



*Universidad de las Ciencias Informáticas*

***Facultad Regional “Mártires de Artemisa”***

***Título: Mercado de datos para la Sección de Colaboración de la Administración Provincial de Artemisa.***

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

***Autoras: Mileyvi Sosa Ricardo y Yisell Travieso Balbuena***

***Tutora: Ing. Lissandra Vigoa García***

***Co Tutor: Lic. Silvano Merced Len***

***Artemisa, Julio 2012, “Año 53 de la Revolución”***



*Todo y cada uno de nosotros paga puntualmente su cuota de sacrificio consciente de recibir el premio en la satisfacción del deber cumplido, conscientes de avanzar con todos hacia el Hombre Nuevo que se vislumbra en el horizonte.*

*Ernesto Che Guevara*

## *Declaración de Autoría*

---

### **Declaración de Autoría.**

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

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

\_\_\_\_\_  
Autora: Mileyvi Sosa Ricardo

\_\_\_\_\_  
Autora: Yisell Travieso Balbuena

Tutor: \_\_\_\_\_

Ing. Lissandra Vigoa García

### **Datos de Contacto.**

**Tutor:** Lissandra Vígoa García

Especialidad de graduación: Ingeniera en Ciencias Informáticas

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

**Autores:** Mileyvi Sosa Ricardo

Yisell Travieso Balbuena

**Correo Electrónico:** [msosa@hab.uci.cu](mailto:msosa@hab.uci.cu)

[ytravieso@hab.uci.cu](mailto:ytravieso@hab.uci.cu)

## *Agradecimientos*

### **De Yisell:**

*A mi querida abuela, sin ti este sueño no hubiera sido posible, gracias por tu dedicación, cariño y sacrificio todos estos años.*

*A mis padres por su amor y confianza, especialmente a mi mamá que me ha apoyado en cada momento, gracias por todo tu esfuerzo, preocupación y por estar siempre cuando más lo he necesitado.*

*A mi hermana Yinett por estar siempre a mi lado y por hacerme reír.*

*A mis tías Magda, Miriam y mi prima Yanet por todo el apoyo y cariño que siempre me han depositado.*

*Al resto de mi familia, que sepan que los quiero mucho.*

*Quiero agradecer muy especialmente a mi novio Gianni por toda su ayuda, amor y por acompañarme en las buenas y en las malas, te quiero.*

*A mi fiel compañera de tesis por su preocupación y confianza durante el desarrollo de la tesis, gracias por ser mi amiga.*

*A todos los profesores que de una forma u otra me ayudaron y apoyaron siempre que lo necesité entre ellos mi co-tutor Silvano.*

*A mi grupo 2, juntos pasamos momentos muy buenos, nunca los voy a olvidar.*

*A todos mis amigos que compartieron conmigo en la universidad y me brindaron su apoyo.*

*A todos los que contribuyeron de una forma u otra en la elaboración de este trabajo, les agradezco de todo corazón.*

## **De Mileyvi:**

*Resulta bastante difícil concentrar en tan solo una página a tantas personas que de una manera u otra han dado a mi vida un sentido. Aún así haré el noble intento de escribir.*

*A mi tesoro como ella me dice, a mi mami que siempre ha estado a mi lado apoyándome en cada decisión que tome, en cada momento de llanto y alegría, ella que supo educarme y guiarme por el camino del saber, sin ella este sueño no se hubiera realizado. Gracias por ser mi madre, mi padre, mi hermana, mi amiga, gracias por tu amor, dedicación y esfuerzo todos estos años.*

*A mi esposo, él que me dio fuerzas para seguir adelante en momentos difíciles justo cuando pensaba que todo estaba perdido.*

*A Raúl por tener mucha paciencia conmigo en el transcurso de todos estos años.*

*A mi familia especialmente a mis tías Mirtha y Mamito, por creer en mí y tener tanta fe en que lo lograría.*

*A la familia de mi novio que me ha acogido como si fuera uno de ellos, por su constante preocupación, por todos sus consejos y poder contar con su apoyo. Gracias por estar presente.*

*A todas mis amistades.*

*A mi cotutor el Lic. Silvano Merced Len gracias por ayudarnos y apoyarnos siempre.*

*A los que de una forma u otra me incitaron a ser mejor cada día.*

*A todas las personas que de una forma espontánea y amable soportaron las molestias para transitar este camino, a todos ellos de corazón muchas gracias y mis más eternos agradecimientos.*

*Por último pero no menos importante a cierta persona nerviosa y tímida a la cual conocí en segundo año y sin querer se convirtió en una amiga más para mí y que este curso fue mi compañera de tesis a Yisell la cual me ha soportado en mis momentos de intranquilidad, gracias por todo.*

## *Dedicatoria*

### **De Mileyvi:**

*Quiero dedicar este trabajo a la persona más especial de mi vida a mi madre porque todo lo que soy se lo debo a ella, por ayudarme en todo momento y confiar siempre en mí. Aquí está el fruto de su sacrificio y amor inmenso.*

*A mi esposo por brindarme tanto amor.*

### **De Yisell:**

*A mis padres y mi abuela por estar a mi lado siempre y ser los ejemplos en mi vida pues sin su apoyo no hubiese sido posible que estuviera presentando esta tesis.*

*A toda mi familia que de una forma u otra se han preocupado por mí todos estos años.*

### Resumen

El presente trabajo de diploma propone la implementación de un mercado de datos para la Sección de Colaboración de la Administración Provincial (AP) de Artemisa que agilice los procesos de consulta de información, permitiendo su análisis desde diversas perspectivas y con grandes velocidades de respuesta. Surge por la necesidad que presenta dicha sección de poder almacenar datos de manera eficiente, permitiendo obtener los resultados esperados para el proceso de toma de decisiones. Para el desarrollo de este mercado de datos se tomó como base la metodología de Ralph Kimball, y las herramientas utilizadas fueron: Visual Paradigm v6.4, PowerArchitect v0.9.13, PostgreSQL v9.1, PgAdmin III v1.12.0, Pentaho Data Integration v4.1, SchemaWorkbench v3.2.1, Pentaho BI Server v3.6, Pentaho Report Design v3.7.0 y Pentaho Metadata Editor v3.7.0. Como resultado se obtiene la estructura dimensional que contiene las dimensiones, jerarquías, niveles, atributos, tablas de hechos y medidas que garantizan los análisis estadísticos. Igualmente se detalla el negocio permitiendo identificar las reglas del mismo. Se realiza el proceso de carga de los datos de las fuentes de la Sección de Colaboración al mercado de datos presentado, se incluyen las estrategias de seguridad, respaldo y recuperación de datos y también se efectúan pruebas que validan la solución en cuestión.

**Palabras Claves:** Mercado de Datos, AP (AP).

## Índice de Contenido

Resumen .....	VIII
Introducción.....	1
CAPÍTULO 1: “Fundamentación teórica de los sistemas dedicados a la toma de decisiones en el mundo.” .....	9
1.1 Estado del Arte.....	9
1.2 Principales conceptos .....	13
1.3 Modos de Almacenamiento de Datos .....	17
1.4 Metodología .....	18
1.5 Modelo Dimensional .....	19
1.6 Herramientas utilizadas para el desarrollo del Mercado de datos.....	21
1.7 Conclusiones del capítulo .....	26
CAPÍTULO 2: “Análisis y diseño del mercado de datos de la Sección de Colaboración de la AP” .....	28
2.1 Análisis.....	28
2.1.1 Especificación de Requerimientos .....	28
2.1.2 Arquitectura .....	34
2.1.3 Definición de las Áreas de Análisis.....	36
2.1.4 Diagrama de Casos de Uso del Sistema.....	37
2.2 Diseño.....	38
2.2.1 Modelo de datos.....	38
2.2.2 Matriz Bus .....	45
2.2.3 Diseño del subsistema de integración.....	46
2.2.4 Diseño de los subsistemas de visualización de datos.....	46
2.3 Conclusiones del capítulo .....	47
CAPÍTULO 3: “Implementación del mercado de datos de la Sección de Colaboración de la AP”:.....	49
3.1 Implementación del subsistema de almacenamiento.....	49
3.2 Implementación de los subsistemas de integración y visualización de datos.....	50
3.2.1 Implementación de las transformaciones .....	50
3.2.2 Implementación de los trabajos (Jobs).....	55
3.2.3 Implementación del subsistema de visualización de datos. ....	56

# *Índice de Contenido*

---

3.3 Vistas Materializadas .....	59
3.4 Seguridad de los datos. ....	59
3.4.1 Política de Respaldo y Seguridad. ....	59
3.5 Conclusiones del capítulo .....	60
Capítulo 4: “Validación de las funcionalidades mercado de datos para la Sección de Colaboración” .....	61
4.1 Validación y Prueba .....	61
4.1.1 Diseño de los Casos de Prueba .....	61
4.1.2 Validación de la investigación.....	63
4.1.3 Estrategias de pruebas para el Plan de Pruebas. ....	63
4.2 Resultados Obtenidos.....	63
4.3 Funcionalidades Obtenidas.....	64
4.4 Aporte Social y Económico .....	64
4.5 Conclusiones del capítulo .....	64
Conclusiones Generales. ....	65
Recomendaciones.....	66
Referencias Bibliográficas. ....	67
Bibliografía consultada. ....	69

## **Introducción**

En el día a día de las operaciones de cualquier organización, se generan datos como producto secundario de todas las transacciones que se llevan a cabo. Típicamente, ésta información se almacena a través de sistemas transaccionales en bases de datos relacionales que cumplen la función de almacenaje y administración de éstos datos. Pero la idea es que estos dejen de ser simples datos para convertirse en información que enriquezca las decisiones de los ejecutivos. Ante esto han surgido los almacenes de datos que permiten almacenar gran cantidad de datos y proveen un ambiente para que las organizaciones hagan un mejor uso de la información.

Los mercados de datos son una versión especial de los almacenes de datos, que contienen una visión de datos operacionales que ayudan a decidir sobre estrategias de negocio basadas en el análisis de tendencias y experiencias pasadas. La diferencia principal es que la creación de un mercado de datos es específica para una necesidad de datos seleccionados, enfatizando el fácil acceso a una información relevante.

Cuba desde la primera mitad del siglo XX se mantiene al tanto de las últimas tecnologías desarrolladas en el mundo y de todo aquello que se relacione con la informática ya que la misma es muy importante en nuestra sociedad, cubriendo ámbitos tan sencillos en nuestra vida diaria, razón por la cual es de vital importancia adquirir nuevos conocimientos y estar a la par con el desarrollo de los mismos.

Las tecnologías y la informática son la base del desarrollo en las empresas que principalmente necesitan digitalizar o mantener almacenada cierta cantidad de información relevante para las mismas. Teniendo como objetivo que los dirigentes puedan analizar la información y así tomar decisiones en un futuro, basándose en esos datos previamente almacenados.

La nueva distribución político-administrativa de la provincia de Artemisa ha puesto en práctica una nueva estructura de la AP, la cual brinda un mayor protagonismo y responsabilidad en la gestión de los recursos y los procesos fundamentales del territorio. La informatización territorial ocupa un elemento clave para el logro de un desarrollo local sostenible.

En este sentido, la utilización de las tecnologías de la informática en la gestión de la información relevante para la toma de decisiones por parte del gobierno provincial, constituye uno de los aspectos primordiales para el logro del desarrollo local.

En este aspecto la AP de Artemisa se ha puesto en marcha hacia la digitalización y almacenamiento de la información de la institución por el tiempo que sea necesario su resguardo, debido a la importancia que esto representa para la toma de decisiones en el futuro de la provincia. Para la realización de este trabajo la AP ha manifestado gran interés en el tema de los almacenes de datos.

La Facultad Regional de la UCI “Mártires de Artemisa”, como centro universitario enclavado en la nueva provincia de Artemisa, ha trazado una estrategia de informatización a la AP mediante el desarrollo de un almacén de datos.

Centro de excelencia en la formación desde la producción de estudiantes de pre grado de la carrera de Ingeniería en Ciencias Informáticas y en el desarrollo de productos, servicios y soluciones integrales. El mismo desarrolla proyectos de investigación y producción con entidades y organismos territoriales, nacionales e internacionales para crear nuevos productos y servicios.

La AP tiene como función principal la de preparar y proponer la política integral del estado en la provincia. Además se encarga de coordinar y controlar su cumplimiento sobre la base de las estrategias de desarrollo del

país. La misma cuenta con 32 direcciones y varias secciones independientes, entre ellas se encuentra la Sección de Colaboración, la cual se encarga de gestionar y asegurar la satisfacción de la Colaboración y de las relaciones internacionales. Para cumplir estas misiones, la Sección de Colaboración realiza un conjunto de funciones entre ellas están:

1. Asegurar la colaboración que la provincia recibe del exterior a través de Proyectos de Colaboración, Donativos Puntuales materiales y en efectivo y su inclusión en el Plan de la Economía Nacional según la Legislación vigente.
2. Asesorar y controlar todo lo relacionado con el desarrollo de la Colaboración que brinda la provincia y la legislación vigente de conjunto con los implicados en esta actividad en el territorio.
3. Controlar el cumplimiento de la política aprobada para los estudiantes extranjeros que cursan estudios en la provincia.
4. Asegurar la coordinación de la atención a las delegaciones extranjeras, embajadores y otros miembros del cuerpo diplomático acreditado en Cuba que visiten la provincia.

Actualmente la Sección de Colaboración de la AP de Artemisa por estar justamente en sus inicios tiene las siguientes dificultades:

- La información generada se archiva acumulando una gran cantidad de documentos que se realizan mediante herramientas obsoletas como Word y Excel o a través de lápiz y papel. Los cuales no son los más idóneos para la gestión de grandes volúmenes de información.
- La información no está centralizada.
- Existencia de errores en la información.
- Existe pérdida y duplicado de la información.

Todas las dificultades antes mencionadas traen como consecuencias que la Sección de Colaboración se vea en la necesidad de enfrentar procesos largos y tediosos en la recopilación y análisis de la información. Impidiéndole al Presidente y a los directivos de la sección controlar el comportamiento de las métricas con exactitud y la toma de decisiones.

Por los motivos antes expuestos se formuló el siguiente **problema de investigación:**

¿Cómo contribuir en la rapidez y seguridad en el análisis de la información como soporte para la toma de decisiones de la Sección de Colaboración de la AP de Artemisa?

A raíz del problema de investigación se define como **Objeto de estudio:** Almacenes de Datos y como **Campo de acción:** Mercado de Datos para la Sección de Colaboración.

Para dar solución al problema de investigación se plantea como **Objetivo general:** Desarrollar un Mercado de Datos que contribuya en la rapidez y seguridad en el análisis de la información como soporte para la toma de decisiones de la Sección de Colaboración de la AP de Artemisa. Se proponen además como **Objetivos Específicos:**

- Fundamentar los criterios teóricos relacionados con los almacenes de datos.
- Caracterizar el proceso de análisis de la información en la Sección de Colaboración de la AP de Artemisa.
- Desarrollar el Mercado de Datos de la Sección de Colaboración de la AP.
- Validar las funcionalidades del Mercado de Datos de la Sección de Colaboración.

Se plantea como **Hipótesis** lo siguiente: Si se implementa el Mercado de datos para la Sección de Colaboración de la AP de Artemisa, se contribuirá en la rapidez y seguridad en el análisis de la información como soporte para la toma de decisiones.

**Variable Independiente:** Mercado de datos para la Sección de Colaboración de la AP.

**Variable Dependiente:** Análisis de la información y la toma de decisiones.

Para dar cumplimiento a los objetivos específicos se proponen las siguientes

**Tareas de investigación:**

- Fundamentación del marco teórico de los almacenes de datos.
- Caracterización del proceso de análisis de la información en la Sección de Colaboración de la AP de Artemisa.
- Desarrollo del mercado de datos para la Sección de Colaboración.
- Validación de las funcionalidades del mercado de datos para la Sección de Colaboración.

**Posibles resultados:** Mercado de Datos para la Sección de Colaboración de la AP de Artemisa.

**Aporte práctico:** Como aporte práctico la investigación realizada brinda la creación de un mercado de datos para la Sección de Colaboración que facilite el análisis de la información como soporte para la toma de decisiones de manera rápida y segura y un informe detallado con toda la base teórico-práctico sobre la cual se sustente la solución propuesta.

