemisa.....	51
Introducción.....	51
4.1 Pruebas.....	51
4.2 Diseño de los casos de prueba .....	51
4.3 Plan de Pruebas.....	53
4.4 Resultados obtenidos.....	57
Conclusiones del Capítulo.....	57
Conclusiones Generales .....	58
Recomendaciones.....	59

Referencias Bibliográficas .....	60
Bibliografía.....	61
Glosario de Términos .....	63
Anexos .....	64



## **Introducción**

Las tecnologías de la información y la comunicación (TIC) desataron una explosión sin precedentes a comienzo de los años '90. A partir de ahí, la Internet pasó de ser un instrumento especializado de la comunidad científica, a ser de fácil uso, lo cual modificó las pautas de interacción social.

Estas siguen su expansión por el mundo de manera imparable. El uso de las TIC en la población, ayuda a disminuir la brecha digital existente ya que aumentaría el conglomerado de usuarios que utilizan las TIC como medio tecnológico para el desarrollo de sus actividades.

Actualmente se sigue perfeccionando el trabajo y ampliando el radio de acción de las nuevas tecnologías en beneficio de todas las personas. En Cuba se habla y planifican metas ambiciosas que están a la altura de los países del primer mundo y que ya hoy no se está muy lejos de poderlas alcanzar, ejemplo de ello es la Industria Cubana del Software. Los altos ejecutivos del gobierno han tenido que buscar alternativas que le permitan al país estar en el mundo tecnológico debido a las presiones de Estados Unidos de América (EUA) que controlan la mayoría de las empresas de software siendo todas de carácter propietario, una de las opciones es emigrar a software libre. La Universidad de las Ciencias Informáticas (UCI), no está ajena a estos cambios ya que todas las facultades de la misma han decidido cambiar su trabajo con software propietario implantando el software libre. Una de estas facultades es la Facultad Regional de Artemisa "Mártires de Artemisa" en la misma desde hace varios años solo se trabaja con herramientas de las comunidades de software libre. Evidenciándose en el uso del lenguaje de programación JAVA, con integraciones de frameworks como Symfony y Jwebsocket, este último trabaja en tiempo real para el caso de los proyectos de Aplicaciones Informáticas.

En la facultad se desarrollan AD con la utilización de herramientas libres como es el caso de Pentaho Data Integration. Anteriormente se realizaron sistemas de apoyo a la toma de decisiones, entre estos se encuentran el proyecto con la Oficina

Nacional de Estadísticas e Información (ONEI) y un mercado de datos para la Dirección de Colaboración Económica del Ministerio del Comercio Exterior y la Inversión extranjera (MINCEX).

Actualmente se trabaja en un AD para la Dirección de Economía y Planificación del MINCEX y con un AD para la Administración Provincial de Artemisa.

La Administración Provincial de Artemisa tiene como función principal la de preparar y proponer la política integral del estado. Además, se encarga de dirigirla y ejecutarla. Está compuesta por 32 direcciones, entre ellas la Dirección de Inspección.

La Dirección de Inspección cuenta con 21 modelos que son los que almacenan toda la información de las inspecciones a los municipios mensualmente. Debido a esto surge la siguiente **situación problemática**: la información es obtenida de forma impresa en Excel o manual lo cual trae como consecuencia pérdida y duplicado de la información, atraso y lentitud a la hora de responder a los pedidos de información, imposibilitando el correcto análisis para la toma de decisiones.

Por lo anteriormente planteado surge el siguiente **problema científico**:

¿Cómo contribuir a la integridad, seguridad, disponibilidad y análisis de la información para la toma de decisiones en la Dirección de Inspección de la Administración Provincial de Artemisa?

**Objeto de estudio:** Almacenes de Datos.

**Campo de acción:** Mercado de Datos en el sector de Inspecciones.

Para dar solución a la problemática antes mencionada se ha definido como **objetivo general**:

Desarrollar un MD que contribuye a la integridad, seguridad, disponibilidad y análisis de la información para la toma de decisiones en la Dirección de Inspección de la Administración Provincial de Artemisa.

### **Objetivos Específicos**

- Investigar los fundamentos teóricos de los MD para gestión de la información

dedicados a la toma de decisiones.

- Fundamentar las metodologías, herramientas y tecnologías a utilizar en el desarrollo de AD.
- Realizar el análisis y diseño del MD para gestión de la información en de la Dirección de Inspección de la Administración Provincial de Artemisa.
- Implementar el MD para gestión de la información en de la Dirección de Inspección de la Administración Provincial de Artemisa.

### **Idea a Defender:**

Con la realización de un MD para la Dirección de Inspección de la Administración Provincial se contribuirá a la integridad, seguridad, disponibilidad y análisis de la información para la toma de decisiones.

### **Tareas de la Investigación**

1. Definición del estudio del estado del arte de los sistemas dedicados a la toma de decisiones.
2. Caracterización de las metodologías, tecnologías y herramientas a utilizar en el desarrollo de MD.
3. Realización del análisis y diseño del MD para gestión de la información en la Dirección de Inspección de la Administración Provincial de Artemisa.
4. Implementación del MD para la Dirección de Inspección de la Administración Provincial de Artemisa.
5. Aplicación de los casos de prueba.

### **Aporte Práctico**

La investigación realizada tendrá como aporte práctico la posibilidad de realizar un MD para la Dirección Inspección detallado con toda la base teórico-práctico sobre la cual se sustente la solución propuesta, contribuyendo a la centralización y organización de la información para facilitar la consulta y utilización de la misma por parte del personal que labora en la institución.

Para llevar a cabo las tareas de esta investigación se emplearon los siguientes métodos.

### **Métodos de Investigación**

Los métodos de investigación son los que permiten llegar a los problemas y soluciones por distintas vías. Son pasos a seguir guiándose por ciertos requisitos y lineamiento.

#### **Métodos teóricos:**

**Método histórico-lógico**, este permite estudiar de forma analítica las trayectorias históricas reales, de las soluciones de Inteligencia de Negocios, su evolución y desarrollo. Durante la implementación del mercado de datos se realizó un estudio relacionado con los almacenes de datos, mercados de datos, metodologías y herramientas seleccionándose así los que permitirían optimizar y beneficiar al cliente, brindándole un producto fácil de manejar y que le sirviera de apoyo para la toma de decisiones.

**Método analítico – sintético**, el cual permite buscar la esencia de los fenómenos que describen la Inteligencia de Negocios, conjuntamente con las metodologías y herramientas a tratar, permite extraer los elementos más importantes que se relacionan con el objeto de estudio. En el proyecto se hizo un estudio donde se analizaron distintas metodologías, herramientas, de se utilizaron las más óptimas y que se ajustarán al proyecto.

**Método lógico – deductivo**, es el razonamiento que, partiendo de casos particulares, se eleva a conocimientos generales. Este método permite la formación de hipótesis, investigación de leyes científicas, y las demostraciones. En la

elaboración del mercado de datos se ha estudiado los problemas existentes el cual con su realización se eliminarán partiendo de las situaciones problemáticas hasta los beneficios que aportaría el mismo.

**Métodos empíricos:** El método empírico es un modelo de investigación científica, que se basa en la experimentación y la lógica empírica, permitiendo la observación y el análisis. Es uno de los métodos más utilizados en la investigación científica, ya que permite percibir directamente, sin intermediarios que deformen la percepción, los hechos de la realidad objetiva.

**Encuesta,** es un método que permite obtener información a través de preguntas relacionados con las problemáticas existentes dentro del negocio. Ayuda a un mejor entendimiento siendo de gran importancia para el diseño e implementación. En la Dirección de Inspección se realizó en una **población** tomada es de 45 trabajadores de la Dirección de Inspección y la **muestra** 19 especialistas de la Dirección de Inspección.

### **Variables Independientes**

- Mercado de Datos para la Dirección de Inspección.

### **Variables Dependiente**

- Integridad, Seguridad, Disponibilidad y análisis de la información.

### **Estructura del Documento**

Este trabajo de diploma consta de una introducción, 4 capítulos, conclusiones, recomendaciones, referencias bibliográficas, bibliografías, glosario de términos y para culminar los anexos.

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

En este capítulo se realiza un estudio del estado del arte de la investigación, se evidencian los principales conceptos de los temas tratados en el desarrollo del

almacén de datos, se presenta la metodología seleccionada y las herramientas utilizadas para el tratamiento necesario de la información.

### **Capítulo 2: Análisis y diseño del MD para la Dirección de Inspección de la Administración Provincial de Artemisa.**

Se definen los requerimientos del sistema, se diseña el diagrama de casos de uso del sistema, los subsistemas de integración y visualización. Se describe la arquitectura del sistema incluyendo la arquitectura de la información del mismo. Se diseñan los modelos dimensionales y las vistas materializadas. Se presentan los reportes y gráficos que serán utilizados así como la estrategia para la elaboración de vistas de análisis.

### **Capítulo 3: Implementación del MD de la Dirección de Inspección 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, los reportes candidatos y la política de seguridad de los usuarios.

### **Capítulo 4: Validación del MD de la Dirección de Inspección de la Administración Provincial de Artemisa.**

Se aplicarán los casos de prueba basados en casos de uso para validar la calidad de los datos cargados, viendo que esto cumpliera con las necesidades de los usuarios de la Dirección de Inspección de la Administración Provincial de Artemisa.

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

### **Introducción**

El uso de sistemas de bases de datos automatizadas, se desarrolla a partir de la necesidad de almacenar grandes cantidades de datos para su posterior consulta, producidas por las nuevas industrias que creaban gran cantidad de información. Pero con el tiempo las empresas fueron acumulando grandes volúmenes de información que hacía muy difícil hacer las consultas, debido a esta problemática se fueron buscando distintas vías que permitieran el fácil manejo y rapidez de la información.

Este capítulo tiene el propósito de describir los principales conceptos, características y las herramientas que se utilizan en los nuevos sistemas de gestión de la información. También permitirá conocer las metodologías a utilizar y la influencia que ha tenido en el mundo y en especial es Cuba.

### **1.1 Estado del arte**

#### **En el Mundo**

Con el transcurso del tiempo las empresas fueron almacenando un gran número de información en diferentes fuentes de datos (documentos de texto, ficheros relacionados, bases de datos), pero con el avance tecnológico se hizo necesario buscar una tecnología que permitiera almacenar los grandes volúmenes de datos y de manera histórica, naciendo así los AD y los directivos de las empresas se dieron cuenta que ésta, podría ser útil ya que reflejaba la mayoría de las operaciones diarias del negocio.

A partir de esta problemática un gran número de compañías se dieron a la tarea de implementar la tecnología de AD para convertir a sus datos en información útil. En el mercado minorista es donde más incidencia ha tenido esta novedosa forma de utilizar los datos debido a la necesidad de estudiar los clientes potenciales y estar actualizado ante la gran competencia que existe en este sentido.

En América Latina existen empresas como Telefónica de Argentina, Visa, Arcor, todas de Argentina; en México existen algunas como Walmart, Procter & Gamble, Whirpol, Tv Azteca, Baxter, GNP, Warner Lambert y Sabre que también han venido incorporando el uso de los AD para la toma de decisiones a nivel gerencial.

También el Instituto Nacional de Estadística de Venezuela posee un conjunto de MD para realizar el análisis de las operaciones estadísticas. Dentro de todos estos mercados también existen empresas como el Banco Interamericano de Desarrollo y la Comunidad Farmacéutica de Madrid que han implementado no solo MD, sino que lo han hecho en el sector de inspecciones, posibilitando un análisis en pos de la implementación del mercado que se desea implementar.

### **En Cuba**

Con esta nueva tecnología en Cuba se ha logrado crear un medio para el análisis y búsquedas de estrategias que permitan a nuestros directivos tomar decisiones sólidas y fundamentadas.

Ejemplos consistentes son la elaboración de almacenes para diversas empresas cubanas como son: CUBAENERGIA la cual ha permitido la asimilación de esta tecnología en el estado del arte para soportar la Gestión del Conocimiento en dicha organización. La ONEI ha desarrollado almacenes y específicamente mercados en el sector de inspecciones, al igual que la empresa de telecomunicaciones ETECSA, sirviendo de base para el estudio de la actual elaboración del mercado.

Después de que se realizó un estudio de los almacenes y mercados de datos tanto en el mundo como en Cuba, es necesario argumentar que ningunos de los creados tiene relación con el que se desea implementar, debido a que no guarda similitud con el sector gubernamental, ni con el sistema de la Administración Provincial de Artemisa.

### **1. 2 Sistemas de almacenes de datos.**

Con la aparición de computación en los años 70, las empresas empezaron a almacenar información que cada día se hacía más grande los volúmenes de la



misma. Para poder guardar ese cúmulo de datos se empezó a utilizar sistemas de ficheros relacionados, lo que fue sustituido posteriormente por los sistemas de bases de datos que prestaban mejor servicio. Al aumentar la tecnología se hizo necesario buscar un sistema que mantuviera el manejo de la información de forma óptima y rápida para ser analizada.

Los AD han facilitado no solo el manejo de la información sino que ayuda a la toma de decisiones de las empresas, lo cual están en constante transformación.

Antes de profundizar en los AD es necesario hablar de las personalidades y las formas de conceptualizar esta terminología, que aunque tienen distintos puntos de vista concuerdan en un único objetivo en el avance de las tecnologías de gestión de información.

“Data Warehouse (AD) es un conjunto de datos integrados, históricos, variantes en el tiempo y unidos alrededor de un tema específico, que es usado por la gerencia para la toma de decisiones”(Bill Inmon, 1992).

