

# Universidad de las Ciencias Informáticas

## Facultad 6



**Título: “Mercado de datos para las Series históricas de comercio interno del Sistema de Información de Gobierno.”**

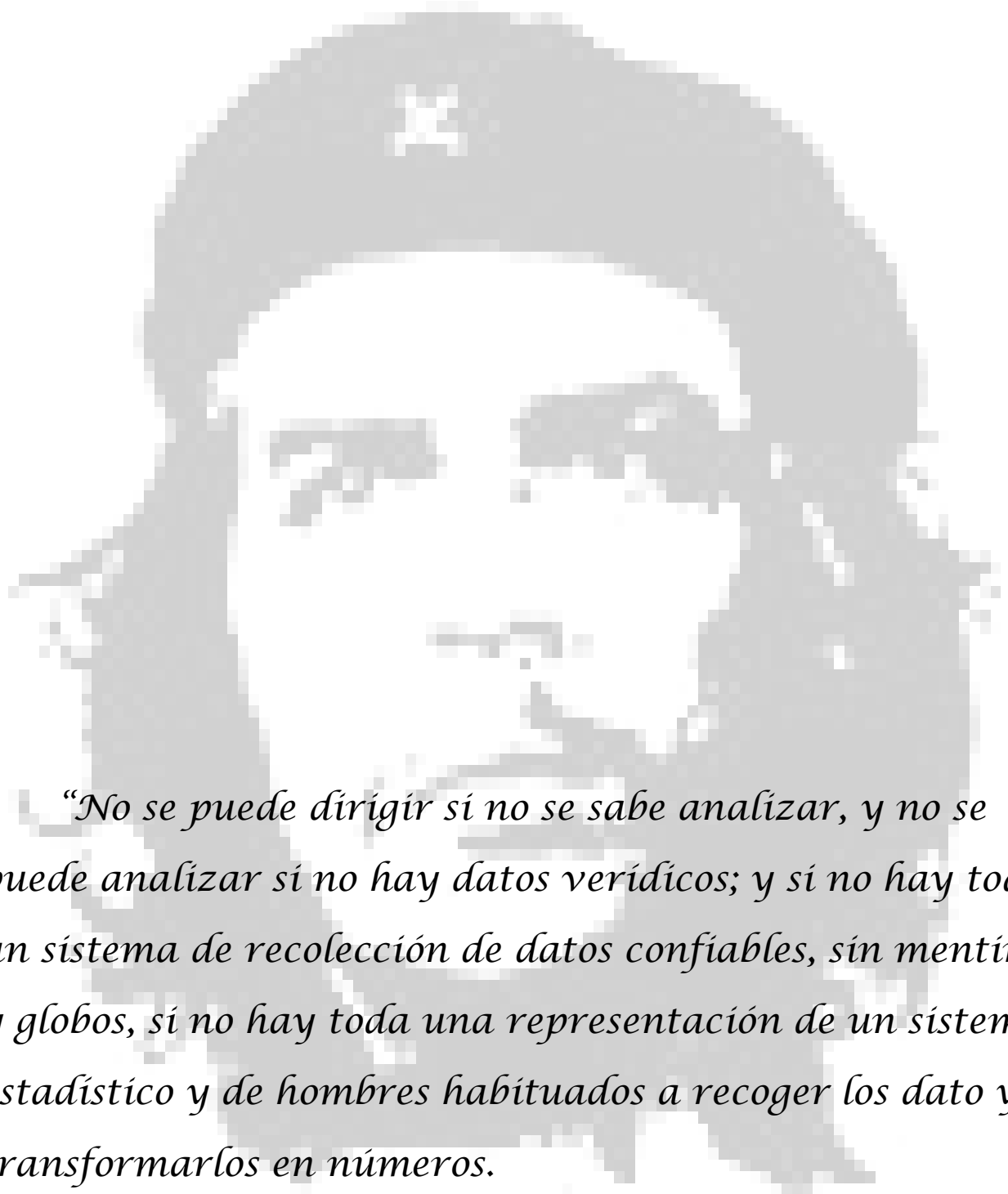
**Trabajo de Diploma para optar por el título de  
Ingeniero en Ciencias Informáticas**

**Autora:** Daylí Alemán Robaina

**Tutores:** Ing. Yonelbys Iznaga González  
Ing. Yordanka Echevarría Melo

La Habana, Junio del 2012

“Año 54 de la Revolución Cubana”



*“No se puede dirigir si no se sabe analizar, y no se puede analizar si no hay datos verídicos; y si no hay todo un sistema de recolección de datos confiables, sin mentiras y globos, si no hay toda una representación de un sistema estadístico y de hombres habituados a recoger los dato y transformarlos en números.*

*Esta es una tarea esencial”*

*Ernesto (Ché) Guevara*

# DECLARACIÓN DE AUTORIA

---

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

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

Daylí Alemán Robaina

---

**Firma de la autora**

Ing. Yonelbys Iznaga González

---

**Firma del tutor**

Ing. Yordanka Echevarría Melo

---

**Firma del tutor**

**Ing. Yonelbys Iznaga González:** Graduado de Ingeniería en Ciencias Informáticas en el 2008 en la UCI. Actualmente se desempeña como trabajador del Departamento de Almacenes de Datos. DATEC. Facultad 6.

Correo: yiznaga@uci.cu

**Ing. Yordanka Echevarría Melo:** Graduada de Ingeniería en Ciencias Informáticas en el 2009 en la UCI. Actualmente se desempeña como trabajadora del Departamento de Almacenes de Datos. DATEC. Facultad 6.

Correo: yordankah@uci.cu

*Agradezco a todas las personas que me guiaron e impulsaron a continuar en este largo camino, y que no me dejaron vencer pese a todas las dificultades y adversidades que estuvieron presentes. A todas aquellas personas que cuando necesite de su ayuda no me la negaron, al contrario me ayudaron con creces.*

*Quiera agradecer muy especialmente a mi familia, por estar en los momentos en que más los necesite. A mi madre y mi padre, que no me dejaron flaquear en ningún momento, a mi hermano que me dio buenos consejos y a mi tutor, que gracias a su ayuda he llegado hasta donde estoy.*

*Dayli*

*Dedico este trabajo de diploma, a mi madre, mi padre y mi hermano que siempre me guiaron por un buen camino, para llegar ser una mejor persona y una gran profesional.*

*A mis abuelos que aunque no entendían mucho de lo que era mi tesis siempre me dieron mucho apoyo sentimental y me dieron muchos consejos que me ayudaron a no perder la fe en que sí podía lograrlo.*

*A mis verdaderos amigos que siempre aunque estén lejos se alegran de mis victorias.*

## RESUMEN

La Oficina Nacional de Estadística e Información (ONEI) encargada del control estadístico en Cuba presenta problemas con el almacenamiento y análisis de las series del Comercio Interno. Por lo que el presente trabajo de diploma enmarcó su desarrollo en los mercados de datos, donde se realizó el análisis de la información perteneciente a esta área. Como guía de este proceso se utilizó la metodología propuesta por el Centro de Tecnología de Gestión de Datos (DATEC). Proyectando como resultado las dimensiones, tablas, hechos y medidas necesaria para el buen desarrollo del mercado; se definieron e implementaron los procesos de extracción, transformación y carga (ETL) de los datos, desarrollándose además una capa de Inteligencia de Negocio (BI) procurando la visualización de los datos a través de los reportes candidatos, y finalmente se llevaron a cabo las pruebas pertinentes, dando paso a un mercado listo para la ayuda en la toma de decisiones.

**PALABRAS CLAVE:** Comercio Interno, BI, ETL, DATEC, ONEI.

**Tabla de contenidos**

<b>RESUMEN</b> .....	V
<b>INTRODUCCIÓN</b> .....	1
<b>CAPÍTULO I: Fundamentación teórica del mercado Series de comercio interno</b> .....	4
<b>1.1 Sistema de Información para Comercio Interno</b> .....	4
1.1.1 Sistema de Información para Comercio Interno en el mundo .....	4
1.1.2 Sistema de Información para Comercio Interno en Cuba .....	5
<b>1.2. Almacén de Datos</b> .....	5
1.2.1 Características de los almacenes de datos .....	6
1.2.2 Estructura de los almacenes de datos .....	6
1.2.3 Ventajas y Desventajas .....	7
<b>1.3. Mercados de Datos</b> .....	7
1.3.1 Características de los mercados de datos .....	8
1.3.2 Ventajas y desventajas de los mercados de datos.....	9
1.3.3 Metodologías para el desarrollo del mercado de datos.....	9
<b>1.4 Integración de Datos</b> .....	12
<b>1.5 Inteligencia de Negocio</b> .....	13
1.5.1 Modelo Multidimensional .....	14
<b>1.6 Modo de Almacenamiento de Datos OLAP</b> .....	16
1.6.1 Características de OLAP .....	16
<b>1.7 Herramientas del proceso de ingeniería</b> .....	18
1.7.1 Herramienta para el modelado .....	18
1.7.2 Sistema Gestor de Base de Datos.....	19
1.7.3 Herramienta de ETL .....	19
1.7.4 Herramientas de Inteligencia de Negocio .....	20
<b>Conclusiones del capítulo</b> .....	21
<b>CAPÍTULO 2: Análisis y Diseño del mercado Series de comercio interno</b> .....	22
<b>2.1 Análisis del mercado de datos Series de comercio interno</b> .....	22
2.1.1 Estudio preliminar.....	22
2.1.2 Tema de análisis .....	22
2.1.3 Necesidades de los usuarios.....	23
2.1.4 Requisitos de Información .....	23
2.1.5 Requisitos Funcionales .....	24
2.1.6 Requisitos no funcionales.....	25
2.1.7 Reglas del Negocio .....	25
2.1.8 Casos de Uso del Sistema (CUS).....	26