**Métodos de Investigación Científicas:**

**Métodos teóricos**

Para una correcta comprensión y realización del trabajo se hace uso de los siguientes métodos científicos:

## **Histórico-Lógico**

Define cómo ha evolucionado y se ha desarrollado el diseño del mercado de datos. Para ello, se caracterizan los temas relacionados con las mejores prácticas en el desarrollo de mercados de datos de las herramientas de código abierto, se establecen las fuentes de datos de la AP de Artemisa para realizar el proceso de análisis.

## **Analítico-Sintético**

Permite elegir la granularidad del proceso del negocio para definir el nivel de análisis de la información, y evaluar la eficiencia del proceso a través de la carta de aceptación del cliente, seleccionar, de las metodologías existentes para el desarrollo de mercados de datos, la más apropiada según los criterios de éxito, para garantizar la calidad del desarrollo de la solución.

## **Modelado**

Se define la arquitectura del sistema para establecer las bases del desarrollo del mercado de datos, se definen las dimensiones y los hechos del mismo, para realizar el diseño de la solución.

## **Métodos empíricos**

Estos métodos describen y explican las características fenomenológicas del objeto, representan un nivel de la investigación cuyo contenido procede de la experiencia y es sometido a cierta elaboración racional.

## **Observación**

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, con lo que eliminan las deformaciones de otros métodos indirectos. Este constituye una forma directa de adquirir información acerca del fenómeno estudiado.

### **Entrevista**

Este método no es más que una conversación previamente definida por el investigador y conciliada con el entrevistado. Con el objetivo de obtener la mayor cantidad de información posible. Además de que será necesario conocer si existen estrategias y procesos definidos previamente, o mecanismos para gestionar los recursos humanos desde el punto de vista de la formación, así como conocer el nivel de responsabilidad que tiene o debe tener un proyecto respecto a dichos temas. Para la realización de este método se confeccionó el artefacto Cuestionario (FRA-Government Cuestionario-Colaboración) v1.0 y un fragmento del mismo se encuentra en los anexos. **Ver Anexo 1.**

### **Estructura del Documento:**

#### **Capítulo 1: “Fundamentación teórica de los sistemas dedicados a la toma de decisiones en el mundo”**

En este capítulo se realiza un estudio del estado de los sistemas dedicados a la toma de decisiones a nivel nacional e internacional, se evidencian los principales conceptos de los temas tratados en el desarrollo del almacén de datos, se presenta la metodología seleccionada y las herramientas de integración utilizadas para el tratamiento necesario de la información que permitirá a la Sección de Colaboración de la AP apoyar el proceso de toma de decisiones.

#### **Capítulo 2: “Análisis y diseño del mercado de datos de la Sección de Colaboración de la AP”:**

Se definen los requerimientos del sistema, se definen áreas de análisis, se especifica la arquitectura, se diseña el diagrama de casos de uso del sistema, el modelo de datos, los subsistemas de integración y visualización.

### **Capítulo 3 “Implementación del mercado de datos de la Sección de Colaboración de la AP”:**

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

### **Capítulo 4: “Validación de las funcionalidades del mercado de datos para la Sección de Colaboración”:**

Finalmente en este capítulo se aplicarán los casos de prueba para validar la calidad de los datos cargados, viendo que esto cumpliera con las necesidades de los usuarios de la Sección de Colaboración de la AP de Artemisa.

# *Capítulo 1: Fundamentación teórica de los sistemas dedicados a la toma de decisiones en el mundo.*

---

## **CAPÍTULO 1: “Fundamentación teórica de los sistemas dedicados a la toma de decisiones en el mundo.”**

Este capítulo contiene el estudio del estado de los sistemas dedicados a la toma de decisiones a nivel nacional e internacional donde se realizó una búsqueda de distintos almacenes y mercados de datos que trataran el tema de la colaboración. Además se evidencian los principales conceptos que son de gran utilidad para el desarrollo del mercado de datos. También se expone la metodología empleada y las herramientas de integración utilizadas para el tratamiento necesario de la información que permitirá a la Sección de Colaboración de la AP apoyar el proceso de toma de decisiones.

### **1.1 Estado del Arte**

Desde un inicio, las bases de datos se convirtieron en una herramienta fundamental de control y manejo de las operaciones comerciales. Fue así como en unos pocos años en grandes empresas y negocios existía un considerable número de información almacenada en diferentes fuentes de datos y estas ya habían alcanzado un tamaño considerablemente grande.

Debido al gran cúmulo de información, los directivos de tales empresas y negocios se dieron cuenta que esta podría tener un fin útil, al estar reflejada la mayoría de sus operaciones comerciales durante los llamados ciclos de negocios propios del mercado.

A su vez, los mercados empresariales han experimentado una transformación radical. Las empresas demandan mayor rapidez y eficiencia en la entrega de productos, y mejora en todos los servicios existentes, por lo que se hace imprescindible encontrar formas más eficaces de distribuir los productos, más facilidades para hacer estudios de mercado basados en la información de las

operaciones comerciales de las empresas y de sus clientes y mayor rapidez a la hora de tomar decisiones.

Por tanto, pensaron en unificar las diferentes fuentes de información de las cuales disponían, en un único lugar, al que sólo se le incorporaría información relevante, sobre la base de una estructura organizada, integrada, lógica, dinámica y de fácil explotación. La respuesta a esto fueron los Almacenes de Datos. (Torres, 2006)

Actualmente muchas instituciones en el mundo utilizan almacenes de datos debido a las ventajas que estos brindan entre ellos el Transporte de Cargas y Pasajeros, Telecomunicaciones, Comercio Minorista y Manufacturera de Bienes de Consumo Masivo.

En la búsqueda de una respuesta a las necesidades de la Sección de Colaboración, se estudiaron diferentes mercados de datos similares al que se diseñará, encontrados en Internet, entidades cubanas y en la Universidad. Este estudio se realizó con el objetivo de tomar experiencias en cuanto a su funcionamiento y a su vez analizar en qué medida dichos mercados brindaban solución a la problemática planteada.

### **1.1.1 Internacional**

En una búsqueda realizada sobre almacenes y mercados de datos a nivel mundial se encontró el almacén de datos del INEGI (Instituto Nacional de Estadísticas Geografía e Informática) que es el que más se asemeja a la Sección de Colaboración en cuanto a la colaboración con otros países.

El INEGI es el organismo responsable de divulgar la balanza comercial de manera oficial en México. Esta organización se encarga de coordinar las actividades del grupo de trabajo, así como de propiciar la aplicación de metodologías adecuadas, con base en las recomendaciones internacionales que permitan generar resultados estadísticos de calidad. (Salazar Luján,

2008)

Uno de los medios de difusión del INEGI es un almacén de datos en Internet que permite realizar consultas a la información del Comercio Exterior de México en forma rápida y eficiente, ofreciendo la posibilidad de explotar los datos requeridos.

Haciendo un análisis de este almacén se pudo evidenciar que no era factible para la utilización del mismo en la AP debido a que este almacén es la fuente de datos única para la presentación de resultados que ahora no se limita a la generación de publicaciones sino que además será la fuente de información para sistemas de consulta interna, desarrollo de productos digitales comercializables y para el sitio del INEGI en Internet por lo tanto sería un trabajo muy engorroso adaptarlo para la Sección de Colaboración de la AP de Artemisa.

## **1.1.2 Nacional**

En Cuba siempre se la ha concedido mucha importancia al hecho de estar actualizados con las últimas tecnologías y el dominio de las herramientas que se van desarrollando en el ámbito informático. A pesar de ser un país subdesarrollado existen empresas, instituciones u organizaciones que han tenido la necesidad de digitalizar toda la información que poseen y mantenerla almacenada por un tiempo determinado para ser más efectiva la toma de decisiones de los directivos.

Un ejemplo es la Corporación CIMEX, la cual implementó un almacén que se especializa en la actividad del comercio con el objetivo de apoyar a la toma de decisiones de la misma. También han trabajado en la implementación de Almacenes de Datos, el Banco de Inversiones S.A para ayudar a la toma de decisiones en estas entidades.

En la Universidad de las Ciencias Informáticas (UCI) existen varios centros de

desarrollo especializados en los almacenes de datos, uno de ellos es el Centro de Tecnologías de Gestión de Datos (DATEC), el cual ha implementado almacenes de datos para distintas empresas e instituciones como la Oficina Nacional de Estadísticas e Información (ONEI).

La Facultad Regional Mártires de Artemisa ha incursionado en el tema de los almacenes de datos para importantes instituciones del país, ejemplo de ello es el almacén desarrollado para el Ministerio del Comercio Exterior y la Inversión Extranjera (MINCEX).

Al realizar un análisis de dicho almacén, se pudo evidenciar que presentaba características similares al que se va a desarrollar para la Sección de Colaboración de la AP de Artemisa, pero es una versión inestable y no presenta los datos de algunos indicadores que se desean analizar además se necesita de un mercado de datos que reúna las temáticas de Proyectos de colaboración y donativos puntuales, proyectos nacionales, empresas mixtas, estudiantes extranjeros en la provincia, colaboradores cumpliendo misión, reclamaciones y atención a familiares de los colaboradores.

Por tanto no se puede hacer uso del almacén desarrollado para el MINCEX en la Sección de Colaboración y se decide implementar una solución que cubra las necesidades específicas de esta Sección.

Los almacenes y mercados de datos anteriormente descritos son insuficientes para suplir las exigencias del producto requerido para la Sección de Colaboración. En ellos, se encuentran solamente fragmentos de los requisitos del mercado de datos que se pretende implementar.

Las características de estos mercados, aunque sean parecidos, no corresponden en totalidad con las deseadas para el mercado de datos que se va a desarrollar para la Sección de Colaboración.

Concluyendo que ninguno de los mercados estudiados contiene las

características principales para ser usado por la Sección de Colaboración. Quedando así establecido iniciar la implementación de un nuevo mercado de datos que le dé respuesta a las necesidades específicas de esta Sección.

## **1.2 Principales conceptos**

### **1.2.1 Almacenes de datos:**

Existen distintas definiciones sobre los almacenes de datos interpretados particularmente por cada uno de los especialistas y profesionales del tema, aunque en ellos no estén de acuerdo en algunos aspectos todos concuerdan en que un almacén de datos es el centro de la arquitectura de los sistemas de información, por ejemplo:

Bill Inmon fue uno de los primeros autores en escribir sobre el tema de los almacenes de datos, define un almacén de datos en términos de las características del repositorio de datos:

#### **Orientado a temas**

Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.

#### **Variante en el tiempo**

Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.

#### **No volátil**

La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas.

#### **Integrado**

La base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.

Ralph Kimball otro gran especialista sobre estas colecciones de datos, reconocido mundialmente y escritor del famoso libro "The building Data Warehouse" afirma que es *"una copia de los datos de la transacción estructurados específicamente para la consulta y el análisis."*

Luego de la investigación realizada para el análisis de los almacenes de datos las autoras se identifican con la definición del prestigioso ingeniero e investigador Ralph Kimball quien define un Almacén de datos como "...una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis...", también determinó que un almacén de datos no era más que "...la unión de todos los mercados de datos de una entidad...".

Para comprender íntegramente el concepto de almacén de datos, es importante entender cuál es el proceso de construcción del mismo, denominado ETL (Extracción, Transformación y Carga) que es el proceso que organiza el flujo de los datos entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un almacén de datos, reformatearlos, limpiarlos y cargarlos en otra base de datos, mercado de datos o bodega de datos.

- ✓ **Extracción:** obtención de información de las distintas fuentes tanto internas como externas.
- ✓ **Transformación:** filtrado, limpieza, depuración, homogeneización y agrupación de la información.
- ✓ **Carga:** organización y actualización de los datos y los metadatos en la base de datos.

## **1.2.2 Mercado de Datos:**

*Un mercado de datos es una base de datos departamental o un subconjunto de la información de una organización que es de valor para un grupo de usuarios en específico. Se caracteriza por guardar los datos en una estructura consistente que permite el análisis de la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Puede ser alimentado desde un almacén de datos y así garantizar la fiabilidad de la información, o integrar por sí mismo un compendio de distintos sistemas operacionales, son fáciles de entender y navegar. Al ser una parte del almacén de datos reduce el tiempo de respuesta a la hora de consultar su contenido debido a que es reducido el volumen de datos para recorrer. (Ralph Kimball 2002).*

## **1.2.3 Dimensiones:**

*Son las variables sobre las que se califican los datos, pueden ser descriptivas o de hechos. Las descriptivas proporcionan una explicación de los datos que serán analizados. En una librería, las dimensiones descriptivas detallarían un producto, como por ejemplo cuadernos, clasificándolos en tamaño, cantidad de hojas, colores. La dimensión de hechos, por su parte, contendría los valores cuantitativos o de medida, como las cantidades, las ventas totales y los precios. (Chinchilla Arley, Ricardo, 2011)*

## **1.2.4 Jerarquías:**

*Thomsen (1997) las describe como: una estructura de árbol, con una raíz en la parte superior y las ramas en la parte inferior. Si todas las ramas se encuentran en un mismo nivel, se denominan jerarquía simétrica, pero si esto no ocurre, se denomina jerarquía asimétrica.*

## **1.2.5 Tablas de hecho:**

Una tabla de hechos captura los datos que miden las operaciones. Las tablas de hechos normalmente se componen de un gran número de filas, sobre todo cuando contienen uno o varios años de historial de un proyecto de equipo grande.

Una característica esencial de las tablas de hechos es que contienen datos numéricos (hechos) que se pueden resumir para proporcionar información sobre el historial de las operaciones de la organización. Cada tabla de hechos también incluye un índice de varias partes que contiene, como claves externas, las claves primarias de las tablas de dimensiones relacionadas así como los atributos de los registros de hechos.

Las tablas de hechos no deben contener información descriptiva ni datos que no procedan de los campos de medición numéricos y los campos de índice que relacionan los hechos con las correspondientes entradas en las tablas de dimensiones. Además es importante saber que las tablas de hecho es el “nivel de detalle” de la información que se almacena. (2011)

## **1.2.6 Medida:**

*Una medida es un tipo de dato cuya información es usada por los analistas (usuarios) en sus consultas para medir el rendimiento del comportamiento de un proceso o un objeto del negocio. Las medidas candidatas son los datos numéricos, pero no cada atributo numérico es una medida candidata. (Mgter. David Luis la Red Martinez, 2007)*

## **1.2.7 Cubo:**

*Principal objeto del proceso analítico en línea (OLAP, Online Analytic Processing), una tecnología que proporciona rápido acceso a los datos de un almacén de datos, conjunto de datos que normalmente se construyen a partir de un subconjunto de un almacén de datos y se organiza y resume en una estructura multidimensional definida por un conjunto de dimensiones y medidas. (Jonathan David Ramos, 2010)*

## 1.3 Modos de Almacenamiento de Datos

Existen tres modelos para el proceso analítico en línea (OLAP, por sus siglas en inglés) de la información ROLAP, MOLAP y HOLAP. El proceso de análisis se realiza de igual forma lo que varía en uno y otro caso es la metodología de almacenamiento. La forma de almacenamiento es crítica para garantizar la velocidad de recuperación de la información, las zonas de ubicación de las agregaciones y el procesamiento de los datos en general.

OLAP es un proceso analítico de datos en línea, que permite al usuario seleccionar y extraer la información desde diferentes puntos de vista, se traduce en un conjunto de técnicas y estándares. Tiene como objetivo agilizar el proceso de consulta de grandes volúmenes de datos. Dependiendo de las técnicas que se utilicen a la hora de obtener los datos, se pueden definir varias herramientas de OLAP, como son:

- **ROLAP** “Procesamiento analítico relacional en línea”: Crean vistas multidimensionales extrayendo los datos de bases de datos SQL relacionales y simulan los datos multidimensionales. (Ronda 2007).
- **MOLAP** “Procesamiento analítico multidimensional en línea”: Son herramientas que acceden a datos que no están almacenados en registros de tablas, sino que almacenan los datos en arreglos de varias dimensiones, llamados cubos. Estos cubos utilizan índices para optimizar el acceso a los datos. (Ronda 2007).
- **HOLAP** “Procesamiento analítico híbrido en línea”. Permiten un análisis híbrido de la información, une lo mejor de las dos herramientas anteriores. El análisis HOLAP ayuda a reducir costes de hardware ya que se necesita menos espacio en disco que en las bases de datos relacionales. Además, la

respuesta de las consultas sobre las bases de datos multidimensionales son más rápidas que sobre las relacionales. Como aspecto negativo, los datos multidimensionales deben ser cargados antes de ser consultados y actualizados cuando se actualizan los datos de la organización. (Ronda 2007).