“Un Data Warehouse (AD) es una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis” (Ralph Kimball, 2002).

En 1993, Susan Osterfeldt publica una definición que sin dudas acierta en la clave del AD: “Yo considero al AD como algo que provee dos beneficios empresariales reales: integración y acceso a los datos. AD elimina una gran cantidad de datos inútiles y no deseados como también el procesamiento desde el ambiente operacional clásico”

Todos los autores convergen de algún modo con la definición de un AD y desde sus percepciones se puede decir que los AD son estructuras que se definen en función de temas específicos donde la información histórica debe estar integrada, robusta ante los cambios que puedan afectar a la organización y que su objetivo principal, y es lo que define su razón de ser, es servir de ayuda a la toma de decisiones empresariales.

Después de los puntos de vistas anteriores, se considera que los almacenes de

datos son el apoyo fundamental para las empresas en la actualidad, ya que permiten grandes volúmenes de almacenamiento, tener información durante mucho tiempo lo cual permite la comparación posibilitando una buena toma de decisiones en los negocio de la misma. Por lo cual se comparte el mismo criterio de Ralph Kimball (2002).

### **Mercados de Datos**

Los MD o Data Mart es una versión especial de AD. Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades. (INMON, 2012)

El MD es un sistema orientado a la consulta, en el que se producen procesos batch de carga de datos (altas) con una frecuencia baja y conocida. Es consultado mediante herramientas OLAP (On line Analytical Processing - Procesamiento Analítico en Línea) que ofrecen una visión multidimensional de la información. Sobre estas bases de datos se pueden construir EIS (Executive Information Systems, Sistemas de Información para Directivos) y DSS (Decision Support Systems, Sistemas de Ayuda a la toma de Decisiones).

Los MD que están dotados con estas estructuras óptimas de análisis presentan las siguientes ventajas:

- Poco volumen de datos
- Mayor rapidez de consulta
- Consultas SQL y/o MDX sencillas
- Validación directa de la información
- Facilidad para la historización de los datos

**AD vs MD**

Mercado de Datos	Almacén de Datos
Se usa en un proceso, departamento o área específica.	Se usa en toda la empresa.
Requiere de 3 a 6 meses para su desarrollo.	Requiere de 1 a 3 años para su desarrollo.
Bases de Datos medianas [100Mb-100Gb].	Bases de Datos grandes[100Gb-100Tb]

**Tabla 1: Comparación entre almacén y mercado de datos**

**1. 3 Propósitos de los MD.**

En la actualidad los modelos de datos son de gran ayuda para predecir el valor de ciertos valores a partir de otros datos. De esta manera es posible establecer una serie de reglas sobre los factores que afectan a cada caso y así tener una base para tomar decisiones en casos futuros.

➤ **La información debe ser accesible.**

La información al ser almacenada, sus niveles, tienen que ser correctos y obvios lo que permita reconocer el punto de llegada con solo un clic. Los contenidos al ser entendibles y navegables traerán consigo un rápido desempeño permitiendo que el tiempo de espera sea nulo. El diseño de las estructuras debe soportar que los usuarios combinen y separen los datos en un sinfín de combinaciones (a este proceso se le denomina slicing y dicing).

➤ **La información de la organización debe presentarse de manera consistente.**

Se define como consistente aquello que une y relaciona las partes de un conjunto y les da unidad y coherencia. De esta manera es que debe ser procesada la información. La información debe presentar alta calidad facilitando la comparación entre procesos de un determinado negocio. Los datos al ser almacenados deben estar limpios permitiendo la accesibilidad a los usuarios.

➤ **La información adaptable y elástica.**

El MD está diseñado para cambios continuos. Cuando se le hacen nuevas preguntas los datos existentes y las tecnologías no cambian ni se corrompen. Al agregar datos nuevos, los datos existentes y las tecnologías tampoco cambian ni se corrompen.

➤ **Un seguro baluarte que protege los valores de la información.**

En el MD no solo controla el acceso efectivo a los datos, también a los dueños de la información evitando el mal uso y abuso de la misma.

➤ **Apoyo crucial para la toma de decisiones.**

Los datos almacenados deben ser altamente confiables para que sean de ayuda a la toma de decisiones empresariales. La organización de la información debe ser lo suficientemente dinámica y efectiva para que pueda ser servida oportunamente a los usuarios.

## **1. 4 Características Generales.**

Los almacenes de datos tienen 4 características fundamentales que lo convierten el mayor apoyo para la toma de decisiones, estas son:

➤ **Orientado al Tema**

El ambiente operacional se diseña alrededor de las aplicaciones y funciones tales como préstamos, ahorros, tarjeta bancaria y depósitos para una institución financiera. Por ejemplo, una aplicación de ingreso de órdenes puede acceder a los datos sobre clientes, productos y cuentas. La base de datos combina estos elementos en una estructura que acomoda las necesidades de la aplicación. (SINNEXUS, 2008)

En el ambiente de MD se organiza alrededor de sujetos tales como cliente, vendedor, producto y actividad. Por ejemplo, para la Dirección de Inspecciones, estos pueden ser las inspecciones, inspectores, multas, gastos y decretos.

Las diferencias entre la orientación de procesos y funciones de las aplicaciones y

la orientación a temas, radican en el contenido de los datos a escala detallada. En el MD se excluye la información que no será usada por el proceso de sistemas de soporte de decisiones.

### ➤ **Integrado**

El aspecto más importante del ambiente de MD es que la información encontrada al interior está siempre integrada. La integración tiene varias formas de mostrarse, puede ser en convenciones de nombres consistentes, en la medida uniforme de variables, fuentes múltiples, entre otros. No importa las distintas codificaciones que existan en los diferentes formatos para almacenar la información en MD se convierten en un único formato permitiendo que la información este en un estado integrado uniforme. (SINNEXUS, 2008)

### ➤ **Variable en el tiempo**

Toda la información del MD es requerida en algún momento. Esta característica básica de los datos en un depósito, es muy diferente de la información encontrada en el ambiente operacional. En éstos, la información se requiere al momento de acceder. Como la información en el MD es solicitada en cualquier momento, los datos encontrados en el depósito se llaman de "tiempo variante".

En el ambiente operacional el tiempo en el que la información es importante es de 2 a 3 años, mientras en el MD como mínimo se establece la estructura de 5 a 10 años. (Immon, 2005).

Otra de las características que hacen al MD variable en el tiempo se debe a la estructura clave, es decir que siempre se encuentra de manera implícita o explícita el elemento tiempo como día, semana, mes. (SINNEXUS, 2008)

### ➤ **No volátil**

La información es útil sólo cuando es estable. Los datos operacionales cambian sobre una base momento a momento. La perspectiva más grande, esencial para el análisis y la toma de decisiones, requiere una base de datos estable.

Mientras en las bases de datos operacionales existen operaciones como insertar,

eliminar y modificar, en los almacenes las únicas que se producen son la carga de los datos y el acceso a los mismos. No existe actualización de los datos. (SINNEXUS, 2008)

### 1. 5 Componentes del MD.

Independientemente que un MD está compuesto por una variedad de elementos específicos, generalmente tienen algunos componentes en común. Estos componentes son:

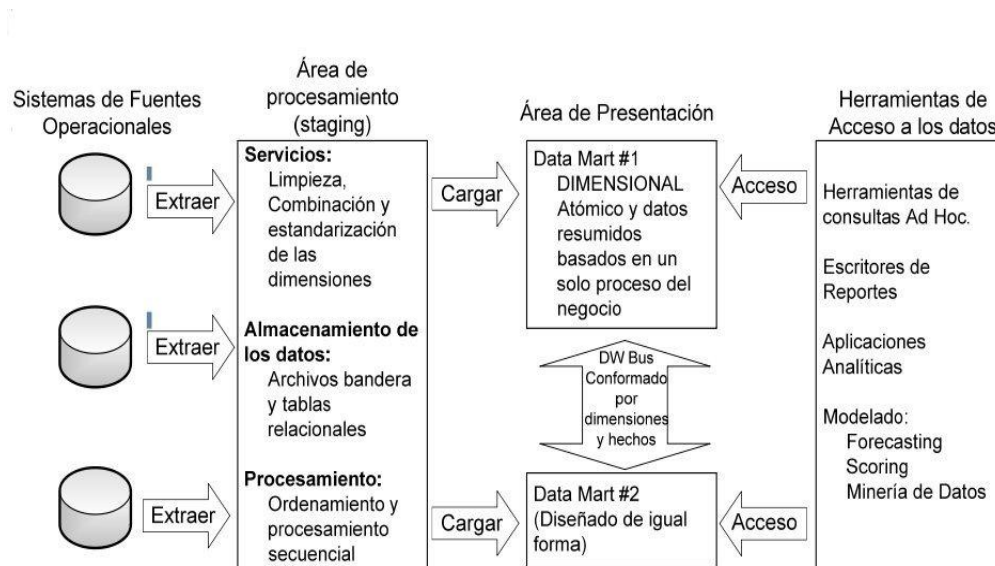


Figura 1: Componentes del Mercado de Datos.

➤ **Fuentes de Datos.**

A partir de este componente es que se realizan las capturas de los datos al AD. Estas pueden ser hechas de varios formatos, desde una base de datos hasta Excel, XML, DBF, textos planos entre otros.

➤ **Extracción y Transformación.**

Este componente es el encargado de mover la información con las transformaciones necesarias desde las fuentes de datos al AD. Se realiza el proceso de extracción de los datos de las diversas fuentes operacionales que se deseen integrar, teniendo como principal tarea la de almacenar toda esa información en bases de datos relacionales, generalmente, para realizar el análisis y procesamiento de los datos.

➤ **Servidor de Datos.**

Los servicios que debe ofrecer incluyen un servicio de mantenimiento de datos y un servicio de distribución para exportar datos del AD a servidores de bases de datos descentralizadas, y a otros sistemas de soporte de decisiones de usuario. Este componente también ofrece servicios de seguridad (archivo, backup, recuperación) y monitorización.

### **1.6 Modelamiento**

#### **Modelo Entidad-Relación**

Un diagrama o modelo entidad-relación (a veces denominado por su siglas, E-R "Entity relationship", o, "DER", Diagrama de Entidad Relación) es un lenguaje para el modelado de datos de un sistema de información. Estos modelos expresan entidades más relevantes para el sistema, sus interrelaciones y propiedades.

Los modelos entidad – relación no son recomendables para el diseño de los almacenes de datos debido a que no garantizan la recuperación óptima del gran cúmulo de información que se almacena. Además estos diagramas tienden a resultar en un diseño normalizado mientras que en un almacén de datos, la normalización no es un requisito a tener muy en cuenta. (Hobbs, y otros, 2005)

#### **Modelo Dimensional**

El modelo dimensional es una adaptación especializada del modelo relacional, solía representar datos en depósitos de datos, en un camino que los datos fácilmente pueden ser resumidos usando consultas OLAP. En el modelo dimensional, una base de datos consiste en una sola tabla grande de hechos que

son descritos usando dimensiones y medidas.

Una dimensión proporciona el contexto de un hecho (como quién participó, cuando y donde pasó, y su tipo). Las dimensiones se toman en cuenta en la formulación de las consultas para agrupar hechos que están relacionados. Las dimensiones tienden a ser discretas y son a menudo jerárquicas; por ejemplo, la ubicación podría incluir el edificio, el estado, y el país.

Una medida es una cantidad que describe el hecho, tales como los ingresos. Es importante que las medidas puedan ser agregados significativamente - por ejemplo, los ingresos provenientes de diferentes lugares pueden sumarse.

### **Justificación del modelo a utilizar.**

En el presente trabajo se va a utilizar el modelo dimensional debido a que en el marco predecible del esquema estrella que a continuación se explica, resiste a los cambios inesperados en el comportamiento del usuario. Cada dimensión es equivalente a las demás y todas las dimensiones pueden ser concebidas como puntos de entrada hacia la tabla de hechos. El diseño lógico puede realizarse independientemente de los patrones de consulta esperados, además es flexible ya que los nuevos elementos de datos y las nuevas decisiones de diseño son fácilmente adaptables.

### **Esquema en estrella**

Es la arquitectura de almacén de datos más simple. En este diseño del almacén de datos la tabla de Variables (Hechos) está rodeada por Dimensiones y juntos forman una estructura que permite implementar mecanismos básicos para poder utilizarla con una herramienta de consultas OLAP. (ETL-TOOLS.INFO, 2011)

Consiste en estructurar la información en procesos, vistas y métricas recordando a una estrella. A nivel de diseño, consiste en una tabla de hechos rodeada de otras tablas que contiene información específica. En el esquema en estrella la tabla de hechos es la única tabla del esquema que tiene múltiples joins que la conectan con otras tablas. El resto de tablas del esquema (tablas de dimensión) únicamente



hacen join con esta tabla de hechos.

Para la materialización física de este tipo de modelo se utiliza comúnmente la propuesta realizada por Ralph Kimball llamada “esquema estrella”, que consiste en una tabla central denominada “tabla de hechos” y un conjunto de tablas llamadas “dimensiones”, que se relacionan con esta tabla central.