2.1.9 Descripción de los Casos de Uso .....	27
2.1.10 Perfilado de datos .....	31
<b>2.2 Diseño del Mercado de datos Series de comercio interno.....</b>	<b>31</b>
2.2.1 Dimensiones identificadas .....	32
2.2.2 Hechos identificados .....	33
2.2.3 Matriz Bus .....	34
2.2.4 Modelo de Datos .....	34
2.2.5 Diseño de los procesos de integración de datos.....	35
2.2.6 Mapa de navegación .....	36
2.2.7 Diseño de los cubos OLAP .....	37
2.2.8 Roles y privilegios en la base de datos.....	39
2.2.9 Políticas de respaldo y recuperación .....	40
<b>Conclusiones del capítulo .....</b>	<b>40</b>
<b>CAPÍTULO 3: Implementación y Prueba del mercado de datos Series de comercio interno .....</b>	<b>41</b>
<b>3.1 Implementación del subsistema de almacenamiento.....</b>	<b>41</b>
<b>3.2 Implementación del modelo físico .....</b>	<b>42</b>
<b>3.3 Implementación del subsistema de integración de datos (ETL).....</b>	<b>43</b>
3.3.1 Arquitectura de integración.....	43
3.3.2 Implementación de los trabajos .....	44
3.3.3 Implementación de las transformaciones.....	45
<b>3.4 Implementación del subsistema de visualización .....</b>	<b>46</b>
3.4.1 Implementación de los reportes candidatos.....	46
<b>3.5 Importancia de la calidad del software .....</b>	<b>48</b>
<b>3.6 Pruebas de software .....</b>	<b>49</b>
3.6.1 Modelo V .....	49
3.6.2 Herramientas de prueba .....	50
3.6.3 Resultado de Pruebas .....	54
<b>Conclusiones del capítulo .....</b>	<b>57</b>
<b>CONCLUSIONES GENERALES .....</b>	<b>58</b>
<b>RECOMENDACIONES.....</b>	<b>59</b>
<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>60</b>
<b>BIBLIOGRAFÍA.....</b>	<b>62</b>
<b>ANEXOS .....</b>	<b>65</b>
<b>GLOSARIO.....</b>	<b>68</b>



## Índice de figuras

Figura 1. Integración de Datos .....	13
Figura 2. Inteligencia de Negocio .....	14
Figura 3: Ejemplo de esquema estrella .....	15
Figura 4: Ejemplo de esquema snowflake o copo de nieve .....	15
Figura 5: Ejemplo de esquema constelación de hechos.....	16
Figura 6: Arquitectura del sistema ROLAP .....	18
Figura 7: Diagrama de Casos de Uso del Sistema.....	27
Figura 8: Modelo de Datos.....	35
Figura 9: Diseño de los procesos de integración de datos .....	35
Figura 10: Diseño de la transformación de la dimensión provincia .....	36
Figura 11: Mapa de Navegación .....	37
Figura 12: Cubo dinero recaudado grupo productos .....	38
Figura 13: Cubo circulación mercantil por unidades físicas .....	38
Figura 14: Cubo ventas.....	38
Figura 15: Cubo dinero recaudado por tipo establecimientos.....	39
Figura 16: Cubo diento recaudado por servicios comerciales prestados.....	39
Figura 17: Subsistema de almacenamiento .....	41
Figura 18: Arquitectura de integración. ....	44
Figura 19: Trabajo principal .....	44
Figura 20: Trabajo de las dimensiones .....	45
Figura 21: Trabajos de los hechos .....	45
Figura 22: Transformación de la dimensión indicador_productos.....	46
Figura 23: Transformación del hecho din_recaud_grupo_prod .....	46
Figura 24: Consulta MDX del reporte 14.2 Circulación mercantil mayorista total y a la red minorista en unidades físicas .....	47
Figura 25: Reporte 14.2 Circulación mercantil mayorista total y a la red minorista en unidades físicas	48
Figura 26: Modelo V.....	49
Figura 27: Resultado de las listas de chequeo.....	55

## Índice de Tablas

Tabla 1: Descripción de los actores y CUS .....	28
Tabla 2: Descripción del CU_ Presentar información de la circulación mercantil mayorista .....	30
Tabla 3: Matriz Bus .....	34
Tabla 4: Hechos y dimensiones del mercado.....	43
Tabla 5: Presentar información de la circulación mercantil mayorista .....	51
Tabla 6: Aplicación de las listas de chequeo.....	54

## INTRODUCCIÓN

En Cuba el Comercio Interno es una de las áreas más importantes para garantizar el sustento de la economía, debido a que este es un país subdesarrollado que presenta una economía planificada donde juega un papel importante el intercambio entre los diferentes organismos y entidades estatales. Toda la información relacionada con esta área es recogida por la Oficina Nacional de Estadísticas e Información (ONEI), organizándola en diversos modelos, de los cuales se extraen las series históricas, las que tienen una frecuencia de captación anual y son guardadas en formato "Excel". Esto provoca que su manejo y comprensión sea complejo debido a que no existe una forma adecuada de almacenar, recuperar y presentar los datos ya que no se encuentran integrados, las fuentes solo pueden ser consultadas por especialistas que tengan conocimientos de informática, por lo que se ve afectado el proceso de recuperación y elaboración de informes, así como, la disponibilidad de las series históricas, provocando retrasos en el proceso de toma de decisiones.

Por lo expuesto anteriormente, surge como **problema de la investigación** ¿Cómo contribuir a la toma de decisiones en el trabajo con series históricas en el área de Comercio Interno del Sistema de Información de Gobierno? Teniendo en cuenta el problema de investigación se define como **objetivo general** desarrollar un mercado de datos en el área de Comercio Interno para el trabajo con series históricas del Sistema de Información de Gobierno, que contribuya a la toma de decisiones. Definiendo de esta forma como **objetivos específicos** fundamentar la selección de la metodología, herramientas y tecnologías a utilizar en el desarrollo del mercado de datos Series de comercio interno, realizar el análisis y diseño del mercado de datos Series de comercio interno, implementar el mercado de datos Series de comercio interno y validar el mercado de datos Series de comercio interno. Se define además como **objeto de estudio** a los Almacenes de Datos y como **campo de acción** al Mercado de datos Series de comercio interno.

Las **tareas de la investigación** que se definen como:

- 1- Caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo del almacén de datos.
- 2- Selección de la metodología, herramientas y tecnologías a utilizar en el desarrollo del almacén de datos.
- 3- Levantamiento de requisitos, descripción de los casos de uso del mercado de datos, definición de los hechos, medidas y dimensiones del mercado de datos.
- 4- Diseño del modelo de datos.

- 5-Definición de la arquitectura del mercado de datos.
- 6-Diseño del subsistema de integración de datos históricos.
- 7-Diseño del subsistema de integración para el Sistema de Información de Gobierno (SIGOB).
- 8-Diseño del subsistema de visualización.
- 9-Diseño de los casos de pruebas.
- 10-Implementación del subsistema de integración de datos históricos.
- 11-Implementación del subsistema de integración para SIGOB
- 12-Implementación del subsistema de visualización
- 13-Aplicación de las listas de chequeo
- 14-Aplicación de los casos de prueba
- 15-Aceptación de la solución por el cliente.

Finalmente se definen como **resultados esperados** la obtención de un Mercado de datos para las Series históricas de Comercio interno del Sistema de Información de Gobierno que contribuya a la toma de decisiones en esta área.

## **Estructura del documento**

El presente documento de tesis estará organizado en introducción, tres capítulos, conclusiones, recomendaciones, referencias bibliográficas, bibliografía, anexos y glosario de términos.

### **Capítulo 1: Fundamentación Teórica del mercado de datos Series de comercio interno**

En este capítulo se definirán los principales conceptos asociados al campo de acción. Se describen las herramientas, metodologías y tecnologías a evaluar para formular la propuesta.

### **Capítulo 2: Análisis y diseño del mercado de datos Series de comercio interno**

Durante el desarrollo de este capítulo se realizará un estudio preliminar del negocio, se efectuará un levantamiento y análisis de los requisitos, los casos de uso, los hechos, las medidas y las dimensiones del mercado de datos que serán identificados durante el desarrollo de este. Una vez que se concluya la etapa inicial, de desarrollo, se construirá el modelo de datos correspondiente. Además, se diseñarán los subsistemas de integración, y visualización y finalmente se construirán los casos de prueba.

## **Capítulo 3: Implementación y pruebas del mercado de datos Series de comercio interno**

Se llevará a cabo la implementación de los procesos de ETL, la realización del modelo de datos, los reportes candidatos, los subsistemas de integración y de visualización. Además, se validará la solución propuesta, mediante la aplicación de diferentes pruebas y herramientas, como los casos de prueba y las listas de chequeo.

## **CAPÍTULO I: Fundamentación teórica del mercado Series de comercio interno**

### **Introducción**

En el presente capítulo se definirán conceptos, herramientas, metodologías y tecnologías que se utilizarán para el desarrollo del mercado de datos Series de comercio interno, fortaleciendo la fundamentación teórica que soporta la investigación. Además, se caracterizarán los procesos de integración de los datos e inteligencia de negocio, parte fundamental para el desarrollo del mismo.

### **1.1 Sistema de Información para Comercio Interno**

El Comercio Interno es el negocio que se realiza al vender, comprar o intercambiar productos para obtener beneficios fundamentales para cada país, se centra en la venta de alimentos y bienes de consumo de producción nacional, obedeciendo a la oferta y la demanda de los bienes y servicios, con el objetivo de cubrir las necesidades de la población. Intervienen en este proceso productos alimenticios, no alimenticios e industriales.

Para la realización de un sistema de información para el Comercio Interno es necesaria la utilización de datos estadísticos organizados por años, provincias, productos, per cápita, promedios anuales, promedios diarios de las ventas y tasas de crecimientos anuales, entre otros datos característicos de esta esfera.

Actualmente en el mundo existen muchos sistemas de información que son los encargados de agilizar el trabajo con datos del Comercio Interno. A continuación se mencionaran brevemente algunos de ellos.

#### **1.1.1. Sistema de Información para Comercio Interno en el mundo**

En el mundo existen muchas aplicaciones para el Comercio Interno debido a que estas se encuentran estrechamente relacionadas con las Tecnologías de la Información y las Comunicaciones(TIC), por lo que existen actualmente a nivel mundial, múltiples Sistemas de Información Automatizados(SIA) que sirven para el registro de las transacciones tanto diarias como de cualquier índole, generando así reportes que presentan información relevante, con un alto nivel de claridad, sencillez e importancia para las personas que trabajan con esta información. Ejemplo de estas tenemos:

- Sistemas ERP (Enterprise-resource planning)
- Sistemas de comercio electrónico (E-commerce systems)

- Aplicaciones de Ventas (General) (1)

## 1.1.2. Sistema de Información para Comercio Interno en Cuba

Anteriormente la Oficina Nacional de Estadísticas e Información (ONEI) conjuntamente con la Universidad de las Ciencias Informáticas (UCI) desarrolló el trabajo de diploma titulado “Sistema de Información de Gobierno. Mercado de datos Comercio Interior”, el cual tuvo como alcance el análisis y diseño del mercado de datos culminando con resultados satisfactorios. Por esa razón ya Cuba cuenta con un Sistema de Información Automatizado (SIA) que guarda información referente al Comercio Interno, pero este mercado no es suficiente debido a que no permite almacenar información histórica referente a esta área, lo que trajo como resultado, que la ONEI conjuntamente con la UCI, decidiera llevar a cabo la realización de un nuevo trabajo de diploma que permitiera almacenar toda la información histórica referente a esta área.