La herramienta que se utilizó para la obtención de los datos fue ROLAP ya que está diseñada para realizar análisis de datos a través del uso de modelos de datos sobre un sistema relacional clásico.

Las herramientas ROLAP acceden a los datos de una base de datos relacional y generan consultas SQL para calcular la información al nivel apropiado cuando un usuario final lo requiere.

## 1.4 Metodología

Existen muchas metodologías de diseño y construcción de un almacén de datos. Cada fabricante de software de inteligencia de negocios busca imponer una metodología con sus productos. Sin embargo, se imponen entre la mayoría dos metodologías, la de Kimball y la de Inmon.

La metodología de Kimball proporciona una base empírica y metodológica adecuada para las implementaciones de almacenes de datos pequeños y medianos, dada su gran versatilidad y su enfoque ascendente, que permite construir los almacenes en forma escalonada. Además presenta una serie de herramientas, tales como planillas, gráficos y documentos, que proporcionan una gran ayuda.

Por otra parte, la metodología de Inmon se basa en conceptos bien conocidos del diseño de bases de datos relacionales (Inmon 02, Imhoff&Galemmo 03); la metodología para la construcción de un sistema de este tipo es la habitual para

construir un sistema de información, utilizando las herramientas habituales, al contrario de la de Kimball, que se basa en un modelado dimensional (no normalizado) (Kimball et al 98, 08).

Desde el punto de vista arquitectónico, la mayor diferencia entre los dos autores es el sentido de la construcción del Almacén de datos, esto es comenzando por los mercados de datos o ascendente (Bottom-up, Kimball) o comenzando con todo el Almacén de datos desde el principio, o descendente (Top-Down, Inmon).

Tomando como base estas metodologías se han desarrollado otras que no siguen obligatoriamente una específica sino que realizan una selección de lo mejor de cada una y definen su propia metodología.

Para el desarrollo del mercado de datos se tomó la propuesta desarrollada por el Centro de Tecnologías de Gestión de Datos (DATEC) de la UCI que permite mitigar las desventajas de la Metodología de Kimball y se ajusta a las condiciones y características de producción. Además integra las mejores prácticas identificadas en las metodologías estudiadas que permiten un adecuado desarrollo de soluciones de Almacenes de Datos.

## **1.5 Modelo Dimensional**

Es considerado una de las técnicas favoritas en la construcción de un almacén de datos, en la cual se constituyen modelos de tablas y relaciones con el propósito de optimizar la toma de decisiones, con base en las consultas hechas en una base de datos relacional que están ligadas con la medición o un conjunto de mediciones de los resultados de los procesos de negocio.

Es una técnica de diseño lógico que tiene como objetivo presentar los datos dentro de un marco de trabajo estándar e intuitivo, para permitir su acceso con un alto rendimiento. Cada modelo dimensional está compuesto por una tabla con una llave combinada, llamada tabla de hechos, y un conjunto de tablas más pequeñas

llamadas tablas de dimensiones. Los elementos de estas tablas se pueden definir de la siguiente manera:

- **Hechos:** Colección de piezas de datos y datos de contexto. Cada hecho representa una parte del negocio, una transacción o un evento.
- **Dimensiones:** Una dimensión es una colección de miembros, unidades o individuos del mismo tipo.
- **Medidas:** Son atributos numéricos de un hecho que representan el comportamiento del negocio relativo a una dimensión. (Kimball, 2009)

## 1.5.1 Esquemas que se encuentran dentro del modelo dimensional

### 1.5.1.1 Esquema Estrella

Una base de datos dimensional se puede concebir como un cubo de tres o cuatro dimensiones, en el que los usuarios pueden acceder a una porción de la base de datos a lo largo de cualquiera de sus dimensiones.

Dado que es muy común representar a un modelo dimensional como una tabla de hechos rodeada por las tablas de dimensiones, frecuentemente se le denomina también como modelo estrella o esquema de estrella-unión. Este modelo también es asimétrico pues hay una tabla dominante en el centro con varias conexiones a otras tablas. **Ver anexo 2**

### 1.5.1.2 Esquema copo de nieve (snowflake)

En este esquema se presentan ramificaciones a partir de las tablas de dimensiones y no sólo a partir de la tabla de hechos. Aunque puede reducir espacio por la mínima redundancia de datos, tiene la contrapartida de peores rendimientos al tener que crear más tablas de dimensiones y más relaciones entre las tablas lo que tiene un impacto directo sobre el rendimiento. Este

esquema es más complejo que el esquema de estrella. **Ver anexo 3**

### 1.5.1.3 Esquema constelación

Su diseño y cualidades son muy similares a las del esquema en estrella pero posee una serie de diferencias con el mismo, que son precisamente las que lo destacan y caracterizan. Entre ellas se pueden mencionar:

- Permite tener más de una tabla de hechos, por lo cual se podrán analizar más aspectos claves del negocio con un mínimo esfuerzo adicional de diseño.
- Contribuye a la reutilización de dimensiones, ya que una misma dimensión puede utilizarse para varias tablas de hechos. **Ver Anexo 4** (Ing. Bernabeu, Ricardo Dario, 2007)

### 1.5.2 Selección del esquema

El esquema utilizado en el diseño del modelo dimensional fue el esquema constelación ya que algunas tablas de hechos comparten las mismas dimensiones.

## 1.6 Herramientas utilizadas para el desarrollo del Mercado de datos

Para la solución de la investigación se decidió utilizar herramientas de código libre, siguiendo la política de software libre del país, de la universidad y del proyecto para el desarrollo de Almacenes de Datos las cuales se muestran a continuación:

### Herramientas de modelado:

#### 1.6.1 Visual Paradigm for UML v6.4

Es una herramienta que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y

despliegue. El software de modelado UML ayuda a una más rápida construcción de aplicaciones de calidad, mejores y a un menor coste. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. Presenta licencia gratuita y comercial. Es fácil de instalar y actualizar, además de ser compatible entre ediciones. (2007)

## 1.6.2 SQL PowerArchitect v0.9.13

Es una herramienta de modelado de datos que fue creada por los diseñadores de almacenamiento de datos.

Con esta herramienta, se podrá abrir múltiples conexiones concurrentes a bases de datos, crear y explorar perfiles de datos fuente, arrastrar y soltar esquemas de datos, tablas y columnas dentro del modelo de datos, y confeccionar la base de datos resultante con su plantilla ETL asociada. Hasta el diseño más complicado, la base de datos más grande o el modelo más inabarcable, podremos gestionarlo y manejarlo con PowerArchitect.

### Algunas Características de SQL PowerArchitect.

- Permite acceder a las bases de datos a través de JDBC<sup>1</sup>.
- Puedes conectarte a múltiples bases de datos al mismo tiempo.
- Todos los proyectos se guardan en formato XML<sup>2</sup>.
- Compara modelos de datos y estructuras de bases de datos e identifica las discrepancias. Drag-and-drop de las tablas origen y las columnas en el área de trabajo.

---

<sup>1</sup> Por sus siglas en inglés (Java Database Connectivity) es una API que permite la ejecución de operaciones sobre bases de datos desde el lenguaje de programación Java, independientemente del sistema operativo donde se ejecute o de la base de datos a la cual se accede, utilizando el dialecto SQL del modelo de base de datos que se utilice.

<sup>2</sup> Por sus siglas en inglés (Extensible Markup Language) es un lenguaje de marcas desarrollado por el World Wide Web Consortium (W3C).

- SQL PowerArchitect edición Community es gratuita bajo licencia Open Source GPL v.3.

SQL PowerArchitect es una herramienta ideal para grupos de desarrollo donde se puede realizar el modelado de datos y poder así tener documentado el modelo de datos de todas las aplicaciones que se desarrollan. (2010)

## **Gestor de Base de Datos**

### **1.6.3 PostgreSQLv9.1**

Esta versión contiene una gran cantidad de mejoras para hacer la administración, consulta y programación en PostgreSQL mucho más fácil que nunca. Con las funcionalidades nuevas o mejoradas en la versión 9.1, hay aún más motivos para escoger PostgreSQL para futuros proyectos.

La mayoría de los cambios en PostgreSQL 9.1 son herramientas y órdenes de administración y monitoreo, nuevas o mejoradas. Cada usuario tiene su funcionalidad favorita que hace su trabajo cotidiano con PostgreSQL más fácil y productivo.

Entre las mejoras más populares están:

- Restauración de bases de datos en procesos paralelos, que acelera recuperación de un respaldo hasta 8 veces.
- Privilegios por columna, que permiten un control más granular de datos confidenciales.
- Configuración de ordenamiento configurable por base de datos, lo cual hace a PostgreSQL más útil en entornos con múltiples idiomas.

- Nuevas herramientas de monitoreo de consultas que le otorgan a los administradores mayor información sobre la actividad del sistema. (Jaime Casanova, 2009)

## **Herramienta manejadora de Base de Datos**

### **1.6.4 PgAdmin v1.12.0**

PgAdmin III está diseñado para responder a las necesidades de todos los usuarios, desde escribir consultas SQL simples hasta desarrollar bases de datos complejas. La interfaz gráfica soporta todas las características de PostgreSQL y facilita enormemente la administración. La conexión al servidor puede hacerse mediante conexión TCP/IP o Unix Domain Sockets (en plataformas \*nix), y puede encriptarse mediante SSL para mayor seguridad.

Administración pgAdmin III es la herramienta Open Source de administración por excelencia para las bases de datos PostgreSQL. Algunas de sus características son: el soporte completo para UNICODE, edición rápida de consultas y datos multihilo y soporte para todos los tipos de objetos de PostgreSQL. (2008)

## **Herramienta para la integración de datos.**

### **1.6.5 Pentaho Data Integration v4.1**

Pentaho Data Integration proporciona la extracción de gran alcance, transformación y carga (ETL) utilizando una innovadora, basada en metadatos enfoque. Con una interfaz intuitiva y gráfica de arrastrar y soltar entorno de diseño y una arquitectura probada, escalable y basada en estándares, Pentaho Data Integration es cada vez más la elección de las organizaciones más tradicionales de ETL, de propiedad o las herramientas de integración de datos. (2011)

## **Herramientas para la Inteligencia de Negocios.**

### **1.6.6 SchemaWorkbench (SW) v3.2.1**

Es una herramienta que permite crear y probar esquemas de cubos OLAP visualmente, además permite publicarlos al servidor BI para que puedan ser utilizados en los análisis por los usuarios de la plataforma. Estos archivos son los modelos de esquemas XML de metadatos que se crean en una estructura específica. Estos modelos XML se pueden considerar las estructuras de forma de cubo que utilizan hechos existentes y tablas de dimensiones.

Con esta aplicación, se puede configurar una conexión JDBC como el modelo físico, para luego elaborar el esquema lógico de manera simple y efectiva. Para ello el entorno ofrece un editor de esquemas con la fuente de datos subyacente para su validación. Permite la ejecución de consultas MDX contra el esquema y la base de datos y la navegación por la base de datos subyacente. (Sherman Wood, JasperSoft, 2007)

## **1.6.7 BI-Server v3.6**

BI Server es una solución de inteligencia de negocios completa y fácil de usar, la cual provee la información requerida para tomar mejores decisiones de manera más rápida. La misma aumenta la consistencia de los datos, permitiendo tomar decisiones basadas en datos. Simplifica la administración y permite la consolidación y estandarización de proveedores con una solución integral de BI.

Proporciona una visualización que permite a los tomadores de decisiones de forma dinámica explorar ideas, investigar patrones, descubrir hechos ocultos y compartir los conocimientos en toda la empresa para una mejor toma de decisiones. Además contiene un extenso conjunto de opciones personalizables y de presentación gráfica de información actual y perspectivas que no son fáciles de detectar en formato tabular. (SAS® Enterprise, 2011)

## **1.6.8 Pentaho Report Desing (PRD) v3.7.0**

Es un editor basado en eclipse con prestaciones profesionales, de calidad y con

capacidad de personalización de informes a las necesidades de negocio destinado a desarrolladores.

Incluye asistentes para facilitar la configuración de propiedades. Está estructurado de forma que los desarrolladores pueden acceder a sus prestaciones de forma rápida:

Incluye un editor de consultas para facilitar la confección de los datos que serán utilizados en un informe. (Díaz Ordaz, 2011)

### **1.6.9 Pentaho Metadata Editor (PME) v3.7.0**

Es una herramienta que permite crear dominios de metadatos y modelos. El objetivo es mapear la estructura física de la base de datos a un modelo lógico de negocio.

## **1.7 Conclusiones del capítulo**

Todo lo expuesto en este capítulo constituye la base sobre la que se trabajará para desarrollar el mercado de datos como vía de solución al problema planteado. Puesto que se realizó una investigación de todas las tecnologías, herramientas, modos de almacenamiento de la información y metodologías que contribuirán al desarrollo del mercado de datos, permitiendo elaborar una propuesta para la selección de aquellas que resulten ser las más acertadas para utilizar en la aplicación. En este caso se utilizará el modo de almacenamiento ROLAP y se tomó la propuesta de metodología desarrollada por DATEC para el diseño e implementación de los almacenes de datos. Así como utilizar las principales herramientas pertenecientes a la familia de software libre como son:

- **CASE Visual Paradigm v6.4:** para realizar el diseño del mercado de datos.
- **PowerArchitect v0.9.13:** para el modelado de datos.

- **PostgreSQL v9.1:** como SGBD<sup>3</sup>.
- **PgAdmin v1.12.0:** como herramienta manejadora de base de datos.
- **Pentaho Data Integration v4.1:** para los procesos de ETL.
- **SchemaWorkbench (SW) v3.2.1:** para crear y probar esquemas de cubos OLAP visualmente.
- **BI-Server v3.6:** para la presentación de la información.
- **Pentaho Report Desing (PRD) v3.7.0:** para la personalización de informes.
- **Pentaho Metadata Editor (PME) 3.7.0:** para crear dominios de metadatos y modelos.

---

<sup>3</sup> Sistema Gestor de Bases de Datos es un tipo de software muy específico, dedicado a servir de interfaz entre la base de datos, el usuario y las aplicaciones que la utilizan.

## *Capítulo 2: Análisis y diseño del Mercado de Datos.*

---

### **CAPÍTULO 2: “Análisis y diseño del mercado de datos de la Sección de Colaboración de la AP”.**

El desarrollo de un mercado de datos no es una tarea fácil puesto que la misma está compuesta por un conjunto de elementos que interactúan entre sí con el objetivo de obtener un diseño robusto y adaptable a las necesidades reales de los usuarios finales.

En este capítulo se abordarán temas relacionados con los requerimientos del sistema, así como la especificación de la arquitectura, además se definirán las distintas áreas de análisis. Se diseñará el diagrama de casos de uso del sistema, el modelo de datos, los subsistemas de integración y visualización. En fin abordará el resultado del análisis y el diseño del mercado de datos de la Sección de Colaboración.

#### **2.1 Análisis**

Este proceso es de vital importancia para la realización de un producto con calidad, para ello es necesario que se definan los requerimientos indispensables para lograr satisfacer correctamente las peticiones del cliente. En el proceso de análisis se determinan los requisitos funcionales, los no funcionales y también los informativos. Por otra parte, se definen los casos de uso del sistema que permiten entender cómo se va a desarrollar el mercado de datos.

##### **2.1.1 Especificación de Requerimientos**

###### **2.1.1.1 Requisitos Informativos**

Las funcionalidades de los requisitos informativos están enfocadas a las necesidades de información que presentan los usuarios. Los requisitos informativos deben mostrar la información que se solicita de la manera en

que el cliente lo requiera.