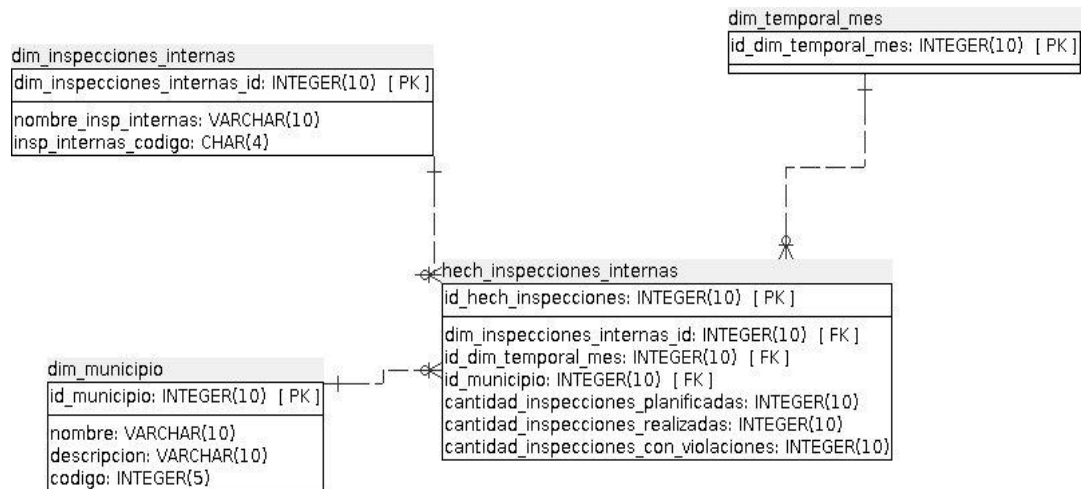


Figura 2: Esquema en Estrella.

Existen otras estructuras que surgen producto a modificaciones realizadas al esquema estrella. En este sentido se tiene el Copo de Nieve (Snowflake, en inglés); la citada estructura tiene como objetivo primordial su uso para el ahorro de espacio de almacenamiento. Se dice que una dimensión se encuentra “snowflaked” cuando los atributos de baja calidad se llevan a tablas separadas. La utilización de este tipo de estructura posee algunas deficiencias debido a que hace las presentaciones más complejas y afecta el rendimiento de la recuperación de las consultas. (CURTO, 2008)

### Esquema en Copo de Nieve

Es un esquema de representación derivado del esquema en estrella, en el que las tablas de dimensión se normalizan en múltiples tablas. Por esta razón, la tabla de hechos deja de ser la única tabla del esquema que se relaciona con otras tablas y

aparecen nuevas joins, gracias a que las dimensiones de análisis se representan ahora en tablas de dimensión normalizadas. En la estructura dimensional normalizada, la tabla que representa el nivel base de la dimensión es la que hace join directamente con la tabla de hechos. La diferencia entre ambos esquemas (star y snowflake) reside entonces en la estructura de las tablas de dimensión. Para conseguir un esquema en copo de nieve se ha de tomar un esquema en estrella y conservar la tabla de hechos, centrándose únicamente en el modelado de las tablas de dimensión, que si bien en el esquema en estrella se encontraban totalmente desnormalizadas, ahora se dividen en subtablas tras un proceso de normalización. Es posible distinguir dos tipos de esquemas en copo de nieve, un snowflake completo (en el que todas las tablas de dimensión en el esquema en estrella aparecen ahora normalizadas en el snowflake) o un snowflake parcial (sólo se lleva a cabo la normalización de algunas de ellas).

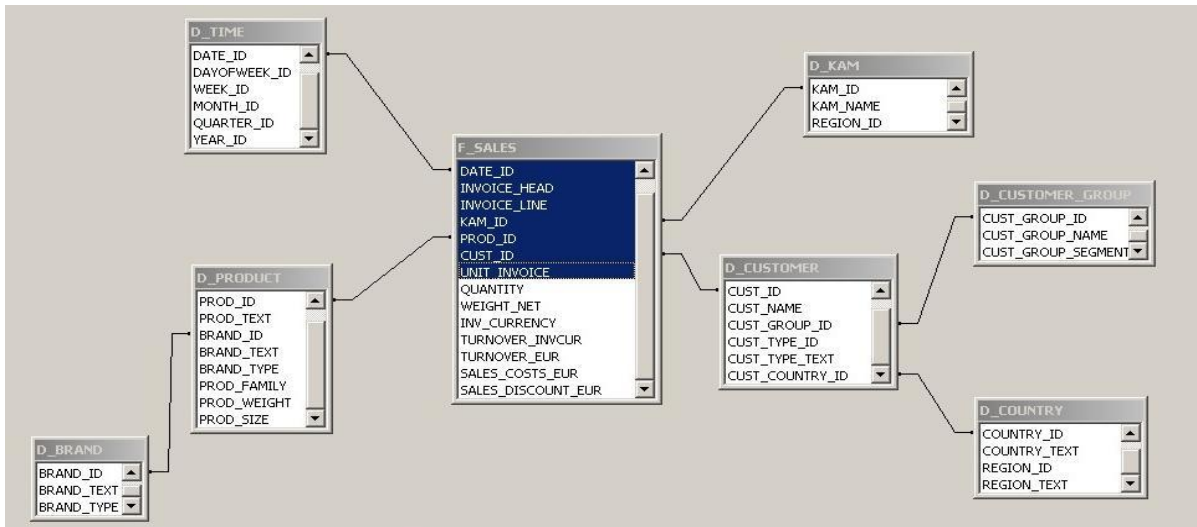


Figura 3: Esquema en Copo de Nieve.

### Justificación del esquema a utilizar.

El esquema utilizado es el esquema en estrella debido a que es más fácil su utilización en cuanto al copo de nieve, es un esquema lineal, las dimensiones no se normalizan. Con ello se logra minimizar el número de uniones y, por

consiguiente, incrementar el rendimiento de las consultas.

### **Tablas de Hechos**

Una tabla de hechos (o *tabla fact*) es la tabla central de un esquema dimensional (en estrella o en copo de nieve) y contiene los valores de las medidas de negocio. La principal condición que deben cumplir las tablas de hechos es que el hecho debe almacenarse de tal forma que su valor sea numérico y a su vez sea aditivo para así poder realizar cálculos sobre él, ya sea %, sumas, igualdades.

### **Tablas de Dimensiones**

Las tablas de dimensiones son las que describen la parte textual del negocio. Estas posee varios atributos lo cual definen una fila en la dimensión en el modelos dimensional. Los atributos de las dimensiones sirven como fuente primaria de las restricciones de las consultas, agrupaciones y las etiquetas de los reportes. Ellos desempeñan un rol de vital importancia dentro del MD debido a que son las llaves que hacen el MD usable y entendible.

### **1. 7 Modos de almacenamiento de datos.**

En una partición, el rendimiento de las consultas y el procesamiento se apoyan mucho en el modo de almacenamiento. OLAP (proceso analítico en línea) se divide en 3 modelos MOLAP, ROLAP, HOLAP. El proceso de análisis se realiza de igual forma lo que varía en uno y otro caso es la metodología de almacenamiento.

#### **MOLAP**

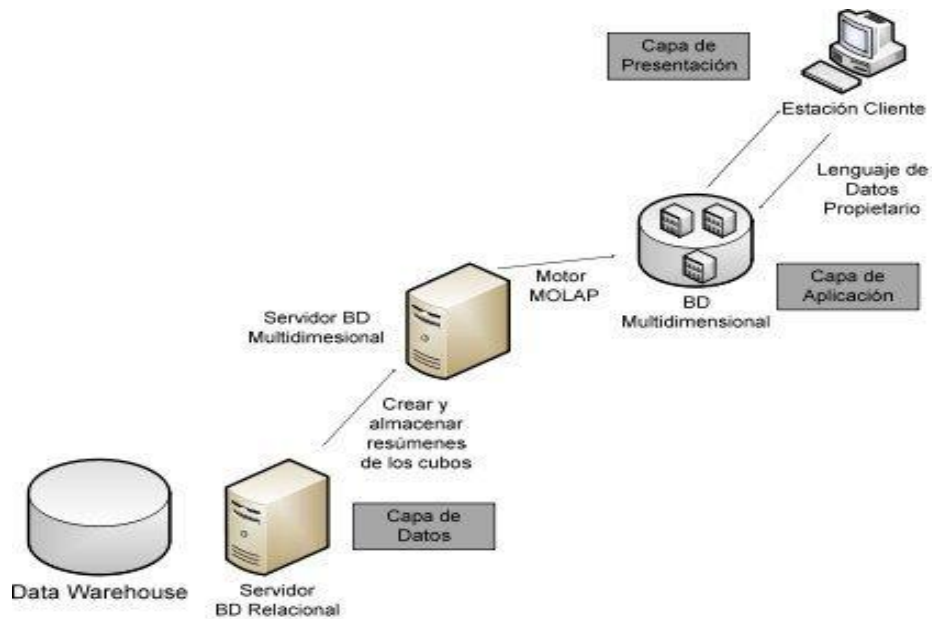
La arquitectura MOLAP usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente. La información se almacena multidimensionalmente, para ser visualizada en varias dimensiones de análisis. La misma cuenta con una arquitectura de dos niveles: la bases de datos multidimensionales y el motor analítico.

La base de datos es la encargada del acceso, obtención y manejo de los datos, mientras que el nivel del motor es responsable de la ejecución los requerimientos

OLAP. En este modelo la información se carga mediante una serie de rutinas por lotes, para calcular los datos agregados, a través de las dimensiones de negocio, rellenando la estructura Base de Datos multidimensional (MDDDB).

La arquitectura MOLAP requiere unos cálculos intensivos de compilación. Lee de datos precompilados, y tiene capacidades limitadas de crear agregaciones dinámicamente o de hallar ratios que no se hayan precalculados y almacenados previamente.

La arquitectura MOLAP requiere unos cálculos intensivos de compilación. Lee de datos precompilados, y tiene capacidades limitadas de crear agregaciones dinámicamente o de hallar ratios que no se hayan precalculados y almacenados previamente. (SINNEXUS, 2008)



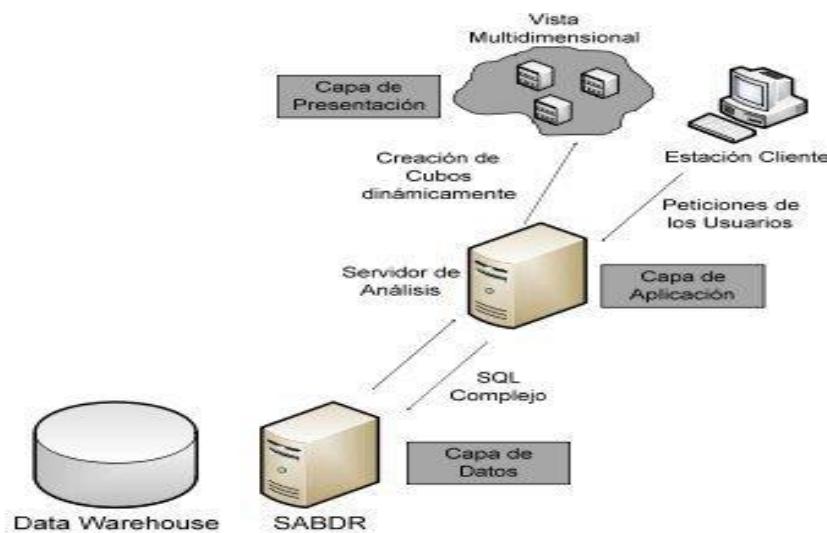
**Figura 4: Arquitectura MOLAP**

## **ROLAP**

Esta arquitectura accede a los datos de los almacenes de datos para proporcionar

los análisis OLAP. La misma utiliza 3 niveles, la base de datos relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica. El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y obtención del dato. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios. Este también se integra con niveles de presentación en los cuales los usuarios realizan sus análisis OLAP. A continuación se ejecutan rutinas de bases de datos para agregar el dato.

Los usuarios finales ejecutan sus análisis multidimensionales, a través del motor ROLAP, que transforma dinámicamente sus consultas a consultas SQL. Se ejecutan estas consultas SQL en las bases de datos relacionales, y sus resultados se relacionan mediante tablas cruzadas y conjuntos multidimensionales para devolver los resultados a los usuarios. (SINNEXUS, 2008)



**Figura 5: Arquitectura ROLAP.**

**Comparación entre MOLAP y ROLAP**

MOLAP	ROLAP
<ul style="list-style-type: none"> <li>• Diversos resúmenes de datos en bases de datos propietarias.</li> <li>• Volumen de datos moderados.</li> <li>• Resúmenes de acceso a datos detallados en bases de datos multidimensionales.</li> <li>• Crea cubos prefabricados por el motor OLAP.</li> <li>• Usa tecnología propietaria para almacenar las vistas multidimensionales en arreglos.</li> <li>• Usa una matriz de alta velocidad para la recuperación de datos.</li> <li>• Maneja poca tecnología de matriz de datos para la gestión de los resúmenes.</li> <li>• Tiene una librería de funciones para el cálculo complejo.</li> <li>• Facilita el análisis independientemente de la cantidad de dimensiones.</li> </ul>	<ul style="list-style-type: none"> <li>• Posee tablas relacionales que resumen los datos disponibles.</li> <li>• Tiene un gran volumen de datos.</li> <li>• Todos los datos de acceso están en la bodega de almacenamiento.</li> <li>• Utiliza SQL complejo para los datos del depósito.</li> <li>• Los cubos de datos son creados sobre la marcha por el servidor de análisis.</li> <li>• Muestra vistas multidimensionales en la capa de presentación.</li> <li>• Tiene un ambiente conocido.</li> <li>• Disponibilidad de herramientas.</li> <li>• Presenta limitaciones en funciones de análisis complejo.</li> </ul>

	<ul style="list-style-type: none"> <li>• Las agregaciones no son factibles.</li> </ul>
--	--

**Tabla 2: Comparación entre modos de almacenamientos**

**HOLAP**

Un desarrollo un poco más reciente ha sido la solución OLAP híbrida (HOLAP), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado. (SINNEXUS, 2008)

**Justificación del modo de almacenamiento a utilizar.**

El modo que se va a utilizar es el ROLAP debido a que el mercado de datos se construye sobre un Sistema Gestor de Base de datos (SGBD) relacional, y la carga de los datos en el proceso de ETL es de una base de datos relacional.