Para llevar a cabo la realización de este mercado es preciso tener conocimiento de diversos conceptos, herramientas y tecnologías de almacenamiento de datos, que sirvan como guía del proceso de construcción del mismo.

## 1.2. Almacén de Datos

Debido al gran avance de la Ciencia de la Información y las Comunicaciones (TIC), el mundo se ha visto inmerso en la búsqueda de cómo poder procesar, almacenar y manejar todo el gran cúmulo de información que diariamente va apareciendo en las organizaciones e instituciones. Por tal motivo, surge la necesidad de digitalizar toda la información para facilitar la manipulación de la misma.

Con el objetivo de dar solución a este problema, surge el término almacén de datos (en inglés Datawarehouse). El concepto de Almacén de Datos (AD) se puede concretar de múltiples formas. A continuación se darán varias definiciones propuestas por diversos autores.

Según William H. Inmon:

***“Un almacén es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza.” (2)***

Ralph Kimball define un almacén de datos como:



**“Una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis.” (2)**

Después de haber examinado las diferentes definiciones de almacén de datos dadas por los expertos se puede concluir que **“un almacén de datos es una colección de datos orientados a temas, integrados, no volátiles e histórico. Contiene información no solo de bases de datos relacionales, sino de otras fuentes relacionadas con la actividad de la empresa y cuya finalidad es la obtención de información estructurada y útil para la toma de decisiones.”**

### 1.2.1 Características de los almacenes de datos

Los almacenes de datos se identifican por sus características, las que lo definen como:

- **Temático:** la información se clasifica en base a los aspectos que son de interés para la empresa.
- **Integrado:** integra datos recogidos de diferentes sistemas operacionales de la organización.
- **No volátil:** el almacén de datos solo permite cargar nuevos datos y acceder a los ya almacenados; pero no permite ni borrar ni modificar los datos.
- **Histórico:** El tiempo debe estar presente en todos los registros contenidos en un almacén de datos. Las bases de datos operacionales contienen los valores actuales de los datos. Un almacén de datos no es más que una serie de instantáneas en el tiempo tomadas periódicamente. (3)

### 1.2.2 Estructura de los almacenes de datos

Para la construcción de los almacenes de datos hay que tener bien claro diferentes conceptos como:

- ✓ **Repositorio de datos:** los repositorios de datos operacionales son la fuente principal donde se encuentran los datos primarios, actuales e integrados, por tal razón son los encargados de suministrar los datos al sistema, estos datos operacionales pueden provenir de fuentes externas, estaciones de trabajo o servidores privados y principalmente de sistemas mainframe.
- ✓ **Gestor de carga:** también conocido como Sistema ETL (Extracción, Transformación y Carga), el cual es encargado de realizar las funciones como indican su nombre de extracción de las fuentes de datos, transformación y limpieza de los datos y finalmente la carga al almacén de datos, refrescando de esta forma el almacén.
- ✓ **Gestor del almacén de datos:** encargado de la gestión de los datos dentro del almacén con la utilización de herramientas específicas para la transformación de los datos, creación de copias

de seguridad y archivado de datos, y así mismo se va realizando el análisis de los datos para mantener la coherencia de los mismos.

- ✓ **Datos muy detallados:** son los que se obtienen directamente del procesado de los datos.
- ✓ **Datos pocos resumidos y muy resumidos:** contienen los datos que el gestor del almacén de datos ha establecido como resúmenes predefinidos.
- ✓ **Datos de archivos/ copia de seguridad:** área en la cual se guardan los datos detallados y resumidos, con el fin de mantener un archivo origen y sus respectivas copias de seguridad.
- ✓ **Metadatos:** describen la estructura de los datos dentro del almacén y pueden ser utilizados por los gestores de carga del almacén y de consultas.
- ✓ **Sistemas gestores de bases de datos (SGBD):** el principal problema al utilizar los almacenes de datos es su gran tamaño, por eso se requiere de una buena elección de un SGBD.
- ✓ **Gestor de consultas:** encargado de gestionar las operaciones asociadas a las consultas.
- ✓ **Herramientas de acceso para usuarios:** el objetivo es proporcionar una herramienta de acceso para los usuarios en la cual se puedan tomar variantes que ayuden a una eficiente toma de decisiones. **(3).**

### 1.2.3 Ventajas y Desventajas

Para una correcta evaluación de la ejecución del almacén de datos hay que considerar las ventajas y desventajas que aporta la implantación del mismo.

#### Ventajas

- Rentabilidad de las inversiones realizadas para su creación.
- Aumento de la productividad de los técnicos de dirección
- Aumento de la competitividad en el mercado.
- Privacidad de los datos.**(4)**

#### Desventajas

- Infravaloración del esfuerzo necesario para su diseño y creación.
- Infravaloración de los recursos necesarios para la captura, carga y almacenamiento de los datos.
- Incremento continuo de los requisitos de los usuarios.**(4)**

### 1.3. Mercados de Datos

Es una Base de Datos (BD) departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer de la estructura óptima de los datos para analizar la información detalladamente desde todas las formas en que afecte los procesos del

departamento. Puede ser alimentado desde los datos de un AD, o integrar por sí mismo uno, comprendido por distintas fuentes de información **(5)**.

Para un mejor entendimiento de este concepto es necesario recurrir al análisis de las características de los mercados de datos.

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

Algunas de las características más significativas y que mejor se ajustan a la justificación del por qué, es importante la utilización de los mercados de datos se expondrán a continuación:

- Poseen las mismas características de integración, no volatilidad, orientación temática e historicidad que un almacén de datos.
- Representan una estrategia de divide y vencerás para ámbitos muy genéricos de un almacén de datos.
- Los mercados de datos se enfocan a los requisitos de los usuarios que están asociados a un departamento específico de la empresa.
- Su utilización y comprensión es sencilla, debido a que contienen menor cantidad de información que los almacenes de datos. **(6)**

Una buena comprensión de los mercados de datos está dada a partir del conocimiento de sus ventajas y desventajas.

### 1.3.2. Ventajas y desventajas de los mercados de datos

Alguna de las ventajas y desventajas que poseen los mercados de datos son:

#### Ventajas

- Son simples de implementar.
- Poco tiempo de construcción y puesta en marcha.
- Permiten manejar información confidencial.
- Reflejan rápidamente sus beneficios y cualidades.
- Reducen la demanda del almacenamiento de datos. **(6)**

#### Desventajas

- Se pierde capacidad de procesamiento debido al crecimiento de los datos.
- Los usuarios necesitan acceder a varios MD.
- Dificultad para construir debido al corto plazo de desarrollo. **(6)**

Otro de los aspectos fundamentales a tener en cuenta viene dado por la metodología que se va a utilizar para la implantación de un correcto mercado de datos. Por tal motivo es necesario tener en cuenta las definiciones dadas por varias personalidades que son autoridades en este tema.

### 1.3.3. Metodologías para el desarrollo del mercado de datos

***Una metodología es una guía que se sigue a fin de realizar las acciones propias de una investigación, es decir, se trata de una guía que va indicando qué hacer y cómo actuar cuando se quiere obtener algún tipo de resultado bien desarrollado. También se puede definir como un enfoque que permite observar un problema de una forma total, sistemática y disciplinada, todo en dependencia de la problemática presentada (7).***

Varios especialistas reconocidos a nivel mundial plantean un enfoque muy personal hacia una metodología más idónea, de acuerdo a las perspectivas de desarrollo de cada almacén o mercado. Por tal razón se tomarán en cuenta dos conceptos fundamentales que se contraponen en sus definiciones a la hora de construir el almacén o el mercado.

De acuerdo con Bill Inmon:

***“Un mercado de datos es una estructura departamental que alimenta un almacén en el cual los datos no están normalizados y se basa en las necesidades de información del departamento.”***  
**(8)**

Derivado de este concepto, Inmon, propone una metodología con un enfoque top-down (descendente) la cual plantea que el almacén de datos debe responder a las necesidades de todos los usuarios en la organización, y no sólo de un determinado grupo. Por otra parte plantea que el almacén de datos debe ser construido primero que los mercados.

Englobando de esta forma, una metodología según la cual, se debe ir desde una visión más general de las distintas partes que componen el almacén y posteriormente ir concretando y refinando cada una de las partes por separado.

Sin embargo, esta definición de metodología, difiere con lo planteado por Ralph Kimball, el cual plantea que:

***“Un mercado de datos es un subconjunto lógico de un almacén de datos completo. Frecuentemente es visto como una restricción de un almacén a un grupo específico del negocio. Está construido, orientado y organizado alrededor de una parte específica del negocio.”*** (9)

Basado en esta definición, Kimball propone un enfoque bottom-up (ascendente), el cual se orienta hacia una metodología rápida que se basa en experimentos y prototipos que da como resultado un método flexible que permite a la organización ir más lejos con menores costos.

La idea es, construir mercados de datos independientes para evaluar las ventajas del nuevo sistema a medida que avanzamos. En él, las partes individuales se diseñan con detalle y luego se enlazan para formar componentes más grandes, que a su vez se enlazan hasta que se forma el sistema completo.

Sin embargo, podría haber problemas tratando de integrar los mercados de datos con el almacén, ya que la primera iteración de definición de datos y las siguientes puede que no sean compatibles. Aún así, es muy práctica si se considera que un almacén de datos tiene que ser entendido fácilmente por los usuarios y ofrecer respuestas correctas a la mayor brevedad posible.