Para el levantamiento de los requisitos se realizaron entrevistas con el cliente, a través de los representantes de la Sección de Colaboración de la AP de Artemisa, en la misma se identificaron 160 Requisitos Informativos agrupados en 5 Casos de Uso que se encuentran descritos de forma detallada en el artefacto de Especificación de Requisitos (FRA-Government-Especificación de Requisitos-Colaboración) v1.0, dentro de los cuales se encuentran los siguientes:

### **Control de Proyectos de Colaboración y Donativos Puntuales**

- Obtener el financiamiento recibido, tasa de cambio, recibido en cuc por proyecto.
- Obtener el financiamiento recibido, tasa de cambio, recibido en cuc por donante.
- Obtener el financiamiento recibido, tasa de cambio, recibido en cuc por año, entidad y municipio.
- Obtener el financiamiento recibido, tasa de cambio, recibido en cuc por año, país y entidad.
- Obtener el financiamiento recibido, tasa de cambio, recibido en cuc por año, clasificación y país.
- Obtener presupuesto inicial, presupuesto del año, financiamiento recibido CUC por donativos colaboración.
- Obtener presupuesto inicial, presupuesto del año, financiamiento recibido CUC por país.
- Obtener presupuesto inicial, presupuesto del año, financiamiento recibido CUC por contraparte cubana.
- Obtener presupuesto inicial, presupuesto del año, financiamiento recibido CUC por contraparte cubana, contraparte extranjera, donante y país.

- Obtener presupuesto inicial, presupuesto del año, financiamiento recibido CUC por año y mes.

### **Proyectos Nacionales**

- Obtener presupuesto aprobado, saldo inicial, ejecución año, saldo actual por organismo.
- Obtener presupuesto aprobado, saldo inicial, ejecución año, saldo actual por año.
- Obtener presupuesto aprobado, saldo inicial, ejecución año, saldo actual por proyecto nacional.
- Obtener presupuesto aprobado, saldo inicial, ejecución año, saldo actual por municipio.
- Obtener presupuesto aprobado, saldo inicial, ejecución año, saldo actual por año, mes y organismo.

### **Control de colaboradores cumpliendo misión.**

- Obtener cantidad de misiones por año.
- Obtener cantidad de misiones por colaborador.
- Obtener cantidad de misiones por país.
- Obtener cantidad de misiones por municipio.
- Obtener cantidad de misiones por año, municipio y entidad.

### **Atención a familiares.**

- Obtener la cantidad de visitas efectuadas por año.
- Obtener la cantidad de visitas efectuadas por entidad.

- Obtener la cantidad de visitas programadas, cantidad de visitas efectuadas por año.
- Obtener la cantidad de visitas programadas, cantidad de visitas efectuadas por municipio.
- Obtener la cantidad de visitas programadas, cantidad de visitas efectuadas por año y mes.

### **Reclamaciones de colaboradores.**

- Obtener la cantidad de reclamaciones totales, la cantidad de reclamaciones en trámite por año.
- Obtener la cantidad de reclamaciones totales, la cantidad de reclamaciones en trámite por entidad.
- Obtener la cantidad de reclamaciones totales, la cantidad de reclamaciones en trámite por tipo de reclamación.
- Obtener la a cantidad de reclamaciones en trámite y cantidad de reclamaciones resueltas por año y tipo de reclamación.
- Obtener la cantidad de reclamaciones totales, la cantidad de reclamaciones en trámite y cantidad de reclamaciones resueltas por año y tipo de reclamación.

### **Control de estudiantes extranjeros.**

- Obtener cantidad de becarios por año, país y centro.
- Obtener cantidad de becarios por mes y año actual.
- Obtener cantidad de becarios por año, mes y país.
- Obtener cantidad de becarios por centro.
- Obtener cantidad de becarios por país del último año.

### **Control de empresas mixtas.**

- Obtener la cantidad de empresas mixtas por entidad.
- Obtener la cantidad de empresas mixtas por organismo.
- Obtener la cantidad de empresas mixtas por municipio.
- Obtener la cantidad de empresas mixtas por año.
- Obtener la cantidad de empresas mixtas por país, entidad y municipio.

### 2.1.1.2 Requisitos Funcionales

Los requisitos funcionales describen las funcionalidades que el sistema debe hacer, ellos definen el comportamiento interno del software y lo establecen según se requiera. Para la investigación de este trabajo se definieron 12 requisitos funcionales agrupados en 10 casos de uso que validan fundamentalmente la autenticación de usuarios y los procesos de extracción, transformación y carga de los datos. Ellos son:

**RF1. Autenticar usuario:** Permite el acceso a la aplicación y define las actividades según el rol. Tiene complejidad Alta.

**RF2. Adicionar usuario:** Permite adicionar usuarios al sistema.

**RF3. Eliminar usuario:** Permite eliminar un usuario existente del sistema.

**RF4. Realizar la transformación y carga de los datos:** Permite realizar la transformación a los datos extraídos, para luego ser cargados en el mercado de datos.

**RF5. Realizar la extracción de los datos:** Permite realizar la extracción de los datos de los archivos fuentes.

**RF6. Graficar la información:** Realizar la obtención de gráficos para que la información obtenida sea más entendible. Tiene complejidad Alta.

**RF7. Disponer de las vistas materializadas referentes a los principales cortes que se realizan a la información:** Realizar la obtención de la información por distintos parámetros. Tiene complejidad Alta.

**RF8. Adicionar los reportes:** Permite adicionar nuevos reportes al sistema.

**RF9. Eliminar reporte:** Permite eliminar un reporte existente en el sistema.

**RF10. Adicionar rol:** Permite adicionar nuevos roles al sistema.

**RF11. Eliminar rol:** Permite eliminar un rol existente en el sistema.

**RF12. Configurar vistas de análisis:** Permite configurar vistas de análisis.

### 2.1.1.3 Requisitos No Funcionales

Por las peculiaridades de los sistemas de almacenamiento y análisis de datos el usuario final debe tener conocimiento en el trabajo con las herramientas de integración de datos. Los mismos especifican criterios que pueden usarse para juzgar la operación de un sistema. En fin los requisitos no funcionales son características o capacidades que el sistema debe tener o cumplir. En la realización de este trabajo se identificaron los siguientes requisitos no funcionales:

- Requisitos de Usabilidad
- Requisitos de Fiabilidad
- Requisitos de Eficiencia
- Restricciones de Diseño
- Requisitos para la documentación de usuarios en línea y ayuda del sistema.

Interfaz

Interfaces de usuario  
Interfaces Hardware  
Interfaces Software  
Interfaces de Comunicación

- Requisitos de Licencia

### 2.1.2 Arquitectura

La arquitectura de un mercado de datos es una arquitectura multidimensional, fue desarrollada por el Dr. Ralph Kimball; está basada en la premisa que todos los análisis de Inteligencia del negocio tiene su origen en una estructura multidimensional. Esta tiene las siguientes características:

Es ascendente (bottom-up).

Está basado principalmente en uso de mercados de datos independientes.

La arquitectura de un mercado de datos está separada en dos capas de procesos y servicios:

- **Habitación Trasera:** A groso modo la cocina del sistema. Se realizan todos los procesos ETL para conseguir los datos de las fuentes de origen, pero además también se consideran aquellos procesos ETL que alimenta cada uno de los mercados de datos independientes existentes en la organización. Kimball distingue además dos tipos de mercados de datos:

**Mercados de datos Atómicos:** Contienen la información al nivel de detalle máximo.

**Mercados de datos Agregados:** Contienen la información por departamento, por áreas o funcional que puede estar alimentado por los anteriores o directamente del Área de Estacionamiento.

- **Habitación Delantera:** Consistente por las herramientas de análisis que usan la información consolidada en los mercados de datos de la habitación trasera. Es por lo tanto claro, que para cada unidad de negocio se creará un mercado de datos (o varios) sin tener en cuenta las necesidades de otra unidad. No prima la visión única del dato.

### 2.1.2.1 Visión general de la arquitectura.

A continuación se explica la arquitectura de la solución propuesta, detallando cada uno de los componentes en los cuales está basada la misma, para tener una visión general del sistema. Para ello se identificaron 3 subsistemas en los cuales el sistema está estructurado:

- Subsistemas Integración.
- Subsistemas Almacenamiento.
- Subsistemas Visualización.

El subsistema de integración es donde se agrupan los procesos que tienen como objetivo realizar tareas relacionadas con la extracción, la transformación y la carga de la información con la que se trabajará. Desde que se cargan los datos de las fuentes hasta que es poblado el mercado de datos se generan un conjunto de transformaciones a implementar que se realizan mediante el proceso de ETL el cual es el encargado de poblar el mercado de datos, dándole así seguimiento al subsistema de almacenamiento que no precisamente debe tener un orden sino que se pueden llevar a cabo paralelamente.

El subsistema de almacenamiento contiene los datos que antes estaban en distintas fuentes para brindar información al usuario mediante el subsistema de visualización, para lograr este objetivo se realizaron 2 tipos de cargas: cargas históricas a partir de la información en ficheros Excel con la cual ya contaba la Sección de Colaboración y cargas incrementales teniendo como fuente de datos

una bases de datos relacional (PostgreSQL).

El subsistema visualización está compuesto por 2 componentes principales los cuales se mencionan a continuación:

- Cubos OLAP.
- Capa de Presentación.

El subsistema de visualización tiene como función principal mostrar los datos almacenados a través de las distintas herramientas, además es una interfaz con la cual interactuará el usuario final la cual se comunica directamente con el servidor de cubos a través de consultas, las cuales retornan la información requerida donde ésta es transformada y presentada para la visualización de los reportes y vistas de análisis.

### 2.1.3 Definición de las Áreas de Análisis

La definición de las Áreas de Análisis (AA) es uno de los pasos más importantes dentro del desarrollo de un mercado de datos. La realización del mismo enfoca el desarrollo hacia el buen cumplimiento de las metas trazadas y garantiza la factibilidad, utilidad y el éxito de las estructuras que se están diseñando. En la solución propuesta se orientan en función de los diferentes cortes a la información que comúnmente la Sección de Colaboración realiza. En este sentido se definieron 4 AA las cuales se muestran a continuación:

- Análisis de los indicadores seleccionados de los donativos y proyectos de colaboración.
- Análisis de los indicadores seleccionados de las empresas mixtas.
- Análisis de los indicadores seleccionados de los colaboradores cumpliendo

misión.

- Análisis de los indicadores seleccionados de los estudiantes extranjeros en la provincia.

### 2.1.4 Diagrama de Casos de Uso del Sistema.

Los diagramas de casos de uso documentan el comportamiento de un sistema desde el punto de vista del usuario. Por lo tanto los casos de uso determinan los requisitos funcionales del sistema, es decir, representan las funciones que un sistema puede ejecutar.

Su ventaja principal es la facilidad para interpretarlos, lo que hace que sean especialmente útiles en la comunicación con el cliente.

A partir de los casos de uso del sistema se obtienen los requisitos de entrada y de salida de datos al almacén a implementar. Estos requisitos servirán de entrada para el diseño del almacén en forma de medidas y dimensiones.

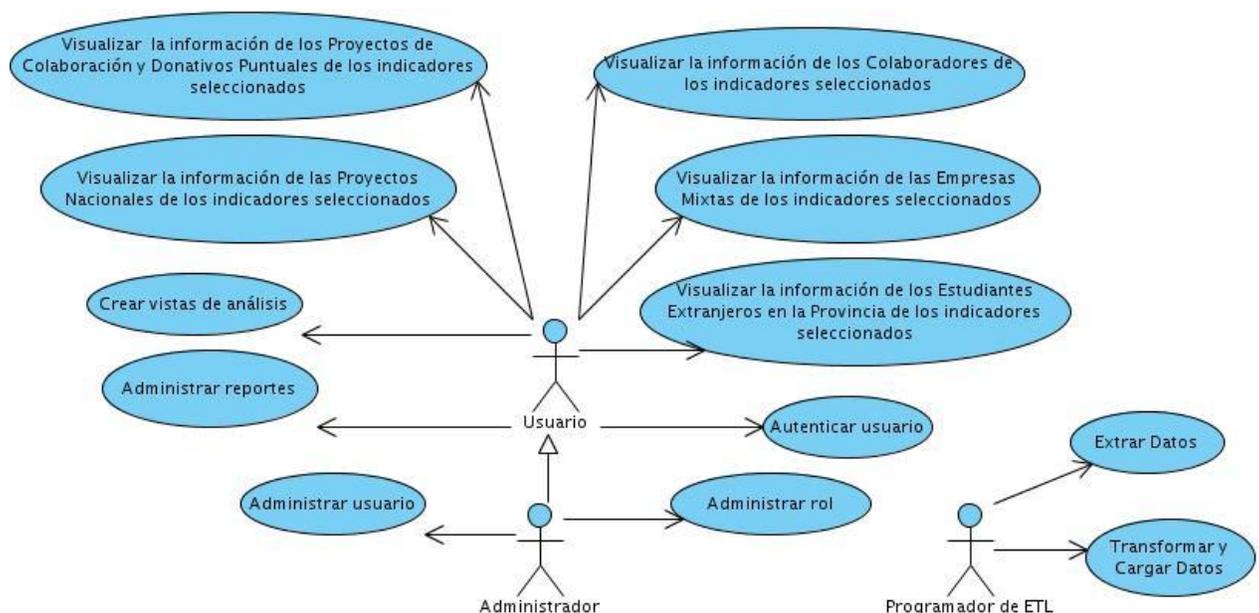


Ilustración 1: Diagrama de Casos de Uso del Sistema.

### 2.2 Diseño

#### 2.2.1 Modelo de datos

Tanto en las bases de datos tradicionales como en los almacenes de datos existen tres niveles de modelado de datos: conceptual, lógico y físico.

##### 2.2.1.1 Modelo de datos conceptual.

El modelo conceptual captura la información fundamental acerca de las entidades del dominio del problema y sus relaciones. Este modelo es más cercano al espacio del problema que al espacio de la solución.

##### 2.2.1.2 Modelo de datos lógico

El modelo lógico describe los datos en detalle, generalmente incluye todas las entidades y relaciones entre ellas, sus atributos y tipos de datos, así como las llaves primarias y extranjeras, sin tener en cuenta cómo ellos se implementarán físicamente en la base de datos. Es un puente entre el nivel conceptual y el físico.

##### 2.2.1.3 Modelo de datos físico

Este modelo describe las estructuras de almacenamiento y los métodos usados para tener un acceso efectivo a los datos.

##### 2.2.1.4 Dimensiones

Las dimensiones son las características de un hecho, que permiten su posterior análisis en el proceso de toma de decisiones; una entidad de negocio respecto a la cual se deben calcular las métricas. La solución propuesta cuenta con 18 dimensiones, algunas de las cuales tienen entre sus características que presentan jerarquías entre sus atributos. A continuación se describen cada una de las

dimensiones del Área de Análisis de Estudiantes Extranjeros:

dim_pais
dim_pais_id: INTEGER(10) [ PK ]
pais_nombre: VARCHAR(20)
region_nombre: VARCHAR(10)

Ilustración 2: Dimensión país.

- **dim\_pais:** Esta dimensión describe el universo de valores bajo los cuales puede clasificarse la información atendiendo el nomenclador del país.

**Jerarquía:** pais ->region

dim_temporal
dim_temporal_id: INTEGER(10) [ PK ]
mes_codigo: VARCHAR(10)
mes_nombre: VARCHAR(10)
mes_numero: INTEGER(10)
trimestre_codigo: VARCHAR(10)
trimestre_nombre: VARCHAR(10)
trimestre_numero: INTEGER(10)
semestre_codigo: INTEGER(10)
semestre_numero: INTEGER(10)
semestre_nombre: VARCHAR(10)
anno_codigo: VARCHAR(10)
anno_nombre: VARCHAR(10)
anno_numero: INTEGER(10)

Ilustración 3: Dimensión temporal.

- **dim\_temporal:** Esta dimensión describe el universo de valores bajo los cuales puede clasificarse la información atendiendo el nomenclador de temporal.

**Jerarquía:** anno ->semestre ->trimestre ->mes

dim_becario	
dim_becario_id:	INTEGER(10) [ PK ]
becario_nombre_apellidos:	VARCHAR(20)
becario_especialidad:	VARCHAR(20)
becario_anno_cursan:	VARCHAR(10)
becario_sexo:	VARCHAR(10)
becario_incidencias:	VARCHAR(10)

Ilustración 4: Dimensión becario.

- **dim\_becario:** Esta dimensión describe el universo de valores bajo los cuales puede clasificarse la información atendiendo el nomenclador del becario.

dim_centro	
dim_centro_id:	INTEGER(10) [ PK ]
centro_nombre:	VARCHAR(20)

Ilustración 5: Dimensión centro.

- **dim\_centro:** Esta dimensión describe el universo de valores bajo los cuales puede clasificarse la información atendiendo el nomenclador del centro.