**1.8 Metodologías para el desarrollo del MD**

Metodología se refiere al conjunto de procedimientos basados en principios lógicos utilizados para alcanzar una serie de objetivos que rigen una investigación científica. La metodología es una etapa específica que procede de una posición teórica y epistemológica, para la selección de técnicas concretas de investigación.

Para abordar problemática y conceptos existen diferentes puntos de vistas, los cuales enriquecen de manera la forma de resolverlos. En los almacenes de datos se han destacado varias metodologías que definen y guían el ciclo de vida del

desarrollo concreto. Existen 2 criterios argumentados los cuales le han sirviéndole de guía a la comunidad mundial en cuanto a este tema.

Bill Inmon y Ralph Kimball son dos de las personalidades referentes y más influyentes en el área de los AD, y responsables de los dos enfoques a los que se hace referencia. Estas tendencias son conocidas como Metodología Kimball y Metodología Inmon.

La principal diferencia que existe entre ambas tendencias está basada en la forma de enfrentar el problema. La visión de Inmon se basa principalmente en un enfoque descendente (top-down) planteando la creación de un repositorio de datos corporativo como fuente de información consolidada, persistente, histórica y de calidad. Al ser construido de manera descendente los MD se nutren del AD 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 es la que nutre los MD. Por lo que no se crea los MD directamente desde el sistema OLTP a través de un área de ensayo. En lugar de ello, se crean a partir de la arquitectura relacional de los datos corporativos.

La propuesta de Kimball de dividir el mundo de BI entre el hecho y las dimensiones es muy eficaz y conduce a una solución completa en una cantidad muy pequeña de tiempo. Además, la técnica de Kimball tiene una gran cantidad de documentación y se puede encontrar una respuesta a casi todas las preguntas que se posean.

Existen otras metodologías como:

**HEFESTO:** Metodología propia que está fundamentada en una amplia investigación, comparación de metodologías y experiencias en procesos de almacenes de datos.



**DM2:** Esta metodología permite la implementación del proyecto por áreas de negocio de la empresa.

**CRISP-MD:** Está descrita en términos de procesos jerárquicos en cuatro niveles de abstracción de lo general a lo específico.

### **Metodología a utilizar.**

La Administración Provincial basándose en su papel como órgano rector en materia estadística en la provincia de Artemisa necesita de una metodología robusta y madura que garantice el éxito de la integración de la información que actualmente disponen. De todo el conjunto de metodologías estudiadas para enfrentar el desarrollo del MD se llega a la conclusión de utilizar la Metodología de Proceso de Desarrollo en la Línea Soluciones de almacenes de datos e inteligencia de negocio, basadas en la Metodología de Kimball, adaptándola, claro está, a la realidad de la Facultad Regional de Artemisa "Mártires de Artemisa", por las siguientes razones:

- La técnica de Kimball posee una gran cantidad de documentación y generalmente se puede encontrar una respuesta a casi todas las problemáticas que puedan presentar.
- Ralph Kimball es una figura emblemática en el mundo de warehousing teniendo publicados alrededor de 100 artículos científicos proponiendo mejoras al proceso, además de innumerables libros que se han posicionado como guías de obligatoria consulta para el desarrollo, ejemplo de esto es su libro Técnicas de Diseño Dimensional que en la actualidad se ha convertido en un "Best Seller" dentro del campo.
- Claridad de las actividades a realizar por cada rol propuesto.
- Esta metodología de dividir el mundo de BI entre el hecho y las dimensiones es muy eficaz y conduce a una solución completa en un tiempo razonable.

- Es iterativo, donde se construye una pieza a la vez (MD) garantizando mayor velocidad de respuesta a los clientes.
- La forma de almacenar la información es de fácil entendimiento por parte del usuario lo que permite mayor comprensión para el análisis de los datos que se encuentran integrados.
- Es una metodología resistente y adaptable ante los cambios.

## 1. 9 Herramientas a utilizar

La Administración Provincial de Artemisa es el órgano que dirige todos los procesos en la provincia de Artemisa por lo cual requiere una emigración hacia las Tecnologías de la Información y la Comunicación. Actualmente la Facultad está trabajando en pos de mejorar su sistema de almacenamiento, quedando seleccionado como Sistema Gestor de Base de Datos el PostgreSQL. Esta decisión ha sido previamente colegiada y aceptada por parte del cliente final debido a que dentro de sus políticas de migración se encuentran las de llevar a todas sus bases de datos hacia dicha plataforma. En este sentido la versión que se utilizará es la 8.4 por mejorar el rendimiento de las aplicaciones. Entre las principales características que avalan esta decisión se encuentran:

- El análisis de datos mucho más sencillo a través de funcionalidades avanzadas de ANSI SQL: 2003.
- Expresiones comunes de tabla y joins recursivos.
- Las estructuras de consulta aumentan sustancialmente la expresividad del dialecto SQL de PostgreSQL, permitiendo a los usuarios hacer preguntas interesantes en una sola consulta, que habría sido imposible de construir antes.

### 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, se podrá gestionar y manejar con Power Architect.

Se ha decidido utilizar estas herramientas ya que una de sus principales ventajas es que la UCI posee las licencias para utilizarla.

### **Visual Paradigm**

Visual Paradigm es una herramienta UML profesional que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. Presenta licencia gratuita y comercial. Es fácil de instalar y actualizar y es compatible entre ediciones.

### **Características:**

- Ingeniería de ida y vuelta.
- Ingeniería inversa - Código a modelo, código a diagrama.
- Generación de código - Modelo a código, diagrama a código.
- Diagramas de flujo de datos.
- Editor de figuras.

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

Además se utilizará el **Pentaho Data Integration (PDI)** en su versión 3.2.0 el mismo:

- Es de formato abierto y de fácil lectura para los xml que recogen transformaciones, tareas programadas y un repositorio relacional de metadatos ETL.
- Es aplicable a diversos tipos de bases de datos (SQL server, PostgreSQL, MySQL, Microsoft Access).
- Posee facilidad para la importación y exportación de datos de un formato a otro cualquiera.
- Su principal fortaleza es la posibilidad que brinda de ser extensible mediante pluggins.
- Alta compatibilidad con la herramienta ETL con gestor PostgreSQL en los llamados Paquetes, donde se ofrecen servicios de importación, exportación, transporte, y transformación de datos.
- Rendimiento eficiente: Alta 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 PDI.
- Capacidad del PDI para extraer y cargar datos utilizando el PostgreSQL. Con estas herramientas se tiene la posibilidad de no sólo almacenar los datos al más atómico de los detalles, sino que también se puedan guardar solamente los agregados necesarios.

### **Pentaho Meta-Editor.**

- El Metadatos es una característica de la Plataforma BI diseñada para facilitar a los usuarios acceder a información en términos de negocio. Muestra cómo crear un reporte ad\_hoc con la interfaz Web para reportes Ad\_hoc, con la posibilidad de elegir elementos específicos que deseen que se muestren en el reporte.

### **Shema workbench**

Es un entorno visual para el desarrollo y prueba de cubos OLAP. Esta herramienta se utiliza para la creación de los archivos XML que se usan para la construcción de los cubos. Además permite la ejecución de consultas MDX contra el esquema y la base de datos.

Después de haber analizado las herramientas para el diseño de los cubos, se elige Shema workbench v3.2.1 como herramienta para OLAP por su alto desempeño, análisis interactivo de grandes o pequeños volúmenes de información y exploración dimensional de los datos así como la amplia documentación que ofrece su comunidad de desarrollo.

### **Conclusiones del Capítulo**

A partir del estudio del estado del arte y el marco teórico realizado se concluyó que la tecnología apropiada para la problemática en cuestión es el MD con la utilización de la metodología de desarrollo adaptada bajo los principios de Kimball. Las herramientas a utilizar son el Sistema Gestor de Base de Datos PostgreSQL, Visual Paradigm for UML, Power Architect y el Pentaho Data Integration.

### **Capítulo 2: Análisis y diseño del MD para la Dirección de Inspección.**

#### **Introducción**

En el presente capítulo se detallaran las reglas del negocio y se brindará los diferentes análisis que se realizaron para dar cumplimiento a las primeras fases de la construcción del MD. Se definirán todas las características del sistema para garantizar su funcionamiento, especificándose sus requerimientos y los diagramas de casos de usos.

#### **2. 1 Reglas del Negocio**

Describe las políticas, normas, operaciones, definiciones y restricciones presentes en una organización que son de vital importancia para alcanzar los objetivos misionales.

Las reglas del negocio identificadas para la construcción del MD son:

- En la Dirección de Inspección cuando un Ministerio necesita información, estos se la solicitan al Presidente y a su vez el Presidente se la solicita al Centro de Información. Luego el Centro de Información se la hace llegar al Presidente el cual la brinda al Ministerio que realizó la solicitud.
- En caso de error en la información el Centro de Información (DOPI) lo soluciona realizando una visita al municipio que brindó la información.
- Cuando la Dirección de Inspección necesita información se la solicita al Presidente, el cual aprueba o no la solicitud. En caso de ser aprobada la Dirección de Inspección se acerca al Presidente de la Dirección Municipal y solicita la información deseada.
- En caso de encontrar errores en algunos campos debe dirigirse al director o subdirector de la dirección correspondiente para ser orientados sobre cómo solucionar el error.
- Se lanzará una alerta con los campos que presenten problemas.
- Los datos nulos se quedan en blanco.

- El responsable de introducir la información debe llenar estos datos nulos.

### **2. 2 Especificación de requerimientos.**

La especificación de requisitos es una descripción completa del comportamiento del sistema a desarrollar, restricciones determinadas de forma precisa que deben satisfacerse. El artefacto especificación de requisitos contiene una lista detallada y completa de los requisitos donde el nivel de detalle de estos se encuentra especificado de forma clara, concreta y consistente.

#### **Requisitos de información.**

Los requisitos informativos representan toda la información que debe estar disponible en el almacén para su consulta. Constituyen la entrada fundamental para todo el proceso de BI y para futuros reporte bases. A continuación se mostrarán los requisitos de información que fueron identificados durante el proceso de análisis.

RF1-Obtener acumulados de actividades por Mes, Municipio, Sistemas y Actividades.

RF2-Obtener acumulados de indicadores por Mes, Municipio, Organismos e Indicadores.

RF3-Obtener acumulado anterior, mes actual y acumulado actual por Mes, Municipio, Organismos de Monedas Libremente Convertibles e Indicadores.

RF4-Obtener acumulado anterior por Mes, Municipio, Organismos de Monedas Libremente Convertibles e Indicadores.

RF5-Obtener mes actual por Mes, Municipio, Organismos de Monedas Libremente Convertibles e Indicadores.

RF6-Obtener acumulado actual por Mes, Municipio, Organismos de Monedas Libremente Convertibles e Indicadores.

RF7-Obtener la cantidad de inspecciones planificadas, la cantidad de inspecciones realizadas, el por ciento de inspecciones realizadas, la cantidad de inspecciones con violaciones y el por ciento de inspecciones con violaciones por Mes, Municipio y Tipo de Inspecciones Internas.

RF8-Obtener la cantidad de inspecciones, la cantidad de apercibimiento, la cantidad de multas, la cantidad de infractores, el índice de detección, la cuantía de las multas, promedio de multas, el índice de imposición, la cantidad de licencias retiradas, la cantidad de decomiso, la cantidad de multas ilegales, la cuantía de las multas ilegales, promedio de multas ilegales por Mes y Municipio.

RF9-Obtener la cantidad de inspectores en plantilla, la cantidad de inspectores real, y la cantidad de inspectores actuantes por Mes, Municipio y Cargos.

RF10-Obtener la cantidad de inspectores actuantes por Mes y Municipio.

RF11-Obtener los gastos de salarios del mes, los gastos de salarios acumulados, los gastos corrientes del mes, los gastos corrientes acumulados, cantidad de gastos del mes y cantidad de gastos acumulados por Mes y Municipios.

RF12-Obtener cantidad de multas en el mes y el acumulado por Mes, Municipio y Concepto.

RF13-Obtener la cantidad de multas\_227 del mes actual y el acumulado anterior por Mes, Municipio, Decretos.

RF14-Obtener la cantidad de multas\_4272 del mes actual y el acumulado anterior por Mes, Municipio, Decretos.

RF15-Obtener la cantidad de multas\_155 del mes actual y el acumulado anterior por Mes, Municipio, Decretos.

RF16-Obtener la cantidad de multas\_272\_17 del mes actual y el acumulado anterior por Mes, Municipio, Decretos.

RF17-Obtener la cantidad de multas\_272\_18 del mes actual y el acumulado anterior por Mes, Municipio, Decretos.

RF18-Obtener la cantidad de multas\_211 del mes actual y el acumulado anterior por Mes, Municipio, Decretos.

RF19-Obtener la cantidad de multas\_272\_ura del mes actual y el acumulado anterior por Mes, Municipio, Decretos.

RF20-Obtener la cantidad de multas\_275\_10\_modif\_171 del mes actual y el acumulado anterior por Mes, Municipio, Decretos.

RF21-Obtener la cantidad de multas\_272\_ci del mes actual y el acumulado



anterior por Mes, Municipio, Decretos.