Este enfoque parte de los requisitos de negocio, mientras que el enfoque top-down propone la validación de los requisitos una vez que se tiene el sistema. Además vale destacar como aspectos positivos de esta metodología que:

- Da lugar a los conceptos de hechos y dimensiones, lo cual contribuye al proceso de toma de decisiones, así como también al proceso de desarrollo.
- Plantea que el almacén de datos se vaya construyendo a medida que se construyan las bases de datos departamentales, lo que da una sensación de organización en la empresa o entidad.
- La documentación existente sobre la misma es abundante, lo que brinda la posibilidad de aclarar todas las dudas relacionadas al tema.
- Es reconocida por todos los que abordan esta disciplina, es una metodología madura y que tiene bien definidas las etapas, actividades, artefactos y roles.

Por tal razón, después de haber analizado los aspectos de cada metodología propuestos por los especialistas, se propone que en la Universidad de las Ciencias Informáticas (UCI) se utilice una adaptación de la metodología Kimball. Esta metodología adaptada conduce a una solución completa en un corto período de tiempo, se tomaron las cuatro fases fundamentales de la metodología de Kimball: análisis, diseño, ETL y BI. Otro aspecto importante es que se adaptó a las características del centro, con la incorporación de un Expediente de Proyecto que cumple con todos los lineamientos establecidos por el Centro de Calidad para Aplicaciones Tecnológicas (CALISOF), incorporándole un grupo de artefactos nuevos, típicos de este tipo de soluciones. También se puede señalar que esta metodología se ajusta a las necesidades del modo de producción establecido por DATEC debido a que organiza los roles y las actividades en cinco grupos de trabajos (Dirección, Análisis, Arquitectura y Base de Datos, ETL y Visualización de datos) estableciendo un balance entre cada grupo y las etapas de desarrollo, de manera que se pudieran reutilizar los recursos humanos en los diferentes proyectos desarrollados por el departamento, además esta metodología cuenta con las fases de desarrollo del ciclo de vida de un almacén de datos, explicadas a continuación:

**Estudio preliminar y planeación:** se realiza un estudio del negocio con el fin de determinar qué se desea construir y qué condiciones existen para su desarrollo y montaje. También en esta etapa se lleva a cabo la planeación del proyecto, se definen los objetivos, el alcance preliminar, los costos estimados y los recursos necesarios.

**Requerimientos:** se identifican las necesidades de información, las reglas de negocio, se realiza el levantamiento de los requerimientos a partir de las necesidades de los usuarios y se identifican las reglas del negocio con los elementos disponibles en las fuentes.

**Arquitectura y Diseño:** se definen las estructuras de almacenamiento, se diseñan las reglas de extracción, transformación y carga, así como la arquitectura de información que regirá el desarrollo de la solución.

**Implementación:** se lleva a cabo el diseño físico de la solución, se crean las estructuras de almacenamiento, se ejecutan las reglas de ETL y se configuran e implementan las herramientas de BI.

**Prueba:** se realizan las pruebas de unidad, luego las pruebas de integración y sistema y las pruebas de aceptación con el cliente final.

**Despliegue:** se configuran los servidores necesarios y se instalan las herramientas según la arquitectura definida, se carga una muestra de los datos en un ambiente controlado, con el fin de demostrarle al cliente final que la solución funciona y una vez aceptada por el cliente, se realiza la carga histórica de los datos.

**Soporte y Mantenimiento:** son los servicios que se brindan luego de implantar la solución, pueden ser desde un soporte en línea, vía telefónica, web, correo u otras, hasta el acompañamiento junto al cliente.

**Gestión y Administración del Proyecto:** transcurre a lo largo de todo el ciclo de vida, es aquí donde se controla, gestiona y chequea todo el desarrollo, los gastos, las utilidades, los recursos, las adquisiciones, y demás actividades relacionadas con la gestión de proyecto **(10)**.

Igualmente, como complemento de la metodología Kimball y fortaleciendo la etapa del levantamiento de requisitos, se tomó lo planteado por Leopoldo Zenaido Zepeda Sánchez en su Tesis de Doctorado, orientando el trabajo hacia los casos de uso guiando el proceso de desarrollo.

## 1.4 Integración de Datos

En la actualidad al hablar de mercados de datos es común escuchar términos como “*integración de datos*” y esto se debe a que este proceso es el que se realiza para lograr la correcta incorporación de los datos que se encuentran en las fuentes, al almacén.

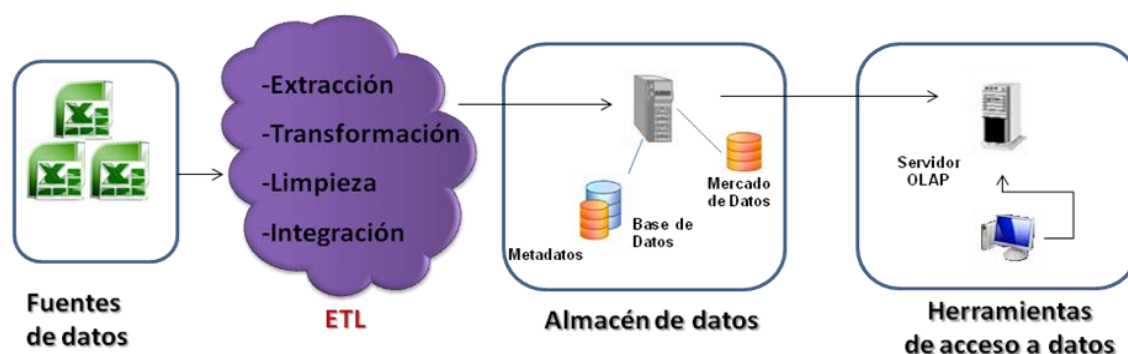


Figura 1: Integración de Datos.

El proceso de “*integración de datos*” es uno de los procesos que más tiempo de trabajo ocupa cuando se quiere acoplar correctamente los datos en el mercado. Para la realización correcta de este proceso es necesaria la utilización de los procesos de Extracción, Transformación y Carga (ETL), dando como resultado que los datos que se encontraban en las fuentes se integren y carguen de forma eficiente en los mercados, logrando como resultado un repositorio listo para el apoyo a la toma de decisiones.

### 1.5 Inteligencia de Negocio

En el mundo actual existen múltiples empresas que poseen grandes volúmenes de información, que necesitan una adecuada forma de manejo y control; pero carecen de métodos y herramientas que permitan el tratamiento apropiado de la misma. Por lo que la inteligencia de negocio puede ser la clave del problema, ya que entre las vías que ofrece para la solución de estos inconvenientes se encuentran las herramientas que ayudan a la visualización de los datos, por medio de gráficos y reportes entre las que se pueden hallar las de Procesamiento Analítico en Línea (OLAP en inglés) (Figura2), a través de las cuales se puede brindar a los usuarios mejores posibilidades para trabajar con los datos y optimizar el proceso de toma de decisiones en las empresas o instituciones, en aras de cumplir con las necesidades de los clientes.



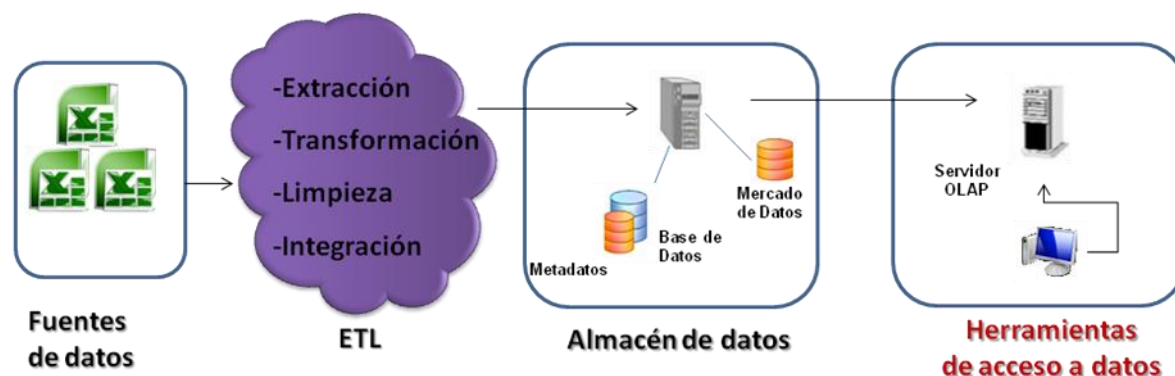


Figura 2: Inteligencia de Negocio.

### 1.5.1. Modelo Multidimensional

El modelo multidimensional almacena la misma información que el Modelo Entidad Relación (MER), lo que este la organiza de una forma diferente con el objetivo de garantizar la velocidad y recuperación de la misma. Esta es una técnica para modelar bases de datos simples y entendibles al usuario final, donde su idea fundamental es que el usuario visualice fácilmente la relación que existe entre los distintos componentes del modelo.

Para un mejor entendimiento de lo que es un modelo multidimensional se hace necesario comentar las posibles variantes existentes y finalmente definir la que se utilizará.

#### 1.5.1.1 Implementación del Modelo Multidimensional

La aplicación de modelos multidimensionales son necesarios para obtener un mejor entendimiento de los datos que almacena, para demostrar cuan explícitos pueden ser estos modelos se explicara algunos de los más usados.

#### Esquema estrella

Este esquema está formado por un modelo asimétrico con una tabla de hechos dominante en el centro del esquema, el cual se encarga de conectar las tablas de dimensiones.

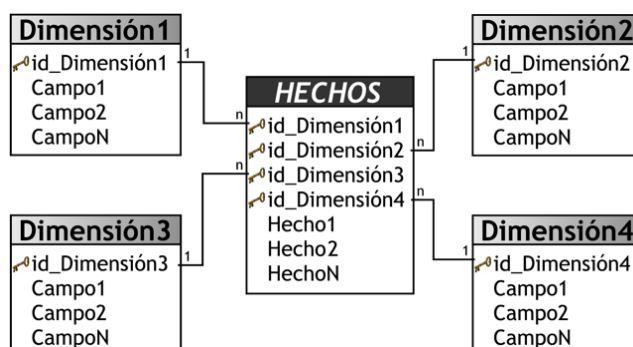


Figura 3: Ejemplo de esquema estrella.

### Esquema snowflake o copo de nieve

Es una variante del esquema de estrella en el que las tablas dimensionales se organizan jerárquicamente mediante su normalización, donde existe una tabla de hechos central que está relacionada con una o más tablas de dimensiones, quienes a su vez pueden estar relacionadas o no con una o más tablas de dimensiones.