### 2.2.1.5 Tablas de Hechos

Las tablas de hechos contienen las medidas y las dimensiones asociadas al mismo. Así como las claves primarias de las tablas de dimensiones relacionadas y los atributos de los registros de hechos. La solución propuesta cuenta con 8 tablas de hechos y 18 medidas que serán analizadas para apoyar el proceso de toma de decisiones.

A continuación se muestra la tabla de hechos para el Área de Análisis de Estudiantes Extranjeros:

hech_becario	
hech_becario_id:	INTEGER(10) [ PK ]
dim_pais_id:	INTEGER(10) [ FK ]
dim_centro_id:	INTEGER(10) [ FK ]
dim_temporal_id:	INTEGER(10) [ FK ]
dim_becario_id:	INTEGER(10) [ FK ]
cant_becarios:	INTEGER(10)

Ilustración 6: Hecho becario.

- **hech\_becarios:** En este hecho se encuentra toda la información referente a becarios.

En el artefacto Especificación del Modelo Dimensional (CFRA-Government-Especificaciones del Modelo de Datos Dimensional-Colaboración) v1.0 se encuentran declarados los gránulos, dimensiones, medidas, y tablas de hechos de cada una de las Áreas de Análisis que componen el mercado de datos.

## 2.2.1.6 Modelo Dimensional del mercado de datos.

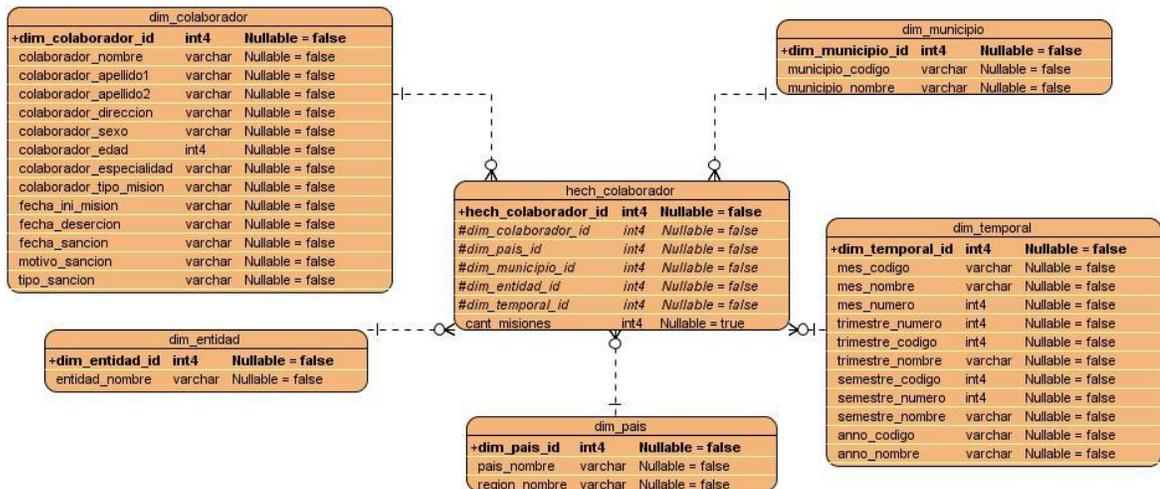


Ilustración 7: Modelo dimensional colaboradores.

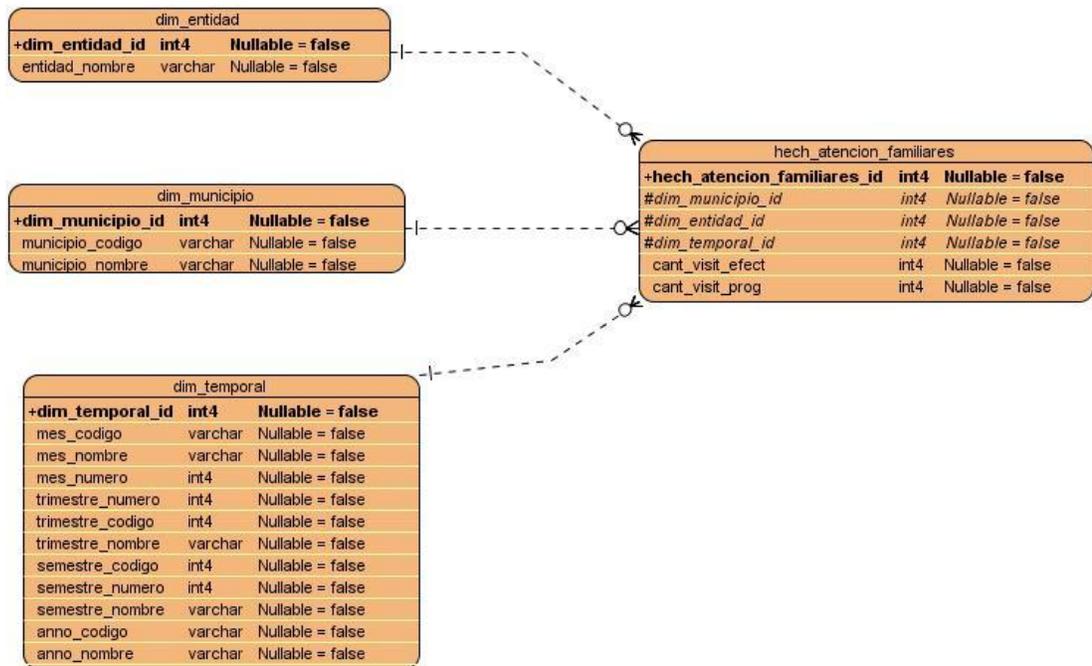


Ilustración 8: Modelo dimensional atención a familiares.

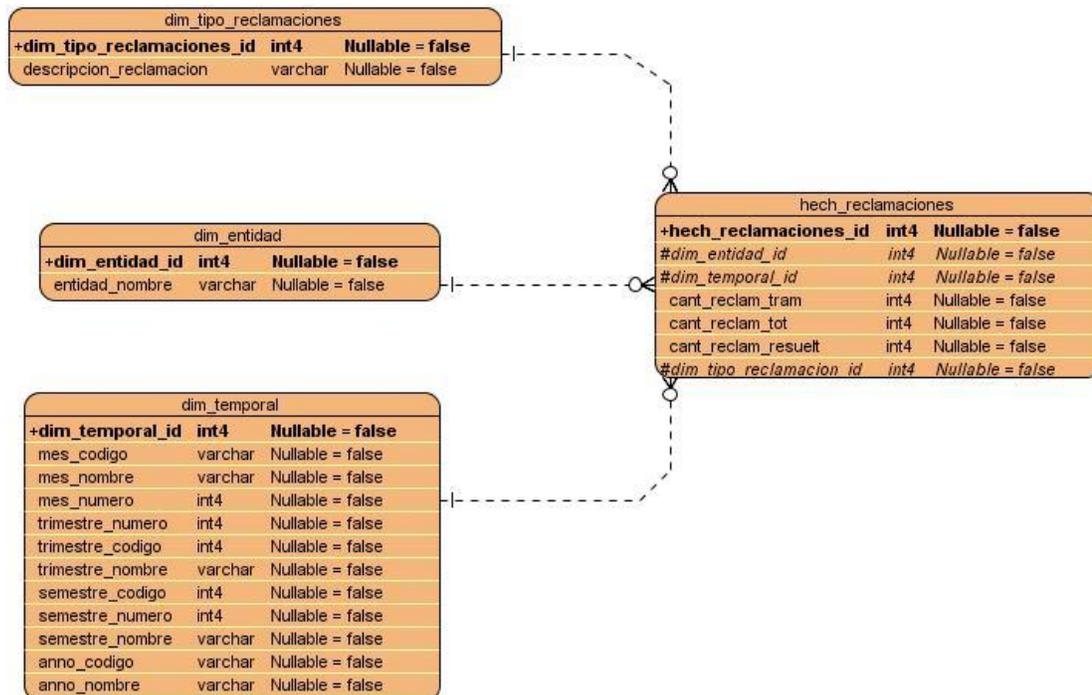


Ilustración 9: Modelo dimensional reclamaciones de colaboradores.

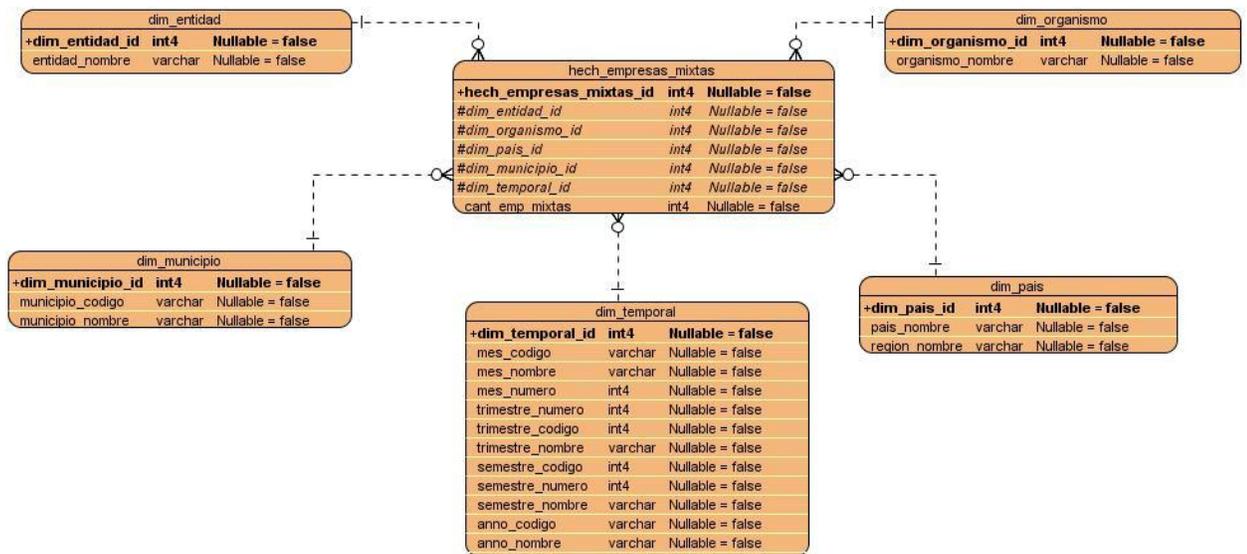


Ilustración 10: Modelo dimensional empresas mixtas.

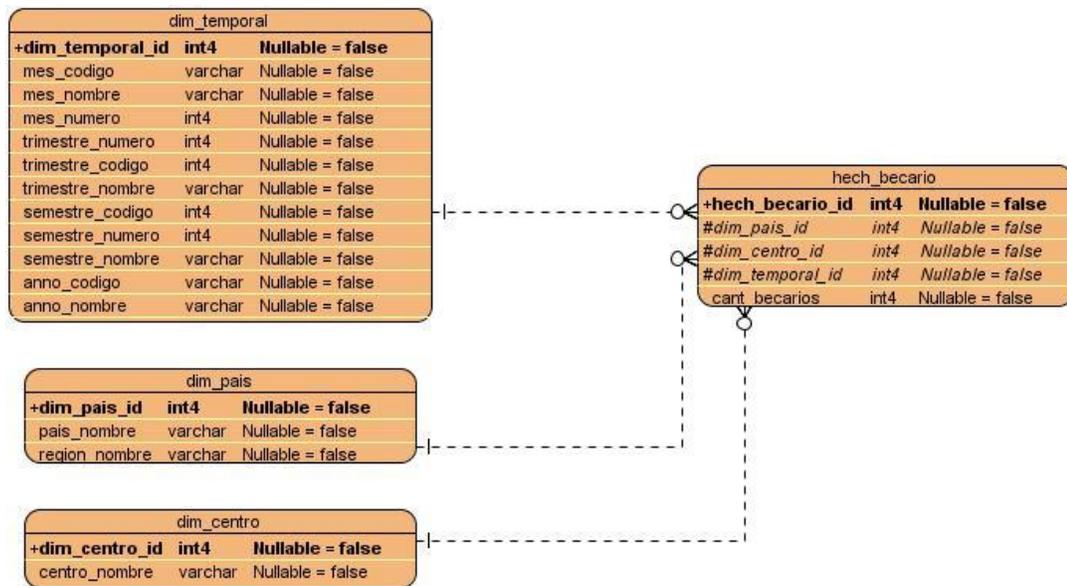


Ilustración 11: Modelo dimensional estudiantes extranjeros.

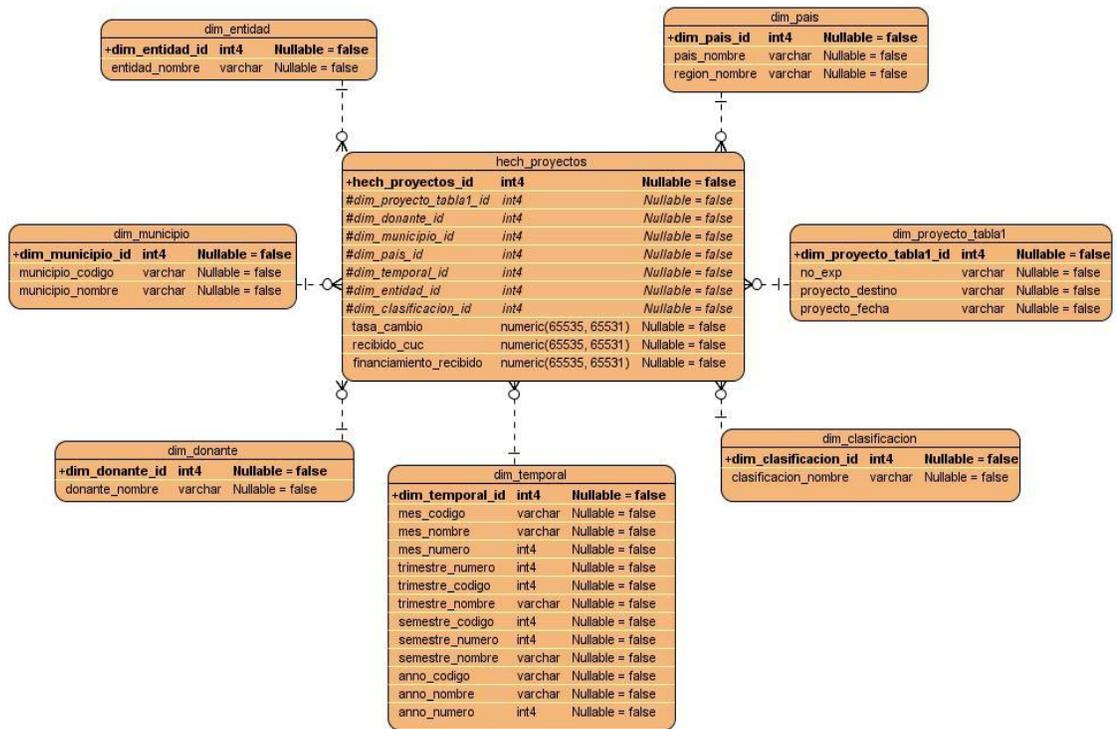


Ilustración 12: Modelo dimensional proyectos.