RF22-Obtener la cantidad de multas\_261\_1 del mes actual y el acumulado anterior por Mes, Municipio, Decretos.

RF23-Obtener la cantidad de multas\_164\_51 del mes actual y el acumulado anterior por Mes, Municipio, Decretos.

RF24-Obtener la cantidad de multas\_274\_mod\_174 del mes actual y el acumulado anterior por Mes, Municipio, Decretos.

### **2.3 Requisitos funcionales.**

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

- Realizar una copia de seguridad al repositorio una vez por mes para brindarle mantenimiento a la información almacenada.
- Asegurar la limpieza de los datos.
- Asegurar la integración de los datos desde distintas fuentes.
- Perfilar datos.
- Permitir la graficar de la información.
- Autenticar usuario.
- Realizar las vistas necesarias.

### **2.4 Requisitos no funcionales.**

Son propiedades o cualidades que el producto debe cumplir. Especifican criterios que pueden usarse para juzgar la operación de un sistema en lugar de sus comportamientos específicos.

#### **Usabilidad**

- Los usuarios deben poseer conocimientos básicos sobre Base de datos.
- Los usuarios deben poseer conocimiento en análisis y tratamiento de la

información.

- Los usuarios deben recibir capacitación de las herramientas utilizadas en la confección de las Tablas de Salida y cómo interpretar la información de los mismos.
- Tener al menos una persona que domine el trabajo con la herramienta “Pentaho Data Integration” para realizar o modificar las transformaciones para la integración de datos.

### **Fiabilidad**

La integridad de los datos que se manejan es fundamental para el éxito de la solución de almacenamiento y análisis de datos.

- Disponibilidad.
- Tiempo medio de reparación.
- Máximo de Errores.

### **Soporte**

Permite saber las herramientas necesarias para el desempeño óptimo del mercado.

Restricciones de diseño

Indicar cualquier restricción de diseño en el sistema a construir.

### **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.

### **Interfaz**

- Los requerimientos de interfaz de usuario se centran en la presentación de la información de cara al cliente.
- Los reportes estadísticos deben contar con una interfaz simple que facilite la

interacción usuario-aplicación.

- Las interfaces de salida no serán cargadas con información innecesaria.

### **Interfaces Almacén de Datos**

- En el proceso de integración es necesario la utilización de una memoria mínima de 2 GB para el proceso de transformación.
- Se debe contar de un área de almacenamiento intermedio de 500 GB máximo.
- Para la visualización y la inteligencia de negocio se necesita una memoria de 8 GB.
- Las estaciones de trabajo (pc clientes) deben contar con impresoras (para garantizar la impresión de las tablas de salida).

### **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.
- El lenguaje para la programación dentro del repositorio será MDX para realizar consultas a la base de datos e implementar las funciones necesarias.
- El perfilado de datos se realizará con el JavaScript, PentaHo, herramienta libre especializada en estas funcionalidades.

### **2. 5 Casos de uso del sistema.**

Los casos de uso del sistema (CUS) se utilizan para capturar los requisitos de un sistema, proporcionado el espacio de interacción del sistema con el usuario o con otro sistema para lograr un objetivo específico.

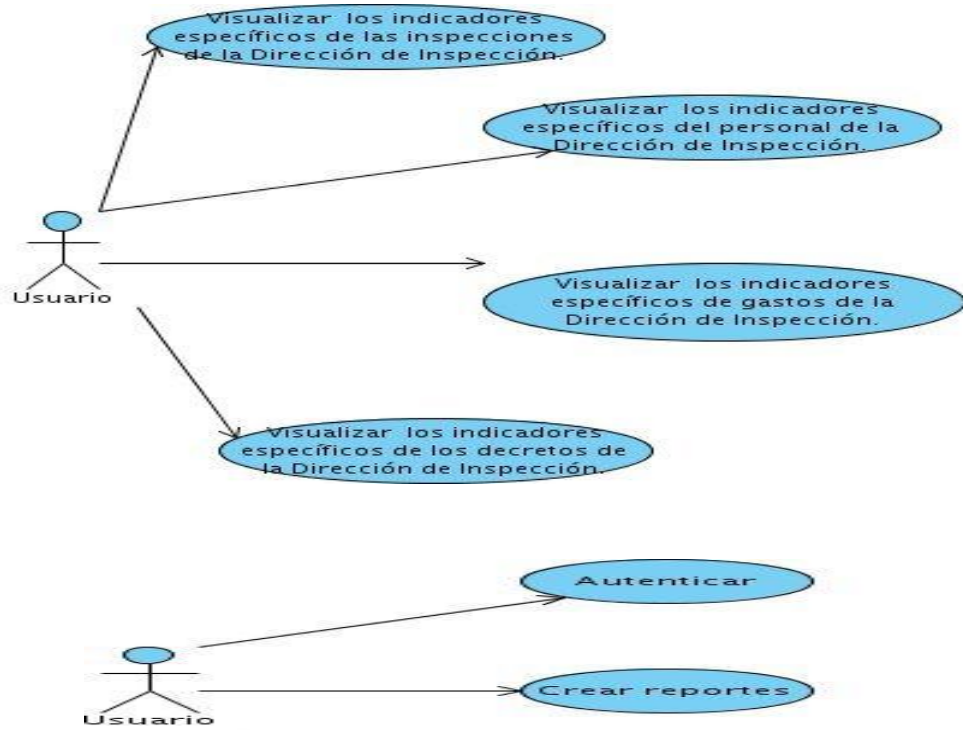


Figura 6: Casos de uso de la Dirección de Inspección

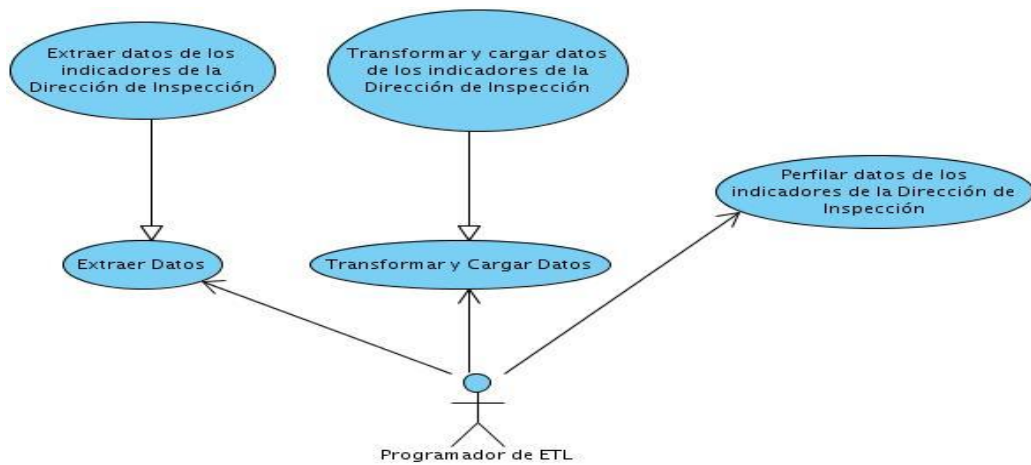


Figura 7: Caso de uso de ETL

Caso de Uso	Descripción
Visualizar los indicadores específicos de las inspecciones de la Dirección de Inspección.	Visualiza todos los reportes de los resultados de las inspecciones.
Visualizar los indicadores específicos del personal de la Dirección de Inspección.	Visualiza todos los reportes del personal.
Visualizar los indicadores específicos de los gastos de la Dirección de Inspección.	Visualiza todos los reportes de los gastos.
Visualizar los indicadores específicos de los decretos de la Dirección de Inspección.	Visualiza todos los reportes de los decretos.
Autenticarse.	Se autentica el usuario en la aplicación.
Gestionar Usuario.	Se insertan y eliminan usuarios.
Gestionar reporte.	Se insertan, actualizan y eliminan reportes.
Extraer datos de los indicadores.	
Transformar y cargar datos de los indicadores.	Realiza la transformación y carga de los datos.
Perfilar datos de los indicadores.	Realiza el perfilado de los datos.

**Tabla 3: Descripción de los casos de uso.**

### **Diseño de la Solución.**

A continuación se abordará todo lo referente al diseño de la propuesta de solución, el cual se convierte en un paso importante para la implementación del sistema, identificándose los puntos necesario para lograr un buen diseño y que se definen a continuación.

**Matriz bus o matriz dimensional.**

La matriz bus, es la relación que existe entre las dimensiones y los hechos del MD. Se define como la habilidad para describir y seguir la vida tanto de una dimensión como de un hecho.

Dimensiones	Hechos																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
dim_actividades	x						x	x														
dim_inciso									x	x	x	x	x	x	x	x	x	x	x	x	x	x
dim_artículo									x	x	x	x	x	x	x	x	x	x	x	x	x	x
dim_cargo						x																
dim_concepto		x																				
dim_inspecciones_internas	x				x																	
dim_municipio	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
dim_temporales	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

**Tabla 4: Relación de hechos con dimensiones.**

**Modelo de Datos**

El modelo de datos permite estructurar y ordenar los datos para cumplir con las necesidades de la organización. Incluyendo las relaciones y agrupación entre las diferentes secciones de datos ordenadamente.

A continuación se muestran unas vistas del modelo de datos para la Dirección de Inspección entre las dimensiones y el hecho para obtener los cargos del personal y de las inspecciones internas en los municipios y la Administración Provincial de Artemisa.

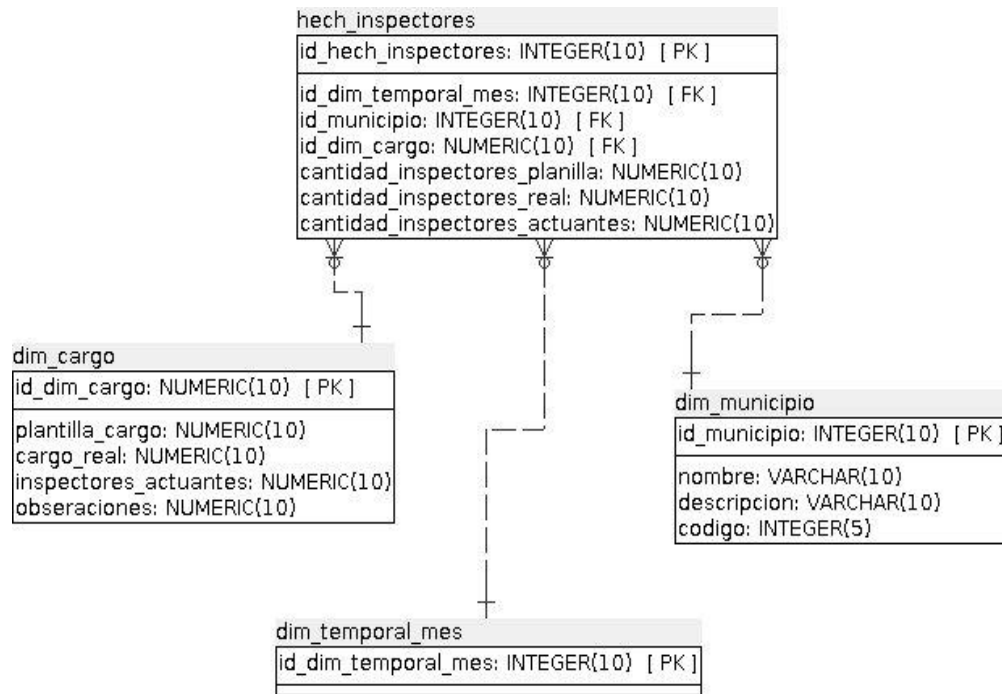


Figura 8: Hechos relacionados con el personal

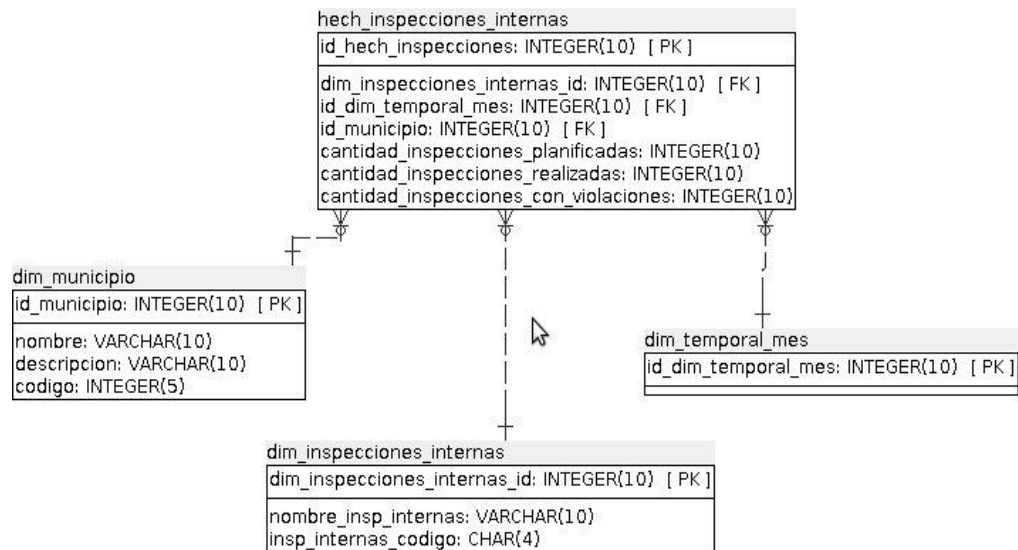


Figura 9: Hechos relacionados con las inspecciones internas

## Dimensiones

Es donde se va a almacenar la información. Estas dimensiones se pueden

encontrar en el artefacto Plantilla Especificaciones del Modelo de Datos Dimensional el cual forma parte del Expediente de Proyecto.