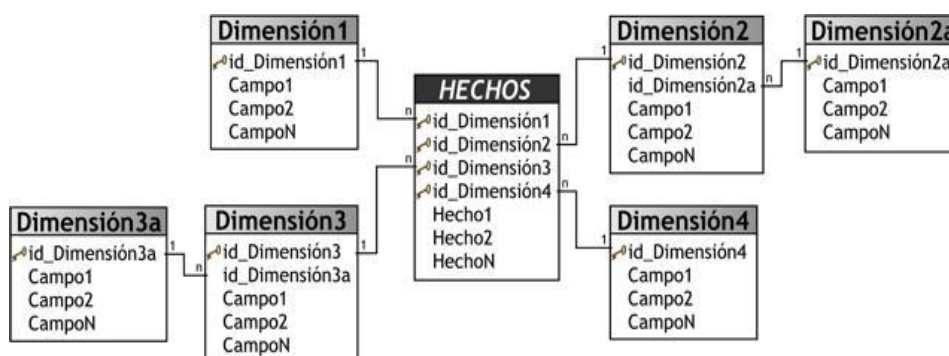


Figura 4: Ejemplo de esquema snowflake o copo de nieve.

### Constelación de hechos

La constelación de hechos es un conjunto de tablas de hechos que comparten algunas tablas de dimensiones, además está compuesto por una serie de esquemas de estrella, es decir, una tabla de hechos central con otras auxiliares y sus respectivas tablas de dimensiones.

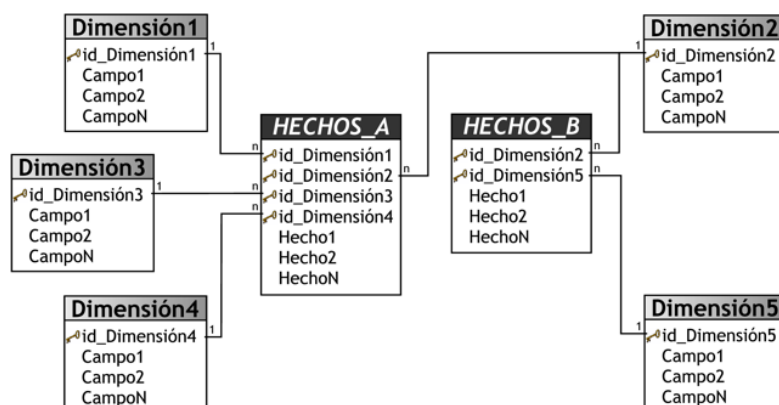


Figura 5: Ejemplo de esquema constelación de hechos.

Después de haber realizado el análisis de los diferentes esquemas, se selecciona el de constelación de hechos como el esquema a utilizar en este mercado, debido a que es un modelo donde se pueden identificar claramente sus componentes y se ve considerablemente toda la información con que se está trabajando.

Para lograr un adecuado desarrollo del mercado de datos se hace necesario conocer qué modo de almacenamiento de datos se va a utilizar.

## 1.6 Modo de Almacenamiento de Datos OLAP

Se decidió utilizar la tecnología de Procesamiento Analítico en Línea (OLAP) debido a que permite clasificar los datos en diferentes dimensiones y pueden ser vistos unos con otros en diferentes combinaciones, para posteriormente realizar el análisis de los datos que contienen.

-En estos modelos los datos son vistos como cubos los cuales consisten en categorías descriptivas (dimensiones) y valores cuantitativos (medidas).

-Estos modelos permiten formular consultas complejas, arreglar datos de un reporte, cambiar los datos resumidos o detallados etc.

Es bueno conocer también las características OLAP para una correcta interpretación de la tecnología a utilizar.

### 1.6.1 Características de OLAP

OLAP tiene muchas características que lo identifican como la herramienta de procesamiento analítico más usada y mejor para clasificar los datos como por ejemplo:

**Rápido:** proporciona la información al usuario a una velocidad constante (mayormente en cinco segundos o menos).

**Análisis:** realiza análisis estadísticos y numéricos básicos de los datos.

**Compartida:** permite compartir los datos potencialmente confidenciales a través de una gran cantidad de usuarios, implementando para esto los requerimientos de seguridad necesarios.

**Multidimensional:** permite ver la información en determinadas vistas o dimensiones.

**Información:** accede a todos los datos necesarios, donde quiera que estos residan, y mientras no esté limitada por el volumen. **(11)**

Además de las características que posee OLAP, vale aclarar que existen tres modelos de información, el **Procesamiento Analítico Relacional en Línea (ROLAP, por sus siglas en inglés)**, que es el que accede a los datos almacenados en un AD para proporcionar los análisis OLAP, este motor se integra con niveles de presentación, a través de los cuales los usuarios realizan los análisis OLAP **(12)**.

También se encuentra el **Procesamiento Analítico Multidimensional en Línea (MOLAP, por sus siglas en inglés)** utiliza una arquitectura de dos niveles: las BD multidimensionales y el motor analítico. La BD multidimensional es la encargada del manejo, acceso y obtención del dato. **(12)**

Otra de las estructuras de almacenamiento es el **Procesamiento Analítico en Línea Híbrido (HOLAP, por sus siglas en inglés)**. En la cual el proceso de análisis se realiza igual, lo que varía es la metodología de almacenamiento, influyendo en la velocidad de recuperación de la información, las zonas de ubicación y el procesamiento de los datos en general. **(12)**

Después de haber realizado un análisis de las estructura de almacenamiento, se decidió usar la arquitectura ROLAP, ya que almacena los datos en una BD relacional, lo que implica que no es necesario que los datos se repliquen en un almacenamiento separado para el análisis. Los cálculos se realizan en una BD relacional, con grandes volúmenes de datos y tiempos de navegación no predecibles. **(12)**

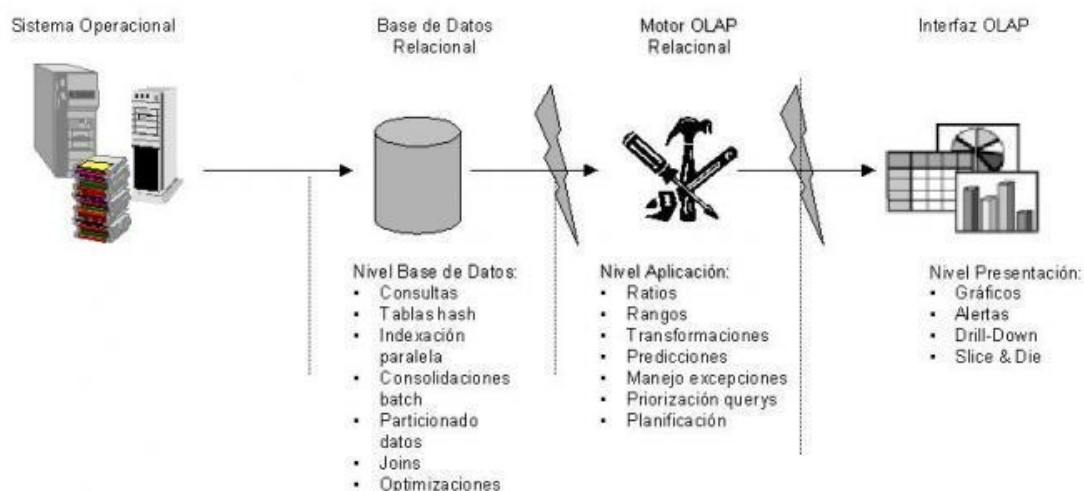


Figura 6: Arquitectura del sistema ROLAP.

## 1.7 Herramientas del proceso de ingeniería

Las herramientas que se utilizan para el proceso de ingeniería dan la posibilidad de realizar un diseño y una mejor construcción de la BD del nivel lógico a un nivel físico. También estas son las que se encargan de plasmar los requerimientos del sistema y la funcionalidad.

### 1.7.1 Herramienta para el modelado

Por ejemplo la herramienta de modelado propuesta por DATEC es **Visual Paradigm 6.4**, porque es una herramienta de diseño que soporta todos los diagramas UML, esquemas y diagramas de entidad-relación y la universidad paga por su licencia debido a su uso y popularidad.

Es una herramienta UML profesional que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. El software de modelado ayuda a una más rápida construcción de aplicaciones de calidad y a un menor coste. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. También proporciona tutoriales y proyectos UML **(13)**

Esta herramienta acelera el desarrollo de aplicaciones, sirve de puente visual entre arquitectos, analistas y diseñadores de sistemas de información, haciendo el trabajo más fácil y dinámico.

### 1.7.2 Sistema Gestor de Base de Datos

Otra de las herramientas más utilizadas son las relacionadas con el trabajo en las bases de datos, como por ejemplo:

**PostgreSQL 9.1** ampliamente considerado como una de las alternativas de sistema de bases de datos de código abierto, donde su código fuente está disponible liberado bajo licencia Berkeley Software Distribución (BSD). Brinda ahorros considerables en costos de operación ya que ha sido diseñado y creado para tener un mantenimiento y ajuste mucho menor que otros productos, conservando todas las características, estabilidad y rendimiento. **(14)**

**PgadminIII 1.14.1** herramienta de propósito general para diseñar, mantener, y administrar las bases de datos de PostgreSQL. Está diseñado para responder a las necesidades de todos los usuarios, desde escribir simples consultas SQL (Lenguaje de Consultas Estructuradas, por sus siglas en inglés Structured Query Language) hasta el diseño de bases de datos complejas. **(14)**

### 1.7.3 Herramienta de ETL

Una de las herramientas más útiles para trabajar con los procesos de ETL es el **Pentaho Data Integration 4.2.1** conocido como Kettle. Es una de las herramientas más utilizadas en el proceso de ETL, de código abierto que reúne un conjunto de componentes que permiten modelar y ejecutar transformaciones sobre flujos de datos. Provee un JDBC que permite la conexión con cualquier BD sin tener que instalar un cliente adicional, su rendimiento puede verse afectado a la hora de realizar operaciones de trabajos con numerosos volúmenes de datos, pues maneja pequeñas cantidades de información en el flujo. Se le pueden incorporar tareas como copiar, eliminar, descompactar y transportar ficheros usando FTP. **(15)**

Para trabajar con el perfilado y calidad de los datos entre otras acciones de limpieza de datos la herramienta más indicada es el **Data Cleaner 1.5.4**. Esta herramienta es una aplicación de código abierto para el perfilado, validación y comparación de los datos, es fácil de utilizar, genera sofisticados informes y gráficos que le permite a los usuarios comprobar la calidad de los datos existentes, almacena los datos en los AD, la investigación estadística, la preparación para el proceso de ETL y otras actividades. **(16)**