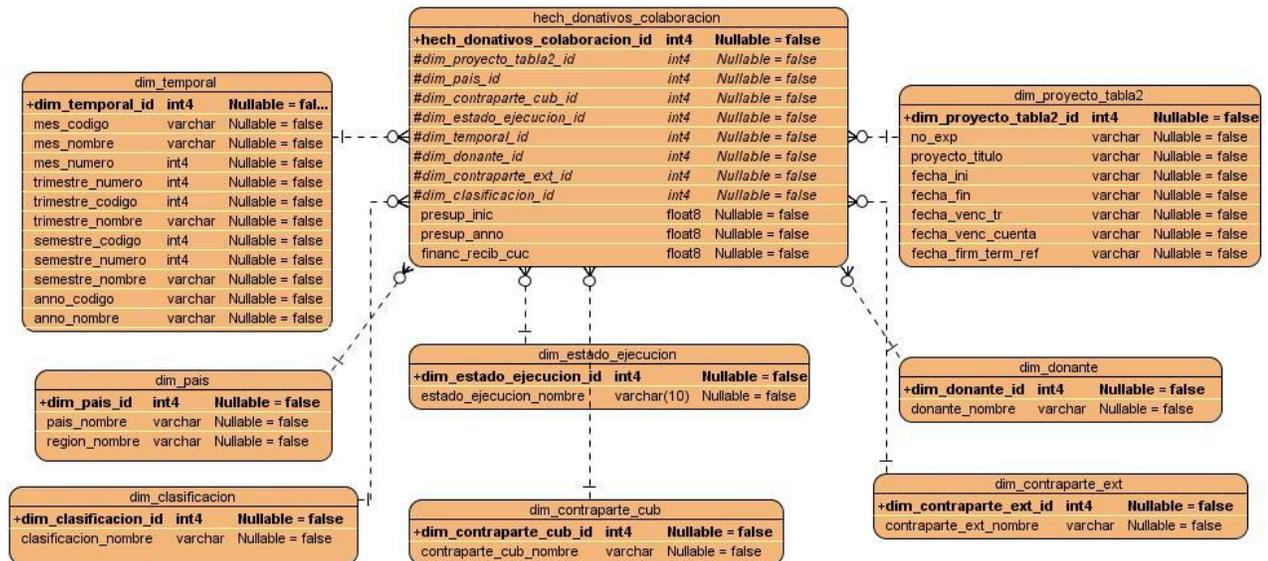


Ilustración 13: Modelo dimensional donativos colaboración.

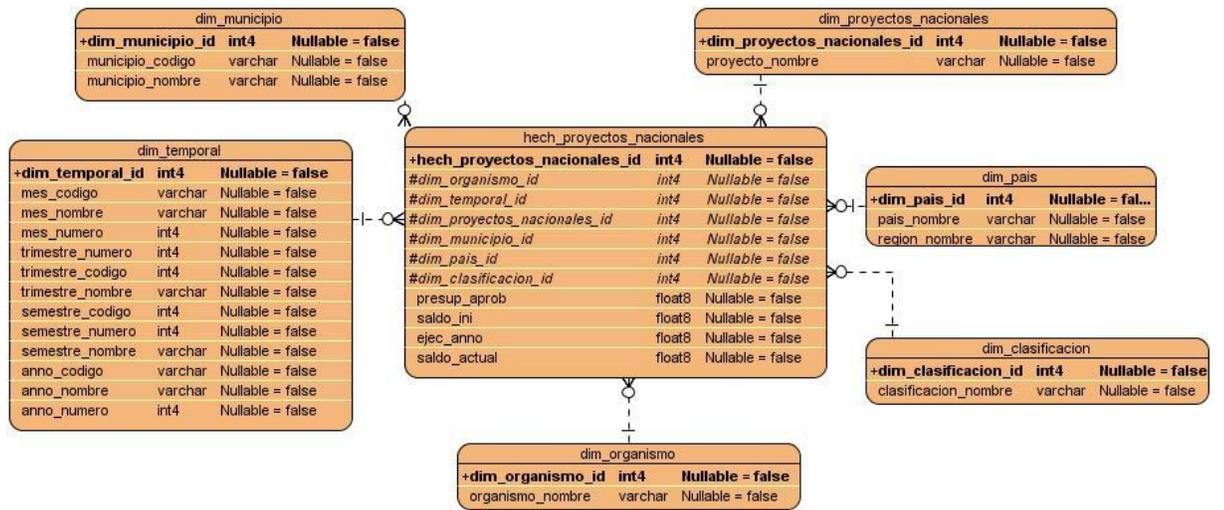


Ilustración 14: Modelo dimensional proyectos nacionales.

## 2.2.2 Matriz Bus

DIM/H	DIM 1	DIM 2	DIM 3	DIM 4	DIM 5	DIM 6	DIM 7	DIM 8	DIM 9	DIM 10	DIM 11	DIM 12	DIM 13	DIM 14	DIM 15	DIM 16	DIM 17	DIM 18
H1	X	X	X	X	X													
H2		X	X	X														
H3			X	X		X												
H4	X			X						X								
H5	X	X	X	X				X										
H6	X	X		X				X					X	X				
H7	X	X	X	X							X			X	X			
H8	X			X								X		X	X	X	X	X

Tabla 1: Matriz Bus.

### **2.2.3 Diseño del subsistema de integración.**

#### **2.2.3.1 Diseño de los procesos de ETL**

El proceso de ETL puebla la estructura del MD que va a servir de fuente para la carga de los cubos. La información almacenada en los cubos es la base para la realización de los reportes bajo la herramienta BI-Server.

El propósito fundamental de este proceso es fusionar los datos en un solo formato para el mercado de datos. Puesto que los datos de origen para la aplicación de la Inteligencia de Negocio se obtienen desde distintas fuentes. Para ello es importante tener en cuenta la fuente de datos y los destinos de los mismos, con el objetivo de identificar de donde provienen los datos y sus características. Una vez seleccionadas las tablas en las que se va a trabajar, es necesario definir el orden y la secuencia de las transformaciones para cada conjunto de datos. Es necesario señalar que todas las tablas de dimensión deben ser cargadas antes que las tablas de hechos.

Todas las transformaciones y trabajos diseñados se almacenarán a modo de seguridad, en el formato que provee la herramienta Pentaho Data Integration (ktr y kjb).

### **2.2.4 Diseño de los subsistemas de visualización de datos.**

Antes del proceso de visualización de los datos se crean los cubos multidimensionales a través de la herramienta Workbench. Por cada tabla de hecho se define un cubo que carga la información.

Lo más importante es como a los usuarios finales les llega la información, es decir la manera en que ellos van a interactuar con los datos que antes tenían en distintos lugares. En este sentido se utilizará la interfaz del BI-Server para la presentación de

la información.

A continuación se muestran los elementos que componen la estructura de navegación del mercado de datos para la Sección de Colaboración. Esta contiene 4 áreas de análisis, con 7 libros de trabajos y 160 reportes asociados a estos.

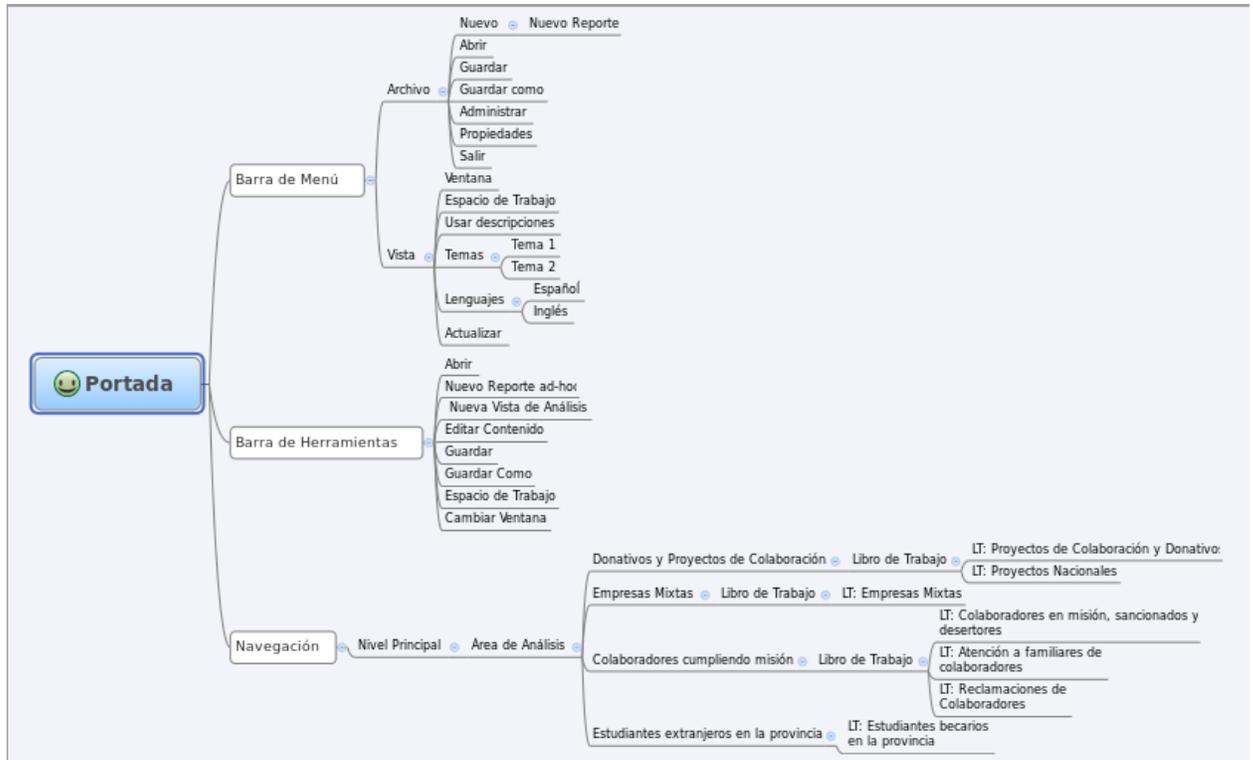


Ilustración 15: Mapa de navegación del mercado de datos para la Sección de Colaboración.

## 2.3 Conclusiones del capítulo

En este capítulo se detallaron los aspectos principales dentro del proceso de análisis y diseño de la solución. Concluyéndose los siguientes resultados:

1. Quedaron definidos los requisitos informativos, funcionales, y no funcionales, para dar respuesta a las necesidades de los usuarios.
2. Se definieron las áreas sobre las cuales se realizarán los análisis de la información.

3. Se definieron 8 tablas de hechos y 18 tablas de dimensiones.
4. Se especificó la arquitectura y se diseñaron los subsistemas de integración y visualización correspondientes con el mercado de datos para la Sección de Colaboración.

## *Capítulo 3: Implementación del Mercado de Datos.*

---

### **CAPÍTULO 3: “Implementación del mercado de datos de la Sección de Colaboración de la AP”:**

Para la realización del mercado de datos es necesario llevar a cabo un conjunto de transformaciones necesarias para 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 subsistema de visualización, los cubos OLAP, los reportes candidatos y la política de seguridad de los usuarios. Dichas acciones se desarrollarán en este capítulo.

#### **3.1 Implementación del subsistema de almacenamiento.**

Uno de los elementos más importantes a la hora de iniciar el desarrollo de cualquier proyecto es el modelado de datos. Cuando se diseña el modelo dimensional, el mismo se transforma a un modelo físico, del cual se genera el script de la base de datos, y es allí donde se evidencian las relaciones que existen entre las diferentes tablas, y a la vez determina si el proyecto va a cumplir con su verdadero objetivo.

##### **3.1.1 Esquemas y Tablas**

Los esquemas son una forma de organizar la información en una base de datos. Dentro de los esquemas se pueden encontrar funciones, operadores y tipos de datos que facilitarán su implementación. Los usuarios solamente tendrán acceso a aquellos que su rol se lo permita.

En el presente trabajo se definió un esquema:

**Esquema mart\_colaboración:** contiene todas las tablas de hechos y las dimensiones propias propuestas en el MD.

El diseño del modelo de datos, consta de 34 tablas en total, 18 dimensiones y 8 hechos. En la figura 8 se muestra utilizando la herramienta pgAdmin la estructura

física del MD.

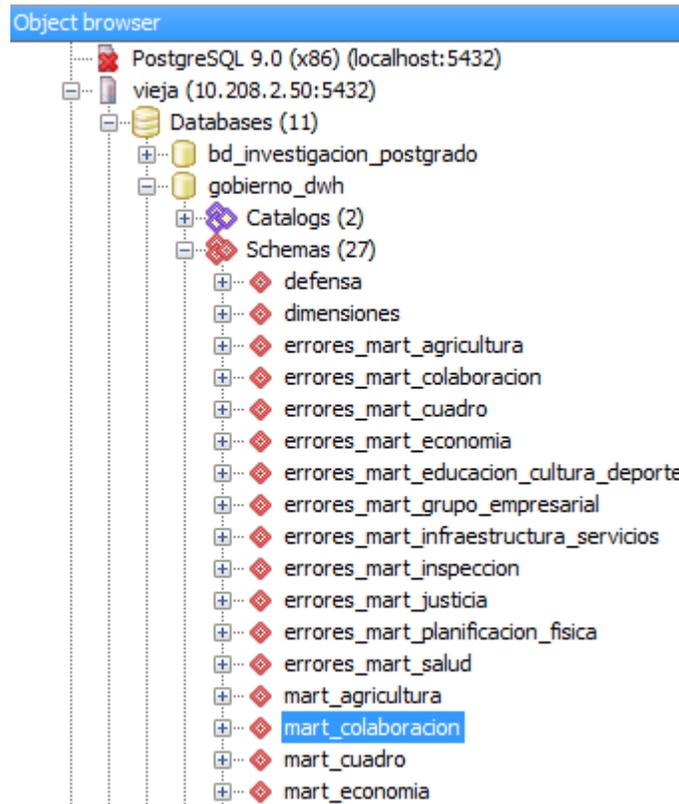


Ilustración 16: Esquema mart\_colaboracion.

## 3.2 Implementación de los subsistemas de integración y visualización de datos.

### 3.2.1 Implementación de las transformaciones

El proceso de Extracción, Transformación y Carga consiste en extraer los datos de las fuentes y se seleccionan los campos necesarios conforme al modelo de datos, luego estos datos se transforman, limpian y estandarizan para eliminar inconsistencias y posibles errores que pudieran llegar a existir. Una vez efectuada la extracción de los datos, se realizan las validaciones necesarias teniendo en cuenta las reglas del negocio identificadas y en caso de encontrarse valores incorrectos, el flujo se desvía hacia un fichero donde es almacenado con la

correspondiente descripción del error. Si se comprueba que los datos poseen la calidad requerida, se les realiza un conjunto de transformaciones y se procede a su inserción en el mercado de datos.

Para poblar el mercado de datos se realizaron dos tipos de cargas: cargas históricas a partir de la información en ficheros Excel y cargas incrementales teniendo como fuente de datos la base de datos relacional del SINAP (Sistema Informativo de AP). Se realizó una transformación por cada tabla de hecho y una por cada dimensión para un total de 45 transformaciones. A continuación, se describe de forma general la estrategia de integración definida para realizar los procesos de ETL.

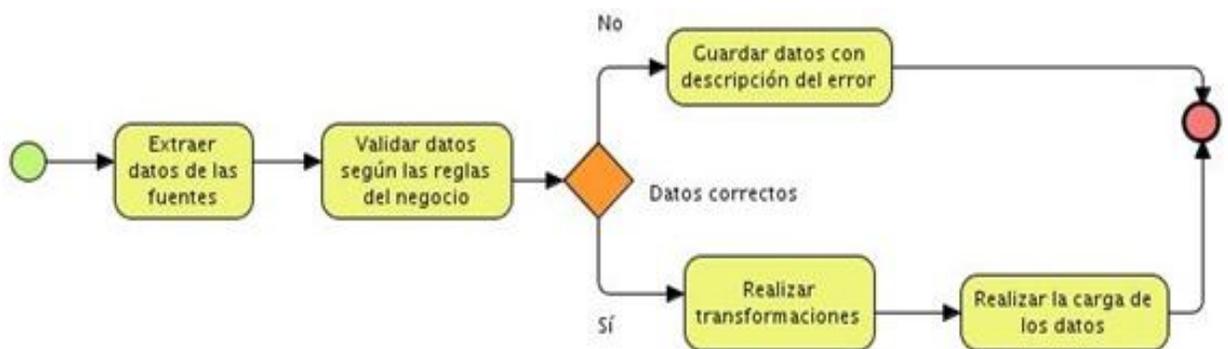


Ilustración 17: Diseño general de las transformaciones.

A continuación se ilustran algunos ejemplos de las transformaciones realizadas para poblar la base de datos correspondiente al mercado de datos para la Sección de Colaboración.