- dim\_actividades
- dim\_inspecciones\_internas
- dim\_articulo
- dim\_concepto
- dim\_cargo
- dim\_municipio
- dim\_temporal\_mes
- dim\_inciso

### **Tabla de Hechos**

Las tablas de hechos son las fuentes de almacenamiento de las medidas numéricas. La tabla de hechos identificada se describe a continuación:

- hech\_actividades
- hech\_gastos
- hech\_inspecciones
- hech\_inspecciones\_internas
- hech\_inspectores
- hech\_organismo
- hech\_organismo\_mlc
- hech\_concepto
- hech\_acuerdo\_4272
- hech\_decreto\_155
- hech\_decreto\_261\_1
- hech\_decreto\_193
- hech\_decreto\_164\_51
- hech\_decreto\_272\_ura
- hech\_decreto\_272\_ci
- hech\_decreto\_272\_17



- hech\_decreto\_272\_18
- hech\_decreto\_211
- hech\_decreto\_227
- hech\_decreto\_274\_tpcp
- hech\_decreto\_275\_10

### **Medidas**

Son las variables de salida del almacén. Estas medidas son encontradas en los artefactos, Especificación de Tablas Hechos y en el documento excel, Herramienta de Recolección de Información.

### **Política de respaldo y recuperación.**

Para garantizar la persistencia de la información y la contribución a no tener almacenada información que no sea útil a los analistas estadísticos de la Dirección de Inspección se realizaran las siguientes estrategias:

- Elaboración de backups (es una forma de guardar la información), con periodicidad mensual, de la información total que posea la base de datos garantizando en todo momento que exista una copia exacta de la información que está vigente en el servidor.
- Periodicidad de las salvas: las salvas de toda la información contenida en la BD se realizan anualmente, así lo define la organización, certificando en todo momento que exista una copia escrita de la información presente en el servidor.
- Tablas involucradas: Las tablas que se involucran en la realización de las salvas son la tabla de hechos y las tablas de las dimensiones pertenecientes a la Dirección de Inspección.

## **2. 6 Subsistemas de integración y visualización.**

### **Subsistemas de integración.**

Para comenzar a construir un software es necesario tener definida una buena arquitectura, se deben tener en cuenta una serie de aspectos o patrones que se

utilicen de guía para la elaboración del mismo.

Para lograr una integración de los datos con resultados satisfactorios, la arquitectura queda definida de la siguiente forma:

-Fuente de Datos: Son archivos de extensión dbf o xls que contienen la información.

-Área temporal: Es el punto intermedio entre la fuente de datos y el MD. Es donde se realiza la integración y transformación de los datos.

-MD: Donde son cargados los datos para su futuro análisis.

### **Subsistemas de visualización.**

El subsistema de visualización permite a los directivos ver de manera consistente la información. Estos se realizan a través de los cubos OLAP en la Inteligencia de Negocios sirviendo de esta manera a la ayuda de la toma de decisiones de la Dirección de Inspección.

### **Conclusiones del Capítulo**

Este capítulo permitió describir la solución mediante el análisis anterior. Se determinaron las reglas del negocio a medir, para lograr una estandarización del envío de la información. También quedaron plasmados los requisitos funcionales, no funcionales y de información, para dar respuesta a las necesidades de los usuarios y conocer las condiciones y propiedades que tanto el sistema como el producto deben cumplir. Se modeló el diseño lógico, definiéndose en los mismo 8 dimensiones y 9 tablas de hechos, así como sus medidas.

## Capítulo 3: Implementación del MD de la Dirección de Inspección.

### Introducción

Con la terminación del proceso de construcción, las pruebas de volumen y carga, resultan tan importantes como el diseño y la implementación misma. El MD al entrar en contacto con los usuarios finales, iniciando un ciclo iterativo e incremental, de lo simple a lo complejo, donde el sistema nunca descansará puesto que a él son adheridos, con el transcurso del tiempo, nuevos años de información, procesos de negocios de la empresa, nuevas necesidades o insatisfacciones del cliente. También se lleva a cabo la capa de Inteligencia de Negocio donde se implementarán los reportes que responden a las necesidades de la Dirección de Inspección, lo cual finalizará con la visualización de tablas y gráficas para su posterior análisis.

### 3. 1 Implementación de la base de datos.

En el desarrollo de un proyecto es de vital importancia el modelado de datos, ya que en el mismo reside la esencia de la aplicación. Al diseñar el modelo dimensional, surgen sus relaciones entre sus tablas. Una buena estructura traerá consigo una aplicación robusta con resultados óptimos.

#### Estructura de los datos.

Una estructura de datos es la forma de organizar los datos con el objetivo de facilitar su manejo, por tanto, para el desarrollo del MD están estructurados por tablas y esquemas, donde los esquemas están organizados por hechos y dimensiones y las tablas están organizadas por políticas de indexado e índices bien definidos.

Esquemas	Tablas
dimensiones	dim_temporal_mes
dimensiones	dim_municipio
mart_inspeccion	dim_actividades

mart_inspeccion	dim_inciso
mart_inspeccion	dim_articulo
mart_inspeccion	dim_concepto
mart_inspeccion	dim_inspecciones_internas
mart_inspeccion	hech_inspectores
mart_inspeccion	hech_organismo
mart_inspeccion	dim_cargo
mart_inspeccion	hech_actividades
mart_inspeccion	hech_gastos
mart_inspeccion	hech_acuerdo_4272
mart_inspeccion	hech_inspecciones
mart_inspeccion	hech_inspecciones_internas
mart_inspeccion	hech_organismo_mlc
mart_inspeccion	hech_concepto
mart_inspeccion	hech_decreto_155
mart_inspeccion	hech_decreto_ura
mart_inspeccion	hech_decreto_211
mart_inspeccion	hech_decreto_164_51
mart_inspeccion	hech_decreto_193
mart_inspeccion	hech_decreto_261_1
mart_inspeccion	hech_decreto_272_17
mart_inspeccion	hech_decreto_272_18
mart_inspeccion	hech_decreto_272_ci
mart_inspeccion	hech_decreto_274_tpcp
mart_inspeccion	hech_decreto_227
mart_inspeccion	hech_decreto_275_10

**Tabla 5: Relación de los Esquemas y tablas**

### 3. 2 Implementación de los subsistemas de integración

Extraer, transformar y cargar (Extract, Transform and Load), es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos, limpiarlos y cargarlos en otra base de datos, MD, o un almacén de datos para analizar, o en otro sistema operacional para apoyar un proceso de negocio.

**Extraer:** Permite extraer los datos desde el sistema origen. Una parte intrínseca del proceso de extracción es la de analizar los datos extraídos, de lo que resulta un chequeo que verifica si los datos cumplen la pauta o estructura que se esperaba. De no ser así los datos son pasados a otro lado para que se revisen.

**Transformar:** En esta fase se aplican una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. Algunas de las funciones que se realizan son:

- Seleccionar datos deseados.
- Partir campos de preferencia.
- Buscar en bases de datos.
- Calcular totales de múltiples filas.

**Carga:** En esta fase es donde los datos de la fase anterior son cargados hacia el sistema de destino. Existen dos formas básicas de desarrollar el proceso de carga:

Acumulación simple: La acumulación simple es la más sencilla y común, y consiste en realizar un resumen de todas las transacciones comprendidas en el período de tiempo seleccionado y transportar el resultado como una única transacción hacia el AD almacenando un valor calculado que consistirá típicamente en un sumatorio o un promedio de la magnitud considerada.

Rolling: Se aplica en los casos en que se opta por mantener varios niveles de granularidad. Para ello se almacena información resumida a distintos niveles, correspondientes a distintas agrupaciones de la unidad de tiempo o diferentes niveles jerárquicos en alguna o varias de las dimensiones de la magnitud

almacenada (por ejemplo, totales diarios, totales semanales, totales mensuales). (INMON, 2012)

Para la implementación de este MD se utilizó la forma de carga rolling debido a que se trabaja con niveles jerárquicos en distintas dimensiones y se almacenan los datos a

A continuación se ilustran algunos ejemplos de las transformaciones realizadas para poblar el mercado correspondiente.

En esta transformación se carga el hecho perteneciente a las actividades realizadas por la Dirección de Inspección durante el mes. Se utiliza una entrada Excel, se capturan los datos donde se escogen los necesario para la transformación, se utiliza una constante donde se almacena la fecha para ser buscada en la base de datos, a continuación se identifica el municipio y el nombre de las actividades, finalmente se filtran los datos validando que no se encuentren nulos y se almacenan en la base de datos en caso de que todos los datos estén correctos, de lo contrario los devuelve para un Excel auxiliar.

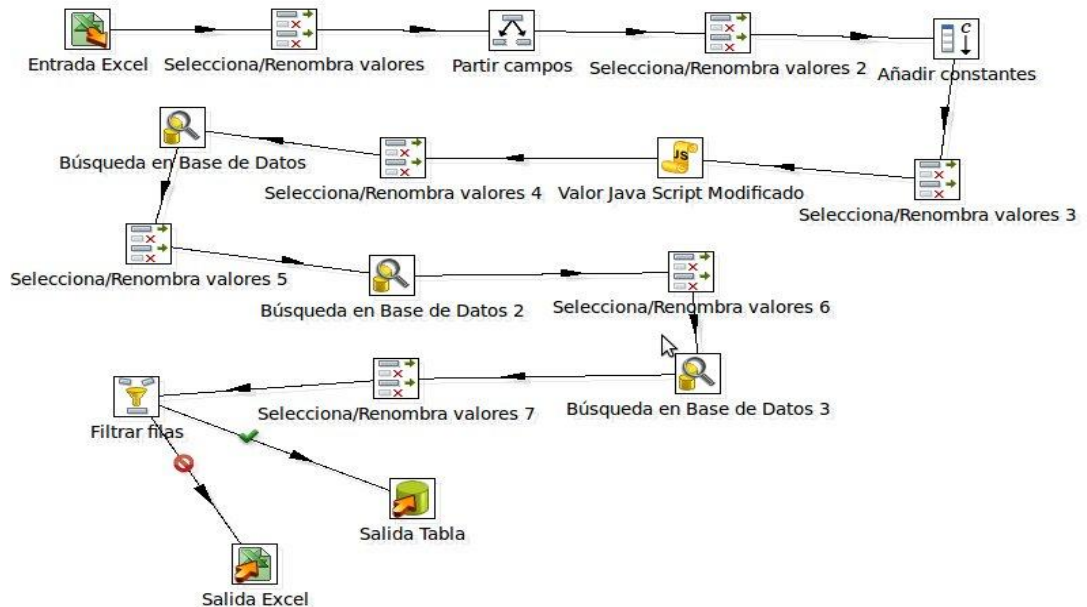
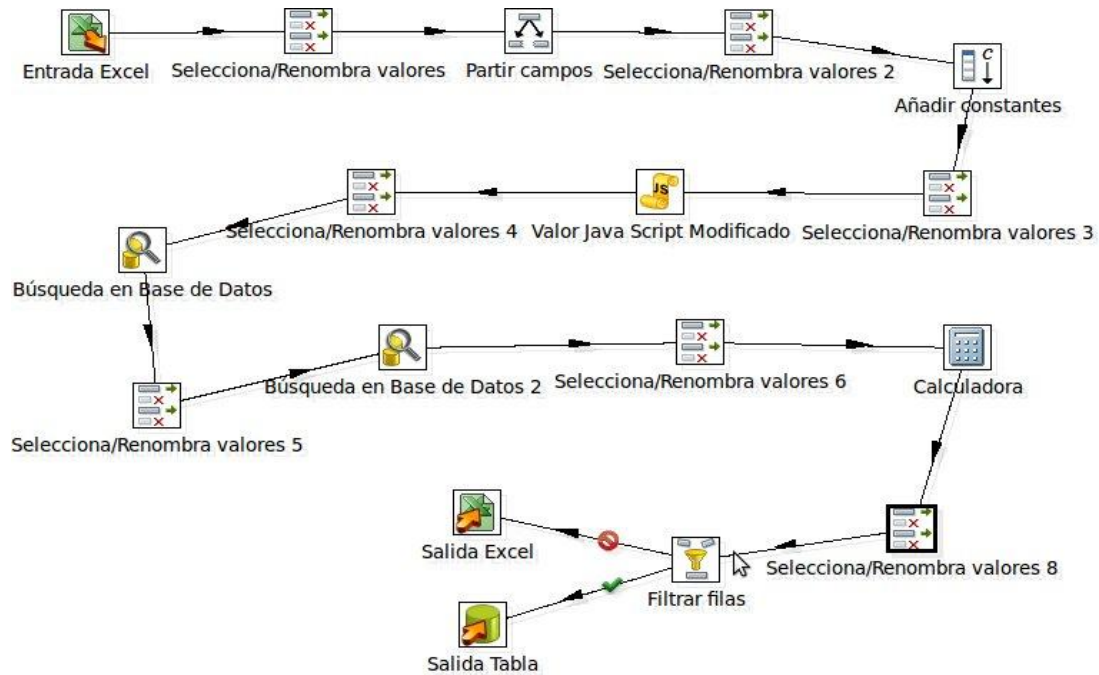


Figura 10: Transformación del Hecho de Actividades

Se le da entrada a los datos por el componente de entrada de Excel, se seleccionan los campos con los cuales se va a trabajar. Se validan los datos y se buscan en las distintas tablas de la base de datos como son la fecha y el municipio, a través del componente calculadora se suman dos cantidad y finalmente se filtran los datos y se exportan para la base de datos en caso de que estén correcto de lo contrario se almacenarán en un excel auxiliar.