#### 1.7.4 Herramientas de Inteligencia de Negocio

La herramienta más idónea para el trabajo con esquemas es el **Pentaho Schema Workbench 3.2.0**. Esta herramienta se utiliza para el desarrollo del esquema del modelo estrella en XML desarrollada en Java. Este programa recientemente publicado en el año 2007, entrega todas las facilidades para poder realizar el modelo lógico del cubo OLAP al cual se le realizarán las consultas. Este se conecta directamente con la BD para así poder diseñar los cubos OLAP que se requieren para que el usuario final pueda visualizar los indicadores. **(17)**

En la búsqueda de propiciar una fuerte plataforma que soporte la infraestructura necesaria para implantar la solución deseada, la herramienta más conveniente a utilizar es **Pentaho BI Server 3.8**. Esta es una plataforma que provee el soporte y la infraestructura necesarios para crear soluciones de inteligencia empresarial a problemas de negocios. El marco proporciona los servicios básicos, incluidos autenticación, registro, auditoría, servicios web y motor de reglas. Incluye un motor de solución que integra reportes, análisis, tableros de comandos y componentes de minería de datos. Funciona como un sistema basado en administración web de informes, el servidor de integración de aplicaciones y un motor de flujo de trabajo ligero (secuencias de acción.) Está diseñado para integrarse fácilmente en cualquier proceso de negocio. **(18)**

Para el soporte de las consultas OLAP es necesaria la utilización de **Mondrian OLAP Server 3.0.4**. Permite realizar consultas al AD y que los resultados sean presentados mediante un navegador de modo que el usuario pueda realizar actividades típicas de navegación. Comprende 4 capas vistas desde el nivel del usuario final: la capa de presentación, la capa dimensional, la capa de estrella y la de almacenamiento. Dentro de la capa de presentación se puede encontrar variadas formas de presentar los datos, como gráficas de pastel, barras o líneas, y la exportación de las mismas. La dimensional analiza gramaticalmente, valida y ejecuta sentencias MDX. La capa de estrella es la responsable de mantener en cache agregado y la de almacenamiento se encuentra representada por un SGBD. Se pueden encontrar en el mismo servidor o en varios aunque la capa dimensional y la de estrella si deben estar en el mismo servidor. **(19)**

Finalmente la herramienta que nos permite realizar acciones para gestionar las soluciones impuestas es la **Apache Tomcat 6.6**. Esta herramienta es un servidor web y de aplicaciones que gestiona solicitudes y respuestas Http y https (incluye el servidor Apache) gracias a sus conectores Http. Es un servidor de aplicaciones o contenedor de Servlets/JSP. Es de código abierto, implementado con tecnología Java bajo la licencia de Apache 2 que le permite funcionar en cualquier sistema operativo que se encuentre la máquina virtual de Java permitiéndoles a los usuarios realizar reportes. **(20)**

## Conclusiones del capítulo

Después de haber realizado el análisis de los fundamentos teóricos del mercado de datos, se llegó a las siguientes conclusiones:

- Se abordó de manera general el proceso de desarrollo de los mercados de datos, con sus respectivas definiciones dadas por los expertos en el tema, constatándose que la implantación de un mercado facilita en muchas ocasiones el trabajo y la organización de la información que es manejada por las empresas.
- Se decidió adoptar la Metodología para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio propuesta por DATEC, la cual es basada en la Metodología de Ralph Kimball, debido a las grandes ventajas que esta proporciona a la hora de desarrollar la solución.
- Para el desarrollo de la solución, se hizo necesario la utilización de varias herramientas de las cuales se expusieron diversos argumentos que las identifican como las más adecuadas, para una correcta y óptima solución, teniendo un peso fundamental el hecho de que todas, excepto Visual Paradigm son herramientas de código abierto, siguiendo la línea de utilización de herramientas de software libre de la UCI.



### **CAPÍTULO 2: Análisis y Diseño del mercado Series de comercio interno**

#### **Introducción**

En este capítulo se realizará un estudio preliminar del negocio, la necesidades de información, las reglas del negocio, descripción de los Casos de Uso del Sistema(CUS), identificación de dimensiones, hechos y medidas, desarrollo de la matriz Bus, modelo de datos, arquitectura de información, así como se definirán los reportes candidatos y el diseño de los procesos de integración.

#### **2.1 Análisis del mercado de datos Series de comercio interno**

##### **2.1.1 Estudio preliminar**

La Oficina Nacional de Estadísticas (ONEI) es la entidad encargada de, organizar, ejecutar, y aplicar las políticas estatales que guían la estadística del país. Todo este proceso es posible mediante la intervención del Sistema de Información Estadístico Nacional (SIEN), que sirve como directriz en cuanto a dirección, ejecución y control de la captación de cifras económicas y sociales, así como su difusión de acuerdo con los requerimientos de la economía y necesidades del país. Esta entidad tiene como misión construir un sistema estadístico de información automatizado, capaz de responder con rapidez y calidad, a las necesidades de información del país, para enfrentar los problemas que diariamente se presentan al manejar la información referente a esta área del Comercio Interno.

El Sistema de Información de Gobierno (SIGOB), nace con la necesidad de organizar y centralizar toda la información existente en la ONEI logrando de esta forma un mejor manejo y control de los datos que en esta se almacenan. Por tal motivo se hace necesaria la creación de una herramienta que permita acceder a toda la información, con el objetivo de apoyar la toma de decisiones en las diferentes áreas de esta entidad y específicamente en el área de Comercio Interno, en la cual se gestionan todas las estadísticas relacionadas con el movimiento de bienes y servicios, para satisfacer todas las necesidades de la población a nivel nacional.

##### **2.1.2 Tema de análisis**

Para el desarrollo de cualquier mercado de dato es necesario identificar el tema de análisis, debido a que este es la forma en que se divide o clasifica la información referente a una determinada organización, tomando en cuenta los distintos temas que se escogen en correspondencia con los objetivos que se quieren desarrollar de acuerdo a las características del área en donde se encuentre el mercado.

En la presente investigación se desarrollará como tema de análisis el área del Comercio Interno en el trabajo con las series históricas, donde se tratará toda la información referente a los indicadores relacionados con los productos alimenticios, no alimenticios e industriales, per cápita, variación anual, ventas promedio diarias, participación en la circulación mercantil minorista, tasa media de crecimiento anual y totales generales o de las redes minoristas. Toda la información manipulada está comprendida entre los años 1989 y 2010.

### **2.1.3 Necesidades de los usuarios**

Para un buen desarrollo del análisis en el proceso de negocio es preciso conocer las necesidades que presentan los usuarios de esta área. La implicación de los mismos durante el desarrollo del mercado es de gran importancia debido a que de ellos depende que los resultados sean o no los esperados.

### **2.1.4 Requisitos de Información**

Estos son la base fundamental sobre la cual se va a trabajar, debido a que constituyen las principales funcionalidades que el sistema debe tener disponible a la hora de realizar el análisis de los datos. Además representan la principal fuente para el proceso de inteligencia de negocio y para la futura salida de los reportes adecuándolos a las necesidades del cliente.

A partir de la comparación entre las reglas de negocio y las necesidades de los usuarios fueron identificados 10 requisitos de información, expuestos a continuación.

**RI 1-** Obtener la cantidad de dinero recaudado por la circulación mercantil mayorista total y a la red minorista por grupo de productos y año.

**RI 2-** Obtener la circulación mercantil mayorista total y a la red minorista en unidades físicas por año y productos.

**RI 3-** Obtener la circulación mercantil minorista total por año.

**RI 4-** Obtener la cantidad de dinero recaudado por la circulación mercantil minorista por año, grupo y provincia.

**RI 5 -** Obtener las ventas en el comercio minorista por año.

**RI 6-** Obtener las ventas en la alimentación pública por año.

**RI 7-** Obtener los ingresos totales de las ventas en el comercio minorista por tipos de establecimientos, año y provincia.

**RI 8-** Obtener los ingresos de las ventas por conceptos en la alimentación pública por año y provincia.

**RI 9-** Obtener la cantidad de servicios comerciales prestados a la población por año y productos.

**RI 10-** Obtener la cantidad de dinero recaudado por los servicios comerciales prestados a la población, por año y productos.

### **2.1.5 Requisitos Funcionales**

Los requisitos funcionales son los que definen las funciones que se van a desarrollar en el sistema, producto a que van orientado hacia las necesidades de los usuarios finales. En la presente investigación se identificaron 25 requisitos funcionales, mostrados a continuación:

**RF 1-** Autenticar usuario.

**RF 2 -** Adicionar roles.

**RF 3 -** Eliminar roles.

**RF 4 -** Adicionar usuarios.

**RF 5 -** Eliminar usuarios.

**RF 6 -** Insertar reportes.

**RF 7 -** Modificar reportes.

**RF 8 -** Eliminar reportes.

**RF 9 -** Extraer datos.

**RF 10 -** Realizar transformación y carga.

**RF 11 -** Abrir navegador OLAP.

**RF 12 -** Mostrar editor MDX.

**RF 13 -** Mostrar Padres.

**RF 14 -** Ocultar repeticiones.

**RF 15** - Intercambiar ejes.

**RF 16** - Mostrar gráfico.

**RF 17** - Configurar gráfico.

**RF 18** - Configurar impresión.

**RF 19** - Exportar a PDF.

**RF 20** - Exportar a Excel.

**RF 21** - Mostrar propiedades.

**RF 22** - Suprimir filas.

**RF 23** - Detallar miembros.

**RF 24** - Entrar en detalles.

**RF 25** - Mostrar datos de origen.

### **2.1.6 Requisitos no funcionales**

Los requisitos no funcionales son propiedades o cualidades que el producto debe cumplir. En la presente investigación se identificaron 10 requisitos no funcionales, relacionados con la usabilidad, confiabilidad, eficiencia, soporte, restricciones de diseño, requisitos para la documentación de usuarios y ayuda del sistema, componentes comprados, interfaz, requisitos de licencia y requisitos legales, de derecho de autor y otros(para consulta de los mismos ver el artefacto Especificación de Requisitos, adjunto al expediente de proyecto).