### 3.2.1.1 Carga de la dimensión dim\_becario desde el Excel.

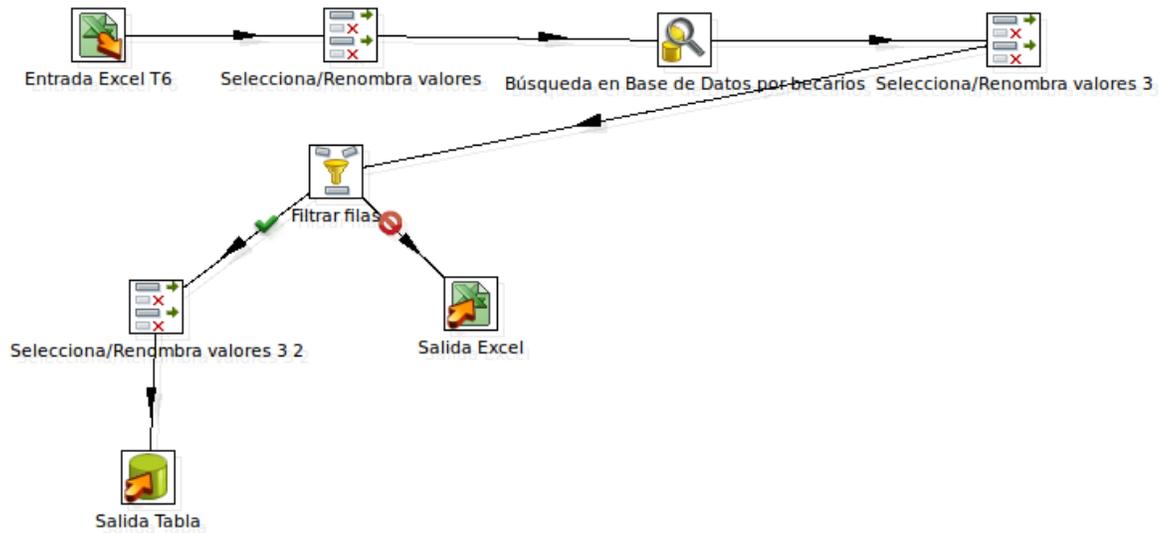


Ilustración 18: Transformación de la dimensión becario desde fichero Excel.

Para hacer la carga de la dimensión dim\_becario se obtienen los datos necesarios desde un fichero con formato Excel, luego se realiza una búsqueda para validar que los datos obtenidos no se repitan en la dimensión, después de obtener los valores que se repiten se pasa a un paso de filtrado donde solamente se insertarán en la dimensión aquellos valores que sean nuevos y los otros irán a un fichero que contendrá los errores o valores repetidos.

### 3.2.1.2 Carga del hecho hech\_proyectos desde el Excel.

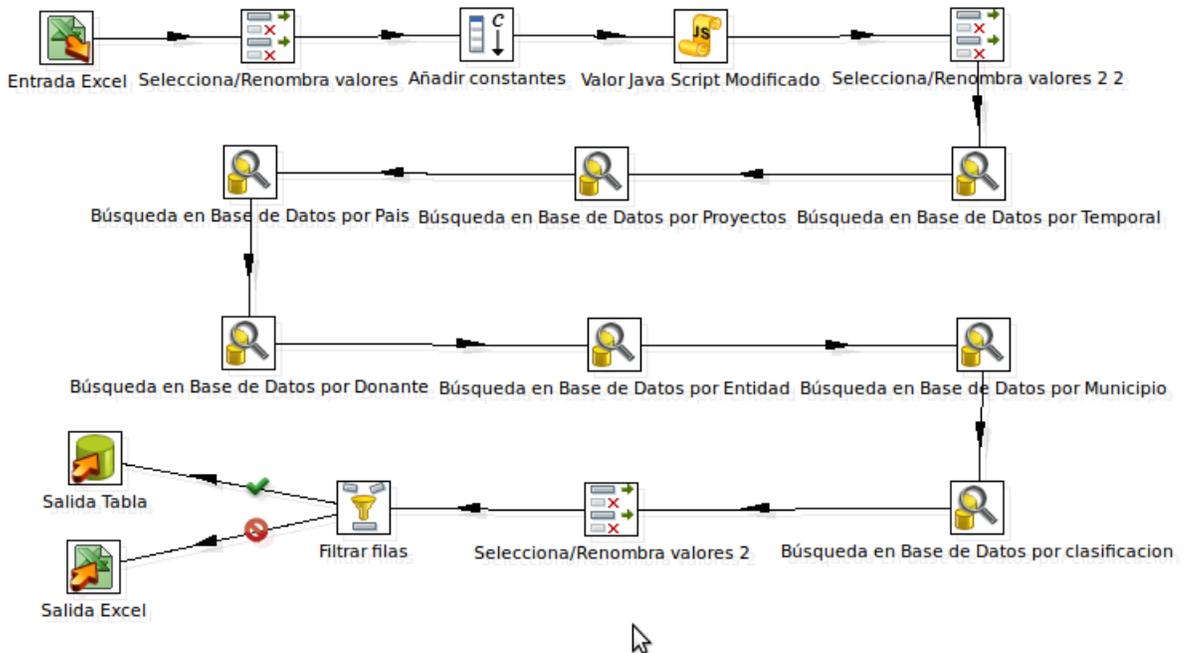


Ilustración 19: Transformación del hecho proyectos desde fichero Excel.

Para hacer la carga del hecho proyectos se obtienen los datos necesarios desde un fichero con formato Excel, se realiza una búsqueda por proyecto\_tabla1, donante, municipio, país, entidad, temporal y clasificación con el objetivo de validar que los id buscados en las dimensiones sean válidos. Para determinar el id de la dimensión tiempo se utilizó el componente añadir constantes para añadir las constantes que no vienen directamente de la fuente y así obtener el mes del campo fecha de la fuente para conocer cuando se realizó la extracción y por último se validan que no existan campos nulos para cargar solamente los campos válidos en la tabla de hecho proyectos.

## 3.2.1.3 Carga del hecho hech\_becarios desde el Excel.

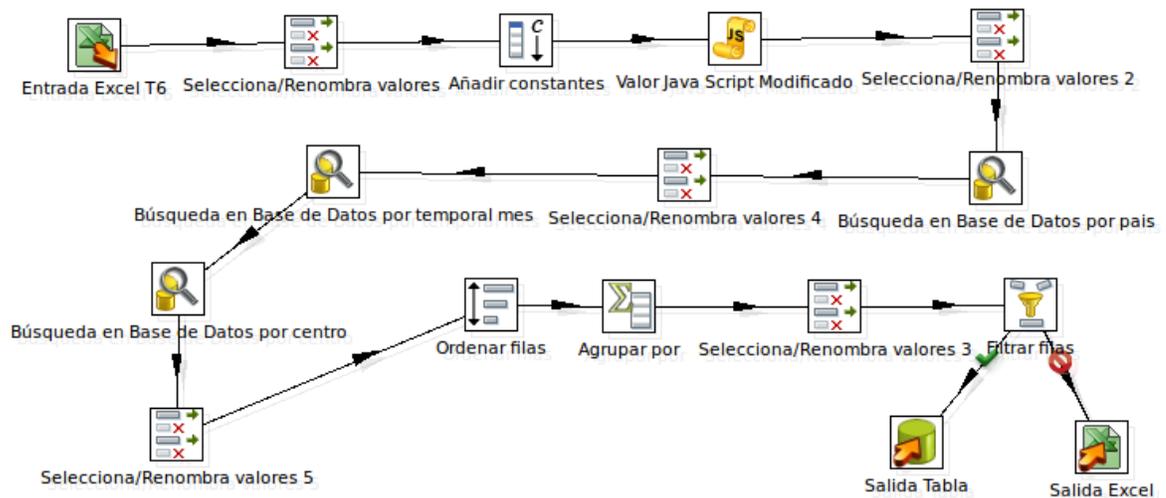


Ilustración 20: Transformación del hecho becarios desde fichero Excel.

Para hacer la carga del hecho becarios se obtienen los datos necesarios desde un fichero con formato Excel, se realiza una búsqueda por país y centros para validar que los id buscados en las dimensiones sean válidos y utilizar el componente agrupar por que agrupa por id de país y devuelve la cantidad de becarios por países. Para determinar el id de la dimensión tiempo se utilizó el componente añadir constantes para añadir las constantes que no vienen directamente de la fuente y así obtener el mes del campo fecha de la fuente para conocer cuando se realizó la extracción y por último se validan que no existan campos nulos para cargar solamente los campos válidos en la tabla de hecho becarios.

## 3.2.1.4 Carga de la dimensión dim\_proyecto\_tabla2 desde el SINAP.

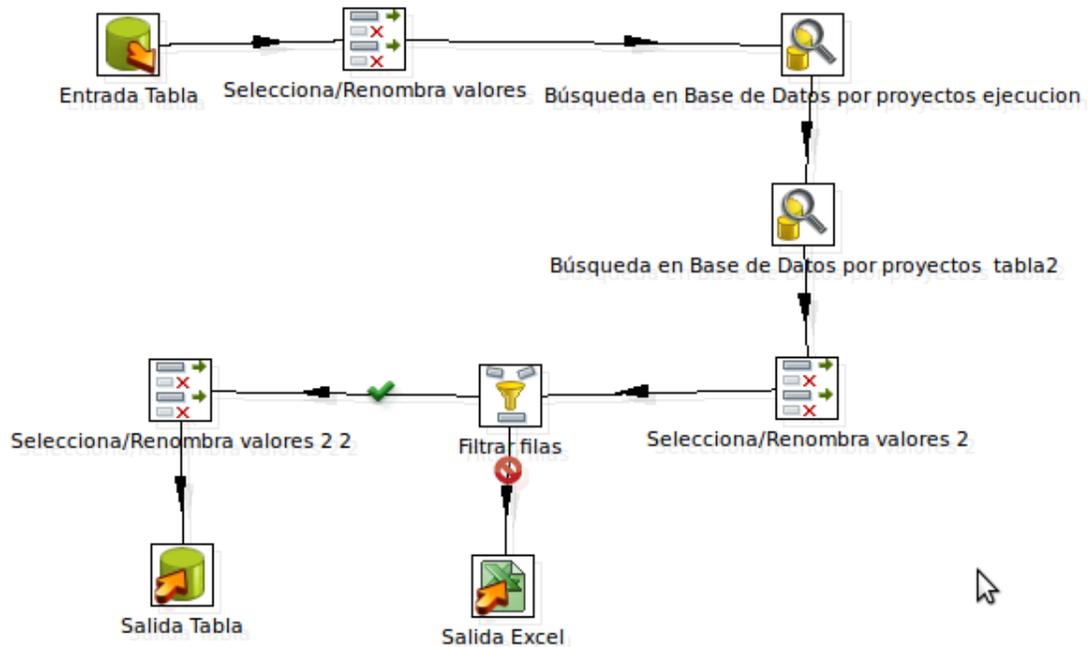


Ilustración 21: Transformación de la dimensión proyectos tabla 2 desde el SINAP.

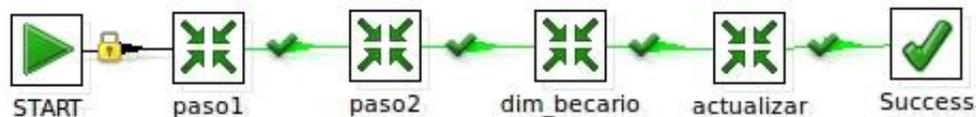
Para hacer la carga de la dimensión dim\_proyecto\_tabla2 se obtienen los datos necesarios de la fuente almacenados en una base de datos del Sistema Informativo de AP (SINAP), luego se procede a realizar una búsqueda para validar que los datos obtenidos no se repitan en la dimensión, después de obtener los valores que se repiten se pasa a un paso de filtrado donde solamente se insertarán en la dimensión aquellos valores que sean nuevos y los otros irán a un fichero que contendrá los errores o valores repetidos.

## 3.2.2 Implementación de los trabajos (Jobs).

Los jobs se utilizan para gestionar la carga de las transformaciones de las dimensiones así como la de los hechos. Este trabajo se realiza con un orden lógico

definido, primero las dimensiones y seguido los hechos. De forma general se realizó un trabajo por cada dimensión, uno por cada tabla de hecho y uno general que ejecute a los demás en dependencia de la frecuencia de carga de la información correspondiente a la Sección de Colaboración.

A continuación se muestra el trabajo realizado para la carga de la dimensión becario, el cual se ejecuta en el siguiente orden: primero se obtiene la fuente de la que se va a cargar (paso 1), el siguiente paso es obtener la fecha del servidor donde se encuentra la tabla fuente y la fecha de la última vez que cargó información al Mercado, esto último se hace consultando la tabla de metadatos del mercado de datos (paso 2), luego se llama a la transformación para llenar la dimensión (fc\_etl\_carga\_dim\_becario) y por último se actualiza la tabla de metadatos (fc\_etl\_actualizar\_metadatos).



*Ilustración 22: Trabajo para la carga de la dimensión becario.*

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

La Inteligencia de Negocio consiste en transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios. Para ello se realizarán un conjunto de tareas encaminadas a satisfacer las necesidades del cliente.

#### 3.2.3.1 Implementación de los Cubos OLAP.

Uno de los factores claves en el procesamiento analítico en línea son los cubos OLAP, estos proveen rápido acceso a los datos almacenados independientemente de la cantidad de datos en el cubo, al mismo tiempo son un subconjunto de datos del almacén de datos. Luego de haber realizado los procesos de ETL se realizó la creación de los cubos correspondientes al mercado de datos. En la presente investigación se diseñaron 8 cubos multidimensionales, uno para cada tabla de hecho, especificando en los cubos las dimensiones y medidas, todas ellas enfocadas a las necesidades del cliente para darle solución al problema planteado. Además se diseñaron 8 cubos multidimensionales correspondientes a los hechos de las vistas materializadas. A continuación se presenta la estructura general de los cubos utilizados dentro de la solución propuesta.



Ilustración 23: Diseño de los cubos OLAP.

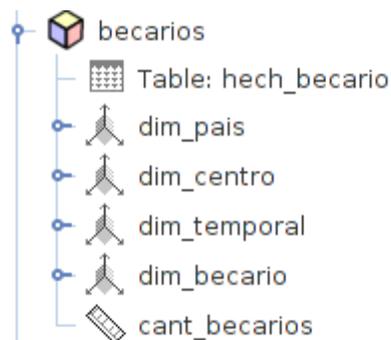


Ilustración 24: Elementos contenidos dentro del cubo becarios.

### 3.2.3.2 Vistas de Análisis.

La solución propuesta cuenta con 120 vistas de análisis, con el objetivo de satisfacer las necesidades de información de la Sección de Colaboración. El resultado del análisis realizado se muestra a través de la herramienta Pentaho BI-Server ya que es posible ver los reportes y analizar la información mediante tablas de datos o gráficas de diversos tipos. A continuación se muestra una vista de análisis correspondiente a las reclamaciones de los colaboradores.

		Medidas	
Entidad	Tiempo	• Cantidad_Reclamaciones_Tramite	• Cantidad_Reclamaciones_Resueltas
Salud	+ 2012	6,00	3,00
Deporte	+ 2012	3,00	2,00
Cultura	+ 2012	2,00	0,00

*Ilustración 25: Vista de análisis de las reclamaciones de los colaboradores.*

En la vista de análisis que se presentó anteriormente se miden dos medidas físicas pertenecientes al hecho reclamaciones y que se definieron en el diseño del cubo OLAP que se corresponde con la tabla de hecho, estas medidas son cantidad de reclamaciones en trámite y cantidad de reclamaciones resueltas por las dimensiones entidad y año todas ellas pueden ser analizadas desde diferentes perspectivas de análisis.

### 3.2.3.3 Reportes.

En la implementación de la solución se generaron 40 reportes diseñados mediante la herramienta Report Designer de Pentaho, la cual permite consultar los datos de distintas fuentes y ponerlos a disposición de los usuarios en diferentes formatos. Además de estos reportes el usuario podrá hacer sus propios reportes a partir de la información de los metadatos definidos en el sistema con el uso de la herramienta Pentaho Metadata Editor.

### 3.3 Vistas Materializadas

Las vistas materializadas se utilizan para analizar datos que pertenecen al año actual o años venideros, según se desee analizar. Además se utilizan para mejorar el rendimiento de la base de datos, las mismas se realizaron en el SGBD PostgreSQL y se definieron 8 vistas materializadas.

### 3.4 Seguridad de los datos.

El esquema de seguridad está respaldado por los niveles de acceso, específicamente por los roles definidos. La arquitectura del sistema de seguridad y alta disponibilidad debe ser de 3 capas:

- ✓ Funcionamiento: dispositivos de seguridad (firewall, inspectores de contenido, sensores).
- ✓ Servidores: servidor de gestión de administración y de base de datos.
- ✓ Presentación: consolas de administración.

La configuración y prueba de todos los equipos debe efectuarse en las instalaciones de la AP de Artemisa, conjuntamente con el personal técnico de informática.