**Figura 11: Transformación del Hecho de Gastos.**

### 3.3 Implementación de los trabajos.

A la vez que la conexión del MD se encuentre lista se realizan los trabajos (job), los cuales se encargan de cargar los datos hacia el mercado en un período de tiempo.

Para la implementación de un job se debe definir correctamente las dimensiones estáticas pues los job solo cargan dimensiones que sus datos sean variables, tanto nuevos como actualizar los que ya existían.

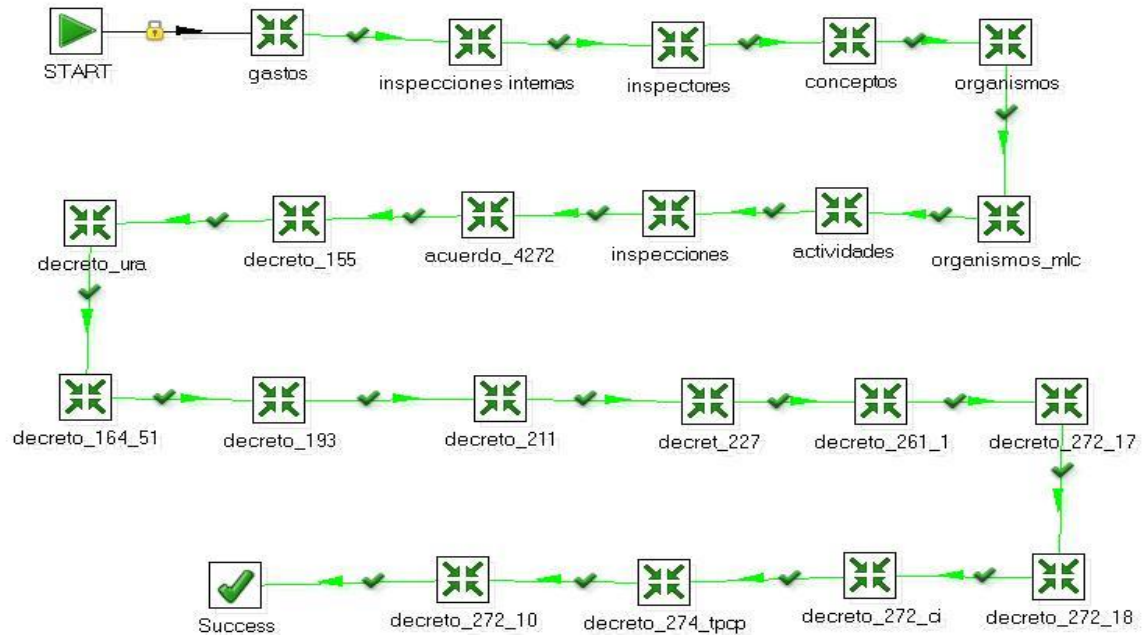


Figura 12: Ejemplo de un job.

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

#### Cubos OLAP

Los cubos OLAP se utilizan para realizar consultas rápidas y agrupar los datos para un mejor análisis de los datos, para obtener los datos más importantes entre toda la información que posee el MD.

Se necesita crear cubos multidimensionales para la implementación de los reportes OLAP, los cuales se realizan mediante la herramienta Pentaho Schema Workbench, la cual genera ficheros de configuración XML. En este fichero de esquema se pueden definir las dimensiones, los niveles de jerarquía de dimensiones, los hechos y conexión con el almacén que sirve los datos para el cubo OLAP.

Se modelaron 9 cubos multidimensionales, en los mismos se especificaron las dimensiones. La siguiente imagen muestra el diseño utilizando la herramienta Pentaho Schema Workbench de unos de los cubos modelados para el control de



los gastos, el cual está formado por sus dimensiones, y medidas.

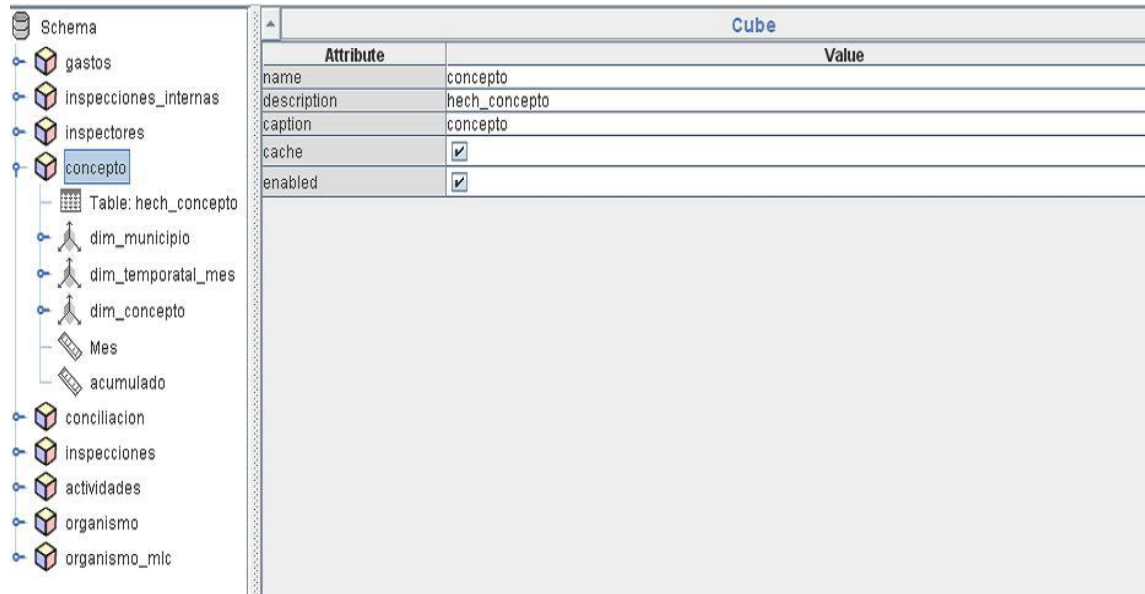


Figura 13: Ejemplo de workbench

Navegación de la capa de visualización.

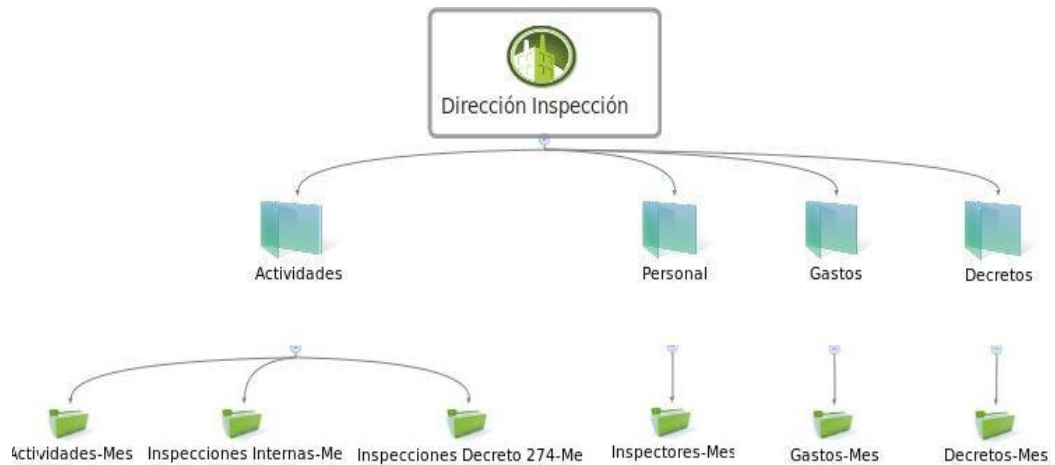


Figura 14: Mapa de Navegación

El mapa de navegación es una representación gráfica donde se encuentra información y la forma de acceder a ella de manera organizada. El MD de la Dirección de Inspección tiene un mapa de navegación compuesto por 4 áreas de análisis, cada una con sus respectivos libros de trabajo y sus reportes.

Áreas de Análisis:

- Análisis de los indicadores de las inspecciones de la Dirección de Inspección.
- Análisis de los indicadores del personal de la Dirección de Inspección.
- Análisis de los indicadores de los gastos de la Dirección de Inspección.
- Análisis de los indicadores específicos de los decretos de la Dirección de Inspección.

Cuando se accede al reporte deseado se pueden observar las vistas de análisis generadas, la siguiente figura muestra la vista de análisis correspondiente al cubo perteneciente al hech\_acuerdo\_4272.

			Medidas	
Municipios	Tiempo	Incisos	● Cantidad Mensual	● acumulado_anterior
+ Municipio	- Fecha	+ Inciso	36,00	50,00
	+ 2011	+ Inciso	36,00	50,00

Figura 15: Ejemplo de visualización

### Conclusiones del Capítulo.

En el presente capítulo se analizó la fuente de datos para lograr implementar el MD, se obtuvo la estructura física del MD, se diseñó el esquema multidimensional con sus respectivos cubos OLAP para agrupar los datos y así facilitar su análisis posteriormente, se identificaron las áreas de análisis, los libros de trabajo y los reportes candidatos, se implementaron y visualizaron los reportes correctamente.

## Capítulo 4. Validación del MD de la Dirección de Inspección del Consejo de la Administración Provincial de Artemisa.

### Introducción

En el presente capítulo se aplicarán los casos de prueba para validar la calidad de los datos cargados, se validan las entradas de datos, verificando que los resultados obtenidos se correspondan con los resultados esperados viendo que esto cumpla con las necesidades de los usuarios de la Dirección de Inspección de la Administración Provincial de Artemisa.

### 4.1 Pruebas

Las pruebas son un conjunto de actividades en las cuales un sistema o componente es ejecutado bajo unas condiciones o requerimientos específicos, donde los resultados son observados y registrados para dar una evaluación de algún aspecto del sistema o componente que es evaluado y determinar la calidad del mismo. En cada versión operacional del sistema se realiza una comprobación durante la integración del mismo, describiendo también como hacer las pruebas al sistema, verificando que todos los requerimientos hayan sido implementados, determinado los defectos del mismo.

Las pruebas que se van a realizar al producto una vez finalizado son los casos de pruebas basados en casos de uso los mismos representan un conjunto de condiciones o variables bajo las cuáles el analista del sistema determina si el requisito de una aplicación es parcial o completamente satisfactorio.

### 4.2 Diseño de los casos de prueba

El diseño de los caso de prueba se realiza para determinar si los requisitos o los caso de uso son completamente satisfactorios. Para el mercado de datos de Inspecciones se diseñaron 4 casos de prueba, uno por cada caso de uso.

Escenario	Descripción	Perfiles de análisis	Indicadores a medir	Respuesta del sistema	Flujo central
-----------	-------------	----------------------	---------------------	-----------------------	---------------

<p>EC1.1 9,R-10</p>	<p>R- Muestra los reportes de los indicadores del personal de inspecciones.</p>	<p>Año</p>	<p>cantidad_plantilla, cantidad_reportes al, cantidad_actuantes</p>	<p>El sistema muestra los reportes de los indicadores de los inspectores</p>	<p>Se abre la aplicación. Se autentifica. Se entra al sistema. En el lateral izquierdo donde aparece el navegador se selecciona las áreas de análisis de <b>A.A</b> <b>Análisis de los indicadores del personal.</b> Se selecciona el libro de trabajo de correspondiente a lo que se desea</p>
-------------------------	---	------------	---	--	---

				buscar. Libro de Trabajo: Análisis de los indicadores de los inspectores. cantidad_pla ntilla,mes, año,cantidad _real,cantida d_actuantes
--	--	--	--	---

**Tabla 6: Casos de Prueba**

### 4.3 Plan de Pruebas.

En este artefacto se describen los roles y las responsabilidades en el proceso de validación del MD Dirección de Inspección. Se realiza el diagrama de Despliegue donde se describen las partes que componen el sistema bajo prueba, y se da un acercamiento a los recursos del sistema.

### Estrategia de Pruebas

#### Prueba de acceso a la información por roles o niveles de usuario.

##### Descripción:

Se verifica el acceso y privilegios que posee cada usuario o rol de la Dirección de Inspección, definido con respecto al sistema.

##### Precondiciones:

- Acceso restringido a la información, solo se puede ver la información necesaria del mercado 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.

**Medida:**

Cumple o no cumple con los privilegios de acceso.

**Prueba de Consistencia de los Datos.**

**Descripción**

Se insertan un número finito de datos a través de proceso de ETL en las tablas de cada hecho, buscando que se ejecute el ciclo completo hasta la capa de visualización, donde se verifica que los datos insertados sean los mismos que los visualizados en la aplicación sistema de estadísticas de la Administración Provincial del usuario final.

**Precondiciones**

- Datos confiables, sin existencia de duplicidad, u otros errores de cálculos en las estructuras

**Medida:**

- Eficiente: el conjunto de datos de entrada debe ser el mismo que el de salida.
- Parcial: el conjunto de datos de salida es el 80 % del conjunto de datos de entrada.
- Deficiente: el conjunto de datos de salida es menor del 80 % del conjunto de datos de entrada.

**Prueba de las estructuras de pre-visualización o estructuras multidimensionales.**

### **Descripción**

Se verifica que las estructuras construidas (cubos, etcétera) respondan a todo el alcance de información que fue definido en el negocio por los clientes y analistas de la Dirección Inspección.

### **Precondiciones**

- Estructuras sólidas que respondan a las necesidades de información que se definieron en el negocio.