### **2.1.7 Reglas del Negocio**

Las reglas del negocio muestran una serie de políticas y normas que se deben seguir, estas son operaciones, definiciones y restricciones presentes en la institución y que son de vital importancia para alcanzar sus objetivos.

En la investigación se definieron seis reglas del negocio:

**RN 1** - En el mercado de datos en las medidas `total_dinero_red_minorista` y `total_dinero_recaudado_en_valor` pueden existir valores nulos, porque en algunos casos la información no se encontraba actualizada desde el año 2002 hasta el 2006 y desde el año 2009 hasta el 2010.

**RN 2** - En el mercado de datos en las medidas `total_dinero_en_unidades_fisicas` y `total_dinero_red_minorista` pueden existir valores nulos, porque en algunos casos la información no se encontraba actualizada desde el año 2008 hasta el 2010.

**RN 3** - En el mercado de datos en las medidas `dinero_recaudado_circMerc_total`, `alimentación_publica` y `comercio_minorista` pueden existir valores nulos, porque en algunos casos la información no se encontraba actualizada en el año 1985.

**RN 4** - En el mercado de datos en la medida `dinero_recaudado_por_tipo_circulacion` pueden existir valores nulos, porque en algunos casos la información no se encontraba actualizada, desde el año 2007 hasta el 2010.

**RN 5** - En el mercado de datos en la medida `cant_servicios_comerciales_prestados_poblacion`, `cant_dinero_recaudado_servicios_comerciales_prestados_poblacion` pueden existir valores nulos, porque en algunos casos la información no se encontraba actualizada, desde el año 1992 hasta el 2009.

**RN 6** - El código de los atributos en cada una de las dimensiones no pueden tomar valores repetidos.

En el artefacto Reglas de negocio y transformación, se realizó una explicación detalla de las mismas.

### **2.1.8 Casos de Uso del Sistema (CUS)**

El diagrama de Casos de Uso del Sistema(CUS) es la representación de la relación entre los CUS y los actores del negocio, donde los CUS son representados generalmente por los requisitos funcionales y los actores del sistema, son los usuarios o programas que interactúan con este. Los casos que se desarrollarán son la representación visual de la información.

Para una mejor comprensión del diagrama se realizará una representación que contiene los CUS y las relaciones que se establecen con los usuarios que interactúan con él.

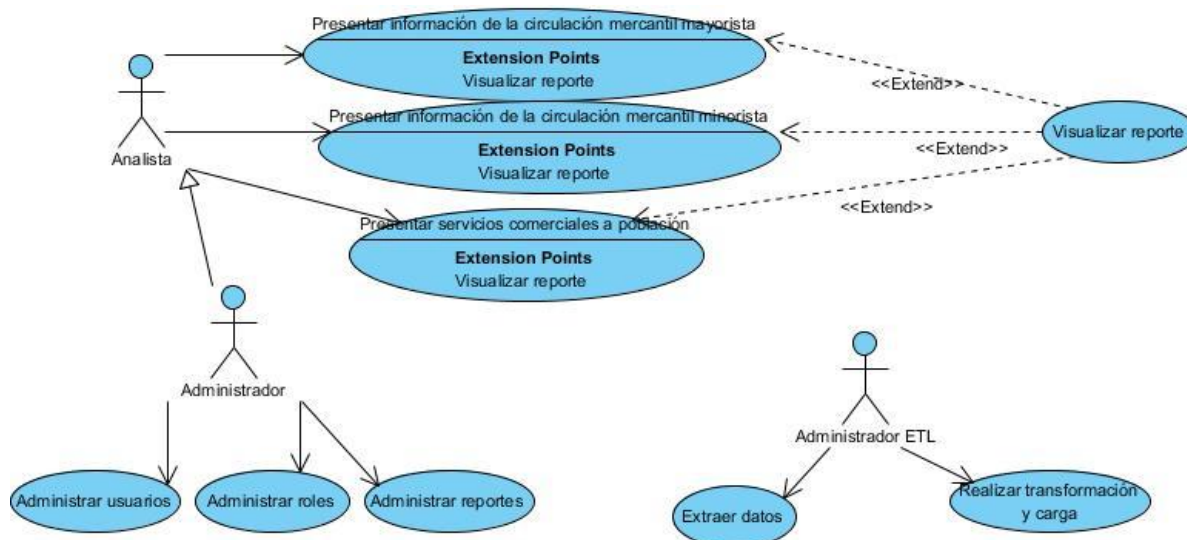


Figura 7: Diagrama de Casos de Uso del Sistema.

### 2.1.9 Descripción de los Casos de Uso.

A continuación se muestra la descripción de los actores y los caso de uso de información.

Actor	Objetivo
Analista	Es el que tiene la tarea de realizar todos los pedidos de información que el cliente quiere que se muestren.
Administrador	Es el encargado de administrar todo lo referente a usuarios, roles y reportes en el AD.
Administrador ETL	Es el encargado de realizar la extracción de la información desde las distintas fuentes, la transformación y carga de los datos al AD.
Caso de Uso	Descripción
CU1_Presentar información de la circulación mercantil mayorista.	Visualiza toda la información referente a la circulación mercantil.
CU2_ Presentar información de la circulación mercantil minorista.	Visualiza todas las ventas que se han realizado por concepto de circulación mercantil minorista.

CU3_Presentar los servicios comerciales a población.	Visualiza todos los servicios comerciales que han sido prestados a la población.
Extraer datos del Comercio Interno.	Realiza la extracción de los datos.
Realizar transformación y carga de los datos del Comercio Interno.	Realiza la transformación y carga de los datos.
Administrar reporte.	Elimina, Inserta y Modifica los reportes OLAP que se visualizan.
Administrar usuario.	Inserta y elimina los usuarios que existen el sistema.
Administrar rol.	Inserta y elimina los roles que existen el sistema.
Autenticar usuario.	Realiza la autenticación de los usuarios en el sistema.
Analizar reporte.	Brinda las diferentes funcionalidades en las que se pueden analizar el reporte.

**Tabla 1: Descripción de los actores y CUS.**

A continuación se describe el CUS “Presentar información de la circulación mercantil mayorista”(para mayor información referente a los otros CUS, consultar expediente de proyecto).

<b>Objetivo</b>	Presentar información de la circulación mercantil mayorista.
<b>Actor</b>	Analista.
<b>Resumen</b>	El caso de uso comienza cuando el Analista selecciona que aspecto de la circulación mercantil desea analizar, luego se generan los reportes concernientes a los datos analizados, y finaliza cuando se muestran los resultados.
<b>Complejidad</b>	Alta.
<b>Prioridad</b>	Media.
<b>Precondiciones</b>	<ul style="list-style-type: none"> <li>✓ Deben estar todos los datos cargados en el mercado, para poder analizar la información deseada. (Complejidad del Mercado y Carga de los Datos.)</li> <li>✓ El analista tiene que estar autenticado en el sistema.</li> <li>✓ Los reportes relacionados con los valores instantáneos fueron creados.</li> </ul>
<b>Postcondiciones</b>	<ul style="list-style-type: none"> <li>✓ Los reportes correspondientes al L.T 1 Circulación mercantil mayorista han</li> </ul>

	sido consultados por el analista.	
<b>Flujo de eventos</b>		
<b>Flujo básico: Presentar información de la circulación mercantil mayorista</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Se autentica en el sistema.	
2.		Muestra la interfaz principal con las áreas de análisis existentes.
3.	Selecciona el área de análisis A.A.G SIGOB.	
4.		Muestra las áreas de análisis que están contenidas dentro del A.A.G SIGOB.
5.	El analista selecciona el área de análisis A.A Series comercio interno.	
6.		Muestra los libros de trabajo que están contenidos dentro del A.A Series comercio interno.
7.	Selecciona el libro de trabajo L.T 1 Circulación mercantil mayorista.	
8.		Muestra los reportes contenidos dentro del L.T 1 Circulación mercantil mayorista.
9.	Selecciona el reporte deseado.	
10.		Muestra la información contenida en el reporte seleccionado y brinda la posibilidad al analista de hacerle cambios al reporte para su análisis. Ir al Caso de Uso <b>Visualizar reporte</b> . Finaliza el caso de uso.
<b>Flujos alternos</b>		
<b>Nº 1 .Los datos son incorrectos</b>		
	<b>Actor</b>	<b>Sistema</b>
10.		Muestra mensaje “Verifique sus datos existen errores en los mismos.”
<b>Opciones de reportes de Presentar información de la circulación mercantil mayorista</b>		
<b>Perspectivas de análisis</b>		<b>Posibles resultados</b>
	<b>Medidas</b>	<b>Periodicidad</b>



<p>Variables de entrada relacionadas con el CU. Presentar información de la circulación mercantil mayorista.</p> <ul style="list-style-type: none"> <li>✓ Temporal Año.</li> <li>✓ Indicador Productos.</li> <li>✓ Indicador Unidades Físicas</li> </ul>	<p>Variables de salida relacionadas con el CU. Presentar información de la circulación mercantil mayorista.</p> <ul style="list-style-type: none"> <li>✓ total dinero por unidades físicas.</li> <li>✓ total dinero recaudado en valor.</li> <li>✓ total de dinero red minorista.</li> </ul>	<p>Rango de tiempo en que se solicitan las variables de salida:</p> <ul style="list-style-type: none"> <li>✓ Anual.</li> </ul>
<p><b>Relaciones</b></p>	<p><b>CU Incluidos</b></p>	<p>No contiene</p>
	<p><b>CU Extendidos</b></p>	<p>Visualizar reporte: 14.1 Valor de la circulación mercantil mayorista total y a la red minorista por grupo de productos. 14.2 Circulación mercantil mayorista total y a la red minorista por unidades físicas.</p>
<p><b>Requisitos no funcionales</b></p>	<ul style="list-style-type: none"> <li>• Cumplir con las pautas de diseño de las interfaces.</li> <li>• Mostrar los mensajes, títulos y demás textos que aparezcan en la interfaz del sistema en idioma español. Los mensajes de error deberán ser lo suficientemente informativos para dar a conocer la severidad del error.</li> <li>• El sistema debe estar disponible durante el horario de trabajo.</li> <li>• Garantizar la conexión de los usuarios al mismo tiempo, según la cantidad de máquinas disponibles.</li> <li>• Utilizar los lenguajes de programación y herramientas definidos durante la investigación.</li> </ul>	
<p><b>Asuntos pendientes</b></p>		

Tabla 2: Descripción del CUS “Presentar información de la circulación mercantil mayorista.”