#### 3.4.1 Política de Respaldo y Seguridad.

La política de respaldo y recuperación que utiliza la solución es sencilla pero a la vez sólida, por ello se miden 3 puntos esenciales:

1. **Periodicidad de las salvas:** mensualmente se realizan las salvas de toda la información contenida en la base de datos.
2. **Tablas involucradas:** las tablas que se involucran son todas las tablas contenidas en la base de datos, tablas de hechos, dimensiones y todas las que se involucren con estas.

3. **Backups existentes:** actualmente no existe Backup ya que no se ha efectuado la primera salva.

### 3.5 Conclusiones del capítulo

En este capítulo se abordaron los elementos que permitieron el desarrollo de la implementación del mercado de datos. Se realizó la implementación del modelo de datos definiéndose las dimensiones, las medidas relacionadas a cada uno de los hechos definidos para lograr una mejor estructuración de la información. Se realizó el proceso de ETL, extrayendo los datos de las series históricas que presentaba la Sección de Colaboración para luego realizar la carga de los datos a la base de datos. Luego de tener los datos cargados se realizó el proceso de inteligencia de negocio, donde se diseñaron los cubos OLAP, se implementaron las vistas de análisis correspondientes y se realizó la visualización de los datos convirtiéndolos en información valiosa para los usuarios.

# *Capítulo 4: Validación de las funcionalidades del mercado de datos para la Sección de Colaboración*

---

## **Capítulo 4: “Validación de las funcionalidades mercado de datos para la Sección de Colaboración”.**

En este capítulo se exponen las pruebas realizadas al mercado de datos para la Sección de Colaboración, así como los resultados obtenidos de cada una de ellas luego de su aplicación. Dichas pruebas son realizadas para garantizar el cumplimiento de las exigencias del cliente y la calidad del producto.

### **4.1 Validación y Prueba**

Desarrollar un producto o servicio de buena calidad y aceptación por el cliente, constituye un requisito indispensable para que el ingeniero de software pueda ganar determinada reputación dentro de la entidad donde se desempeña como profesional y también la empresa u organización a la que pertenecen.

La forma más eficaz de evitar que los errores se produzcan y se propaguen durante el desarrollo de un producto de software es disponer de procedimientos de calidad y pruebas que acompañen al producto a lo largo de su ciclo de vida.

#### **4.1.1 Diseño de los Casos de Prueba**

Para lograr la calidad del producto de software es necesario realizar un conjunto de evaluaciones durante todo el proceso de desarrollo que implique al cliente y desarrollador. Para ello se diseñaron casos de pruebas basados en los casos de uso, para verificar hasta qué punto cumplía con las necesidades del cliente. Para el mercado de datos para la Sección de Colaboración se diseñaron 5 casos de prueba, a continuación se muestra un escenario del diseño de caso de prueba correspondiente al caso de uso “Visualizar la información de los Colaboradores de los indicadores seleccionados”.

Escenario	Descripción	Perfiles de análisis	Indicadores a medir	Respuesta del sistema	Flujo central	Resultado de la prueba
<b>C1 Atención a Familiares</b>	Muestra los reportes de los indicadores de atención a familiares.	municipio	Cantidad de visitas programadas	El sistema muestra los reportes de los indicadores de atención a familiares.	<p>Se abre la aplicación.</p> <p>Se autentifica.</p> <p>Se entra al sistema.</p> <p>En el lateral izquierdo donde aparece el navegador se selecciona el área de análisis de A.A Colaboradores cumpliendo misión.</p> <p>Se selecciona el libro de trabajo de L.T Atención a Familiares.</p> <p>En la parte inferior izquierda se selecciona el reporte deseado.</p> <p>En el área de trabajo se visualiza la tabla correspondiente al reporte.</p>	Satisfactorio
		entidad	Cantidad de visitas			
		tiempo	efectuadas			

*Tabla 2: Escenario del diseño de caso de prueba correspondiente al caso de uso “Visualizar la información de los Colaboradores de los indicadores seleccionados”.*

Partiendo de los roles y permisos que los usuarios poseen en su interacción con la aplicación, se realizó el caso de prueba de seguridad Administrar Rol. **Ver anexo 5**

También se realizaron las pruebas unitarias y de integración del mercado de datos para verificar el correcto funcionamiento de los procesos de ETL obteniendo resultados satisfactorios.

### **4.1.2 Validación de la investigación.**

Para validar la presente investigación se realizaron casos de pruebas funcionales que no son más que los requisitos funcionales que debe cumplir el mercado de datos diseñado, seguidamente se pasó a un proceso de certificación de calidad del software por el grupo de calidad del centro de desarrollo de la Facultad Regional donde este grupo hizo la certificación principalmente basada en la funcionalidad, estandarización y limpieza del código y revisó los artefactos documentales que son generados por la metodología adoptada por la universidad, la cual fue seleccionada para el desarrollo de esta investigación.

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

Se estableció el Plan de pruebas para probar las funcionalidades y características del mercado de datos para la Sección de Colaboración. En este documento se identifican los escenarios de prueba y se establecen los roles y responsabilidades en esta etapa.

## **4.2 Resultados Obtenidos**

Se obtuvo el mercado de datos disponible en su versión 1.0, el cual cumple con todos los requisitos, para contribuir en la rapidez y seguridad en el análisis de la información en apoyo a la toma de decisiones de la Sección de Colaboración del AP de Artemisa.

### 4.3 Funcionalidades Obtenidas

Las funcionalidades obtenidas más importantes del mercado de datos en su versión 1.0 permiten:

- Realizar los reportes de los modelos.
- Crear Vistas de Análisis.
- Crear Reportes ad-hoc.
- Proporcionar al usuario una interfaz consolidada, única para los datos, que hace más fácil el trabajo con las consultas para la toma de decisiones.

### 4.4 Aporte Social y Económico

El trabajo que se realiza en la Sección de Colaboración de la AP de Artemisa, según la problemática planteada, ha traído consecuencias negativas en cuanto a la calidad y análisis de la información que se maneja. El mercado de datos para dicha sección brinda a los especialistas de la misma guardar los datos durante períodos de tiempo extensos permitiendo el acceso a datos históricos por años, además de lograr un análisis de la información en apoyo a la toma de decisiones y brinda una vía más cómoda, eficiente y rápida de realizar su trabajo. Debido a la eficiencia y optimización de este sistema, se verán reducidas las pérdidas de datos contables que son fundamentales en la economía de la provincia.

### 4.5 Conclusiones del capítulo

En este capítulo se expuso la validación del mercado de datos. Obteniéndose los siguientes resultados:

1. Se diseñaron y aplicaron cinco casos de pruebas basados en casos de uso.
2. Se realizaron las pruebas unitarias y de integración para validar la calidad de los procesos de ETL.
3. Se elaboró un Plan de pruebas para probar las funcionalidades del mercado de datos para la Sección de Colaboración.

### **Conclusiones Generales.**

Los objetivos propuestos en el presente trabajo de diploma han sido cumplidos satisfactoriamente. Se logró obtener un mercado de datos para la Sección de Colaboración de la AP de Artemisa, facilitando el análisis estadístico para la toma de decisiones.

- Se realizó la fundamentación teórica de los almacenes de datos, el cual brindó un aporte para la solución de la problemática planteada.
- Se caracterizó el proceso de análisis de la información en la Sección de Colaboración de la AP de Artemisa.
- Se desarrolló del mercado de datos para la Sección de Colaboración.
- Se validó funcionalmente el mercado de datos para la Sección de Colaboración.

### **Recomendaciones.**

Se recomienda para el desarrollo futuro del sistema:

- ✓ Poblar el mercado de datos con toda la información existente en la Sección de Colaboración de la AP de Artemisa.
- ✓ Una vez poblado el mercado de datos, mantener actualizado el contenido del mismo.

## Referencias Bibliográficas.

**2007.** [Online] marzo 5, 2007. Disponible en: [http://www.freedownloadmanager.org/es/downloads/Visual\\_Paradigm\\_for\\_UML\\_%28LE%29\\_%5BWindows%5D\\_14728\\_p/](http://www.freedownloadmanager.org/es/downloads/Visual_Paradigm_for_UML_%28LE%29_%5BWindows%5D_14728_p/).

**2009.** [Online] 2009. Disponible en: [http://www.leroot.com/site/index.php?option=com\\_content&view=article&id=63&Itemid=69](http://www.leroot.com/site/index.php?option=com_content&view=article&id=63&Itemid=69).

**2010.** [Online] octubre 19, 2010. Disponible en: <http://www.tuinformaticafacil.com/herramientas-desarrollo/sql-power-architect-herramienta-de-modelado-de-datos>.

**2011.** [Online] msdn, 2011. Disponible en: <http://msdn.microsoft.com/es-es/library/ms244679%28v=vs.80%29.aspx>.

**2011.** [Online] 2011. Disponible en: <http://kettle.pentaho.com/>.

**Beatriz Lamancha Pérez. 2006.** [Online] 2006. [Cited: abril 4, 2012.] Disponible en: <http://www.fing.edu.uy>.

**Chinchilla Arley, Ricardo. 2011.** Mercado de datos: conceptos y metodologías de desarrollo. [Online] Julio-Setiembre 2011. Disponible en: [http://www.tec.ac.cr/sitios/Vicerrectoria/vie/editorial\\_tecnologica/Revista\\_Tecnologia\\_Marcha/pdf/tecnologia\\_marcha\\_24-3/TM%2024-3%20art%206.pdf](http://www.tec.ac.cr/sitios/Vicerrectoria/vie/editorial_tecnologica/Revista_Tecnologia_Marcha/pdf/tecnologia_marcha_24-3/TM%2024-3%20art%206.pdf).

**Diaz Ordaz . 2011.** [Online] Gravatar, 2011. Disponible en: <http://www.gravatar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.

**Ing. Bernabeu, Ricardo Dario. 2007.** [Online] noviembre 2007. Disponible en: <http://es.scribd.com/doc/55234594/45/Esquema-Constelacion>.

**Jaime Casanova. 2009.** [Online] Asesoría y desarrollo de sistemas, julio 1, 2009. Disponible en: <http://archives.postgresql.org/pgsql-es-fomento/2009-07/msg00000.php>.

**Johan. 2011.** [Online] Inqbaton, marzo 7, 2011. Disponible en: <http://www.inqbaton.com/pentaho-data-integration/>.

**Jonathan David Ramos. 2010.** SOLUCIONES OLAP CON MICROSOFT SQL SERVER ANALYSIS SERVICES. [Online] 2010. Disponible en: <http://www.eumed.net/libros/2009c/574/TRABAJANDO%20CON%20CUBOS.htm>.

## Referencias Bibliográficas

---

**Mgter. David Luis la Red Martinez. 2007.** *Procesamiento Analítico en Línea (OLAP)*. [Online] 2007. Disponible en: <http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/OLAPMonog.pdf>.

**Outeiro, Ronda. 2011.** [Online] Sinnexus, 2011. Disponible en: [http://www.sinnexus.com/business\\_intelligence/datawarehouse.aspx](http://www.sinnexus.com/business_intelligence/datawarehouse.aspx).

**Salazar Luján, Ricardo. 2008.** Data Warehouse para la prestación del servicio público de información estadística. [Online] 2008.

**SAS® Enterprise. 2011.** [Online] 2011. Disponible en: <http://www.sas.com/offices/latinamerica/mexico/technologies/bi/entbiserver/index.html>.

**Sherman Wood, JasperSoft. 2007.** [Online] abril 2007. Disponible en: <http://mondrian.pentaho.com/documentation/workbench.php>.

**Torres, Lic. Liudmila Padrón. 2006.** [Online] Empresa de Telecomunicaciones de Cuba S.A (ETECSA V.C.), 01 01, 2006. Disponible en: <http://www.ilustrados.com/tema/8205/Almacenes-Datos-Importancia-Estandarizacion-direcciones-para.html>.

## Bibliografía consultada.

**2007.** [Online] marzo 5, 2007. Disponible en: [http://www.freedownloadmanager.org/es/downloads/Visual\\_Paradigm\\_for\\_UML\\_%28LE%29\\_%5BWindows%5D\\_14728\\_p/](http://www.freedownloadmanager.org/es/downloads/Visual_Paradigm_for_UML_%28LE%29_%5BWindows%5D_14728_p/).

**2009.** [Online] 2009. Disponible en: [http://www.leroot.com/site/index.php?option=com\\_content&view=article&id=63&Itemid=69](http://www.leroot.com/site/index.php?option=com_content&view=article&id=63&Itemid=69).

**2010.** [Online] octubre 19, 2010. Disponible en: <http://www.tuinformaticafacil.com/herramientas-desarrollo/sql-power-architect-herramienta-de-modelado-de-datos>.

**2011.** [Online] msdn, 2011. Disponible en: <http://msdn.microsoft.com/es-es/library/ms244679%28v=vs.80%29.aspx>.

**2011.** [Online] 2011. Disponible en: <http://kettle.pentaho.com/>.

**Beatriz Lamancha Pérez. 2006.** [Online] 2006. [Cited: abril 4, 2012.] Disponible en: <http://www.fing.edu.uy>.

**Chinchilla Arley, Ricardo. 2011.** Mercado de datos: conceptos y metodologías de desarrollo. [Online] Julio-Setiembre 2011. Disponible en: [http://www.tec.ac.cr/sitios/Vicerrectoria/vie/editorial\\_tecnologica/Revista\\_Tecnologia\\_Marcha/pdf/tecnologia\\_marcha\\_24-3/TM%2024-3%20art%206.pdf](http://www.tec.ac.cr/sitios/Vicerrectoria/vie/editorial_tecnologica/Revista_Tecnologia_Marcha/pdf/tecnologia_marcha_24-3/TM%2024-3%20art%206.pdf).

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

**Diaz Ordaz . 2011.** [Online] Gravatar, 2011. Disponible en: <http://www.gravatar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.

**Ing. Bernabeu, Ricardo Dario. 2007.** [Online] noviembre 2007. Disponible en: <http://es.scribd.com/doc/55234594/45/Esquema-Constelacion>.

**Jaime Casanova. 2009.** [Online] Asesoría y desarrollo de sistemas, julio 1, 2009. Disponible en: <http://archives.postgresql.org/pgsql-es-fomento/2009-07/msg00000.php>.

**Johan. 2011.** [Online] Inqbaton, marzo 7, 2011. Disponible en: <http://www.inqbaton.com/pentaho-data-integration/>.

**Jonathan David Ramos. 2010.** SOLUCIONES OLAP CON MICROSOFT SQL SERVER ANALYSIS SERVICES. [Online] 2010. Disponible en: <http://www.eumed.net/libros/2009c/574/TRABAJANDO%20CON%20CUBOS.htm>.

**Kimball, Ralph. 2008.** The Data Warehouse Lifecycle Toolkit. EUA : Wiley Publishing Inc, 2008.

**Lopez, Fabian. 2011.** *Sistema de Información de Gobierno. Mercado de datos para las áreas de Cultura y Deporte.* 2011.

**Lora, Heidy Carmona. 2007.** *Modelo y Diseño de un DWH usando vistas materializadas.* 2007.

**Mgter. David Luis la Red Martinez. 2007.** *Procesamiento Analítico en Línea (OLAP).* [Online] 2007. Disponible en: <http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/OLAPMonog.pdf>.

**Outeiro, Ronda. 2011.** [Online] Sinnexus, 2011. Disponible en: [http://www.sinnexus.com/business\\_intelligence/datawarehouse.aspx](http://www.sinnexus.com/business_intelligence/datawarehouse.aspx).

**Salazar Luján, Ricardo. 2008.** Data Warehouse para la prestación del servicio público de información estadística. [Online] 2008.

**SAS® Enterprise. 2011.** [Online] 2011. Disponible en: <http://www.sas.com/offices/latinamerica/mexico/technologies/bi/entbiserver/index.html>.

**Sherman Wood, JasperSoft. 2007.** [Online] abril 2007. Disponible en: <http://mondrian.pentaho.com/documentation/workbench.php>.

**Torres, Lic. Liudmila Padrón. 2006.** [Online] Empresa de Telecomunicaciones de Cuba S.A (ETECSA V.C.), 01 01, 2006. Disponible en: <http://www.ilustrados.com/tema/8205/Almacenes-Datos-Importancia-Estandarizacion-direcciones-para.html>.