### **Medida**

- Eficiente: Las estructuras cumplen con el alcance de información previsto
- Parcial: Las estructuras cumplen en un 80% con el alcance de información previsto
- Deficiente: Las estructuras cumplen menos del 80% con el alcance de información previsto.

## **Prueba de capacitación a los usuarios sobre el manejo de los metadatos.**

### **Descripción**

Se verifica que los usuarios tengan pleno conocimiento sobre el manejo de las herramientas de visualización y el manejo de las estructuras multidimensionales.

### **Precondiciones**

- Usuarios con pleno dominio de las herramientas de visualización y análisis de información, y de las estructuras multidimensionales

### **Medida**

Después de efectuar el periodo de capacitación se les realizan pruebas a los usuarios para verificar el conocimiento alcanzado.

- Eficiente: Usuarios que cumplen con el 100% de los conocimientos adquiridos en la capacitación.
- Parcial: Usuarios que cumplen con el 80 % de los conocimientos adquiridos en la capacitación.
- Deficiente: Usuarios que cumplen con menos del 80% de los conocimientos adquiridos en la capacitación.

### **Prueba de respaldo de la información de los metadatos.**

#### **Descripción**

Verificar que el sistema de respaldo de la estructura multidimensional es completamente efectivo.

#### **Precondiciones**

Con una frecuencia determinada se realizan copias de respaldo del repositorio de metadatos.

#### **Medida**

Cumple o no cumple

### **Pruebas de la extracción, carga y transformación de los datos (ETL).**

Se realizan en los procesos de ETL.

#### **Objetivo**

El objetivo fundamental para las pruebas relacionadas con el proceso ETL es que los datos que se almacenarán en el destino, en este caso en el MD para la Dirección Inspección, se encuentren de la manera esperada, con la calidad, la validez y la completitud requeridas.

#### **Técnica**

La técnica a utilizar será la ejecución de consultas en lenguaje SQL para lograr el objetivo esperado de las Pruebas. Esta técnica se puede realizar cíclicamente.



### **Entorno de prueba**

El entorno en el cual se desarrollarán las pruebas será a través del gestor Postgres, específicamente el PgAdmin III, bajo el SO Linux, en la distribución que se utilice, ya sea Ubuntu, Debian, u otra.

### **Proceso**

El proceso estará compuesto por 3 pruebas fundamentales, que utilizarán código SQL para seleccionar datos del MD y realizar las verificaciones necesarias y previstas por las pruebas.

### **4.4 Resultados obtenidos.**

Como resultado de este trabajo el MD de Inspección queda disponible en su versión 1.0. Obteniéndose un sistema que cumple con todas las funcionalidades descritas. Además el presente MD cuenta con la validación del producto emitido por la Facultad. (Anexo 2)

### **Conclusiones del Capítulo**

En el presente capítulo con el objetivo de mejorar la organización y la ejecución de las pruebas se diseñaron y aplicaron casos de pruebas para validar la calidad del mercado de datos de la Dirección de Inspección. También se estableció el plan de pruebas para verificar la calidad del proceso de implementación.

### **Conclusiones Generales**

Al concluir este trabajo se puede plantear que se ha cumplido con los objetivos del mismo así como con las tareas de la investigación. A partir de esta investigación donde se realiza la creación de un MD para la Administración Provincial de Artemisa relacionado con los indicadores sobre las inspecciones se arribó a las siguientes conclusiones:

- La definición del marco teórico desarrollado propició un total entendimiento de las tendencias de la implementación de los AD y los MD.
- El análisis de los documentos Excel relacionados con las inspecciones permitió que se identificaran eficientemente todos los procesos, requisitos funcionales e informativos.
- El diseño del MD para los indicadores sobre las inspecciones se desarrolló satisfactoriamente obteniéndose un modelo que soporta los requisitos definidos para el almacén de datos de la Administración Provincial de Artemisa.
- La implementación y carga de los clasificadores produjo una integración más completa de la información y una mejor disponibilidad de la misma.
- La validación de la solución desarrollada mediante la realización de pruebas demuestra que el sistema cumple con las expectativas del cliente.

### **Recomendaciones**

Se recomienda que para futuras versiones se le realice al MD para la Dirección de Inspección la integración con la base de datos.

Se recomienda su implementación en otras Administraciones Provinciales.

### Referencias Bibliográficas

- CURTO, J.** *Diseño de un data warehouse: estrella y copo de nieve* «[Information Management](#) .ed. Última actualización: November 19, 2008. ().  
Disponible en:<http://informationmanagement.wordpress.com/2007/11/19/dise-no-de-un-data-warehouse-estrella-y-copo-de-nieve/>.
- ETL-TOOLS.INFO.** *Esquema de la estrella* .ed. Última actualización: 2011. ().  
Disponible en:[http://etl-tools.info/es/bi/almacenedatos\\_esquema-estrella.htm](http://etl-tools.info/es/bi/almacenedatos_esquema-estrella.htm).
- INMON.** *Data mart* .ed. Última actualización: 2012. (). Disponible en:  
<http://es.scribd.com/doc/52805589/Data-mart>.
- SINNEXUS.** *Datawarehouse*. ed. Última actualización: 20011 2008a. (). Disponible en: [http://www.sinnexus.com/business\\_intelligence/datawarehouse.aspx](http://www.sinnexus.com/business_intelligence/datawarehouse.aspx).
- MOLAP, ROLAP, HOLAP** ed. Última actualización: 2011 2008. (). Disponible en: [http://www.sinnexus.com/business\\_intelligence/olap\\_avanzado.aspx](http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx).

### Bibliografía.

**Zepeda Sánchez**, Leopoldo. Tesis. [En línea] [http://tesis.com.es/autores/zepeda-sanchez-Leopoldo\\_zenaido/](http://tesis.com.es/autores/zepeda-sanchez-Leopoldo_zenaido/).

**Bussiness Inteligence** 2007. [En línea] 2007 **Inmon, Bill** 2007. [En línea] 2007 <http://www.kimballgroup.com>

<http://www.bi-bestpractices.com>

**Enrique Hernández Orallo**. El lenguaje Unificado de Modelado (UML). Departamento de Informática. Universidad Politécnica de Valencia. <http://www.disca.upv.es/enheror/pdf/ActaUML.PDF>.

Herramientas CASE. Instituto Nacional de estadísticas e Informática. <http://informática.gonzalonazareno.org/file/php/8/case.pdf>.

**Liudmila Padrón Torres**. 2009. Almacenes de datos. <http://www.zonaeconomica.com/informática/almacenesdedatos>

**Repinosa Mila**. 2010. Kimball vs Inmon. <http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimENSIONAL>.

Pruebas de software <http://pruebasdesoftware.com/laspruebasdesoftware.htm>

**Garia, Maria Isabel Guzmán**. Implementación de un datawarehouse para el soporte de toma de decisiones

Data warehouse. Manual para la construcción. <http://www.elprisma.com>

Análisis de Dimensiones y Hechos. Modelo lógico <http://churriwifi.wordpress.com/2010/04/22/15-3-analisis-dimensiones-hechos>.

**Herrera, Cristian**. Todo lo que querías saber sobre Data Warehouse. 2010. <http://adictosaltrabajo.com/tutoriales>.

**Martinez, Manuela de Jesus Bonilla**. 2009. SISTESIS DE ANALYSIS SERVICES SQL SERVER. [En línea] 2009. <http://es.scribd.com>.

**Mondrian**. Mondrian. [En línea] <http://jasperserver.sourceforge.net/docs/3-7-0/Mondrian-3.0-Technical-Guide.pdf>.

**Negocios, Inteligencia de**. 2009. Inteligencia de Negocios. [En línea] 2001.

<http://www.idensa.com>.

**Paradigm. 2010.** Boost Productivity with innovative and intuitive technologies. [En línea] 2010. <http://www.visual-paradigm.com/product/?favor=vpuml>.

**Pentaho.** Pentaho Data Integrator. [En línea] <http://sourceforge.net/projects/pentaho/files/Data%20Integration>.

**Pgadmin. 2009.** Pgadmin. [En línea] 2009. <http://www.pgadmin.org>.

**PostgreSQL. 2010.** Tinysofa Copyright-grupo de desarrollo global de postgresQL. [En línea] 2010. <http://www.postgresql.org>.

**Rivadera\*, Gustavo R. 2010.** La metodología de Kimball para el diseño de almacenes de. [En línea] 2010. <http://www.ucasal.net/templates/unid-academicas/ingenieria/apps/5-p56-rivadera-formateado.pdf>.

### **Glosario de Términos**

BI: Business Intelligence del español Inteligencia del Negocio.

DSS: Decision Support Systems, Sistemas de Ayuda a la toma de Decisiones.

EIS: Executive Information Systems, Sistemas de Información para Directivos.

ETL: Extracción, Transformación y Carga de datos.

Excel: Aplicación para manejar hojas de cálculo.

HOLAP: On line Analytical Processing Hybrid del español Procesamiento Analítico en Línea Híbrido.

JAVA: Lenguaje de programación orientado a objetos.

WebSocket: Marco de trabajo de código abierto en tiempo real.

MOLAP: On line Analytical Processing Multi Dimensional del español Procesamiento Analítico en Línea Multidimensional.

OLAP: On line Analytical Processing del español Procesamiento Analítico en Línea.

Procesos batch: es un ciclo donde se procesa información sin que intervenga el usuario.

ROLAP: On line Analytical Processing Relational del español Procesamiento Analítico en Línea Relacional.

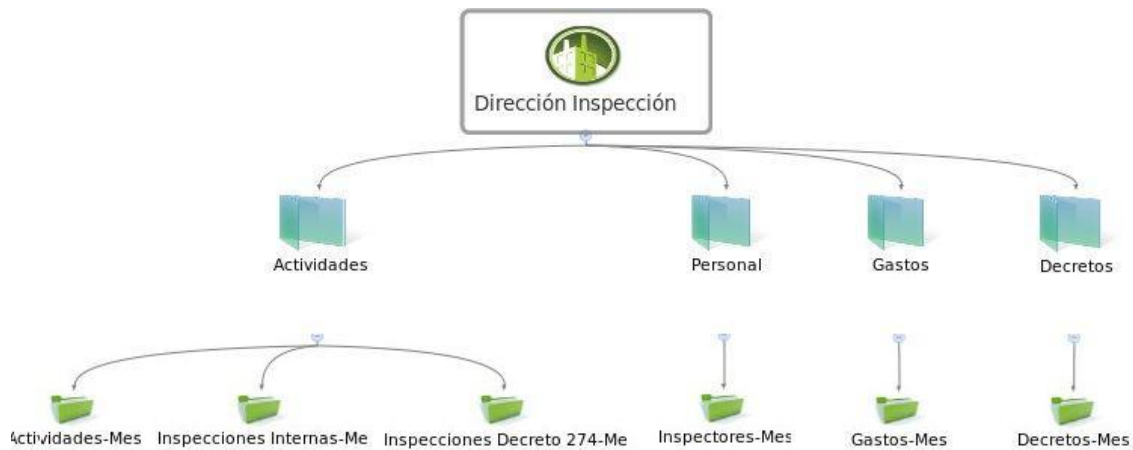
Symfony: Framework diseñado para optimizar el desarrollo de las aplicaciones web.

TIC: Tecnologías de la información y la comunicación.

UCI: Universidad de las Ciencias Informáticas.

**Anexos**

Anexo 1 Mapa de navegación



Anexo 2 Aval de calidad



**Aval de Calidad de Software**


El grupo de Calidad de Software del Centro de Desarrollo de la Facultad Regional "Mártires de Artemisa" conformado por:

- **Asesora de Calidad:** Ing. Maidel Ojeda Cruz
- **Asesor de Tecnología:** Ing. Domma Moreno Dager
- **Especialista de Calidad:** Ing. Yenisleydi Rodríguez Martínez

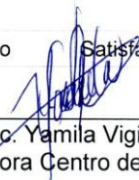
emite el presente **Aval de Calidad de Software** en colaboración con los especialistas del Centro de Desarrollo a: Alejandro Heredia Molina, como resultado satisfactorio de su desempeño en la tareas asociadas al proyecto: Almacén de Datos para la Administración Provincial de Artemisa.

Para emitir el presente aval se valoraron un conjunto de elementos evaluados de manera individual teniendo en cuenta los parámetros de calidad de software del proyecto. A continuación se presenta los resultados en cada uno de los aspectos valorados:

Elementos evaluados	Resultado
Validar la coincidencia entre requisitos informativos y estructuras de visualización.	Satisfactorio
Validar el correcto funcionamiento de las estructuras de visualización.	Satisfactorio
Validar el cumplimiento de la arquitectura de la información.	Satisfactorio
Validar estructura y funcionamiento de los metadatos y reportes ad-hoc.	Satisfactorio
Validar estructura y funcionamiento de los esquemas de mondrian.	Satisfactorio
Verificar que los reportes brindaran información coherente.	Satisfactorio
Validar el funcionamiento interno de las vistas materializadas.	Satisfactorio
Validar el correcto funcionamiento de los procesos de ETL.	Satisfactorio
Validar el tratamiento de errores en los procesos de ETL.	Satisfactorio
Generación de todos los artefactos de la metodología Kimball.	Satisfactorio
Cumplimiento de las plantillas establecidas para cada artefacto.	Satisfactorio
Ortografía, concordancia y redacción de cada uno de los artefactos.	Satisfactorio
Efectividad de los casos de pruebas definidos en el proyecto	Satisfactorio

  
Ing. Maidel Ojeda Cruz  
Asesora de Calidad de SW



  
Msc. Yamila Vigil Regalado  
Directora Centro de Desarrollo