### 2.1.10 Perfilado de datos

Este proceso consiste en hacer un análisis de los datos origen, con el objetivo de conocer el formato, la estructura y la calidad de los datos, para de esta forma establecer reglas adecuadas para erradicar algunos problemas que puedan existir con la calidad de los datos de las fuentes y de esta manera garantizar la aceptación de los datos almacenados en la BD.

Para el correcto tratamiento de estos errores fue necesaria la utilización de la herramienta DataCleaner, anteriormente explicada.

Durante el análisis de los reportes del mercado de datos Series de comercio interno, en el reporte estándar de medidas se puede ver la cantidad de filas que poseen las fuentes de datos, además se puede presenciar los campos con valores nulos, vacíos, el mayor y menor valor que trae cada uno de estos en la fuente de datos. Así, con esta información poder realizar una correcta limpieza y transformación de los datos (**Anexo 1**). Como resultado del perfilado de los datos se obtuvo:

- Que existen valores nulos o vacíos.
- El tipo de dato identificado es varchar.
- Existen valores duplicados, distintos y únicos.

Después de haber consultado y analizado con el cliente los resultados del perfilado, los problemas en los datos de las fuentes quedaron erradicados.

### 2.2 Diseño del Mercado de datos Series de comercio interno

En un mercado de datos, la parte más importante en el desarrollo de la solución es el diseño, por lo que se hace necesario llevar a cabo una buena realización del mismo.

A diferencia de los clásicos sistemas de BD que presentan sus estructuras diseñadas mediante el modelo Entidad-Relación, los almacenes de datos se diseñan mediante un modelo dimensional. Además poseen la misma información que el Diagrama de Entidad Relación (DER), pero la organiza de forma diferente para garantizar la velocidad y eficiencia en la recuperación de la misma. Una de sus características principales es que no necesita una predefinición de los reportes, debido a que se diseñan de forma tal que cubra el universo de variantes que los usuarios necesiten consultar en la información almacenada.

El modelo dimensional divide el mundo de los datos en dos grandes conjuntos: las medidas y las descripciones del entorno de estas medidas. Las medidas, que generalmente son numéricas, se

almacenan en las tablas de hechos y las descripciones de los entornos, que son textuales, se almacenan en las tablas de dimensiones. Las tablas de hechos son las tablas primarias en el modelo dimensional y contienen los valores del negocio. **(21)**

### Hechos

Las tablas de hechos son las tablas primarias en el modelo dimensional, donde el rendimiento de las mediciones numéricas del negocio es almacenado. Generalmente cada tabla de hechos define un mercado de datos determinado debido a que en ellas se almacena la información concerniente al tema en cuestión **(22)**.

### Dimensiones

Las tablas de dimensiones contienen la descripción textual del negocio. En el modelo dimensional, las tablas de dimensiones poseen varios atributos que en su conjunto definen una fila en la tabla. Los atributos de las dimensiones sirven como fuente primaria de las restricciones de las consultas, agrupaciones y las etiquetas de los reportes. Ellos desempeñan un rol de vital importancia dentro del AD debido a que son las llaves que hacen que este sea usable y entendible. Estos atributos son las llaves de entrada a los hechos o medidas almacenadas.

La calidad de todo AD se mide por la definición de los atributos de las dimensiones. Su poder es directamente proporcional a la calidad y profundidad de estos atributos. **(23)**

#### 2.2.1 Dimensiones identificadas

Durante el desarrollo de la investigación se identificaron siete dimensiones:

**Indicadores de los productos:**(dim\_indicador\_productos) la dimensión describe la cantidad de dinero recaudado por concepto de circulación mercantil mayorista y a la red minorista por grupo de productos.

**Indicadores de las unidades físicas:** (dim\_indicador\_unidades\_fisicas) la dimensión describe la circulación mercantil mayorista y minorista en unidades físicas.

**Indicadores de las ventas:**(dim\_indicador\_ventas) la dimensión describe las diferentes ventas que se pueden llevar a cabo en la circulación mercantil.

**Indicadores de establecimientos:** (dim\_establecimientos) la dimensión describe la cantidad de dinero recaudado por la circulación mercantil minorista por tipo de establecimientos.

**Indicadores de los servicios comerciales:**(dim\_servicios\_comerciales) la dimensión describe la cantidad de dinero recaudado por concepto de servicios comerciales prestados ya sea a la población o en valor.

**Indicadores de provincias:**(dim\_provincia) la dimensión describe como esta dividido el país por provincias.

**Indicadores de los años:**(dim\_temporal\_anno) la dimensión describe los diferentes años que se van a tomar en cuenta para el desarrollo del sistema.

### 2.2.2 Hechos identificados

Una tabla de hechos es la tabla central en un esquema dimensional y contienen los valores y las medidas, acorde a lo que se quiere resolver.

El modelo consta de 5 tablas de hechos:

**Cantidad dinero recaudado por grupo de productos:** (hech\_din\_recaudado\_grup\_prod) este hecho almacena toda la información referente a la cantidad de dinero recaudado por la circulación mercantil mayorista total y a la red minorista por grupo de productos y año.

**Cantidad de unidades físicas:** (hech\_circMerc\_unidades\_fisicas) este hecho almacena toda la información referente a la circulación mercantil mayorista total y a la red minorista en unidades físicas por año y productos.

**Cantidad de dinero recaudado por las ventas:** (hech\_ventas) este hecho almacena todo la información referente a la circulación mercantil minorista total y por concepto de la alimentación pública por año.

**Ingresos de las ventas por establecimientos:** (hech\_dinrecau\_tipo\_establecimiento) este hecho almacena toda la información referente a los ingresos totales de las ventas por conceptos en la alimentación pública en el comercio minorista por tipos de establecimientos.

**Cantidad de dinero recaudado por servicios comerciales:** (hech\_dinrecau\_servcomer\_prestados) este hecho almacena la cantidad de dinero recaudado por los servicios comerciales prestados a la población.

### 2.2.3 Matriz Bus

A continuación se muestra la matriz dimensional en la cual se representa la relación entre los hechos y las dimensiones del mercado de datos Series de comercio interno.

Hechos/Dimensiones	D1	D2	D3	D4	D5	D6	D7
H1	x						x
H2		x					x
H3			x				x
H4				x		x	x
H5					x		x

Tabla 3: Matriz Bus.

#### Leyenda:

**D1:** indicador\_productos

**D2:** indicador\_unidades\_fisicas

**D3:** indicador\_ventas

**D4:** establecimientos

**D5:** servicios\_comerciales

**D6:** provincial

**D7:** temporal\_anno

**H1:** din\_recaudado\_grup\_prod

**H2:** circMerc\_unidades\_fisicas

**H3:** ventas

**H4:** dinrecau\_tipo\_establecimiento

**H5:** dinrecau\_servcomer\_prestados

Después de haber analizado la Matriz Bus se pudo constatar que, no existe solapamiento entre hechos y dimensiones.

### 2.2.4 Modelo de Datos

A continuación se expondrá el modelo de datos compuesto por los hechos, dimensiones y medidas que han sido identificados durante la investigación.

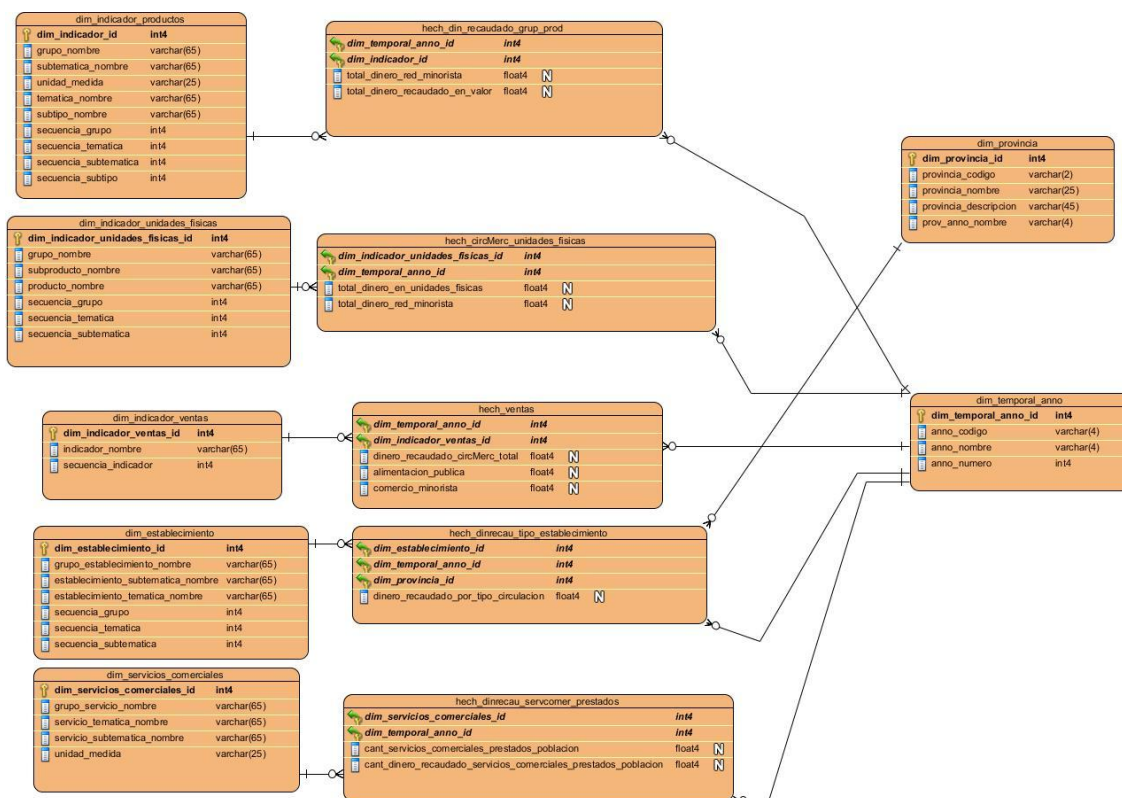


Figura 8: Modelo de Datos.

### 2.2.5 Diseño de los procesos de integración de datos

Para lograr una correcta organización de los procesos de integración de datos, se decidió realizar un diseño de cómo podrían ejecutarse los mismos.

A continuación, se muestra el flujo general del proceso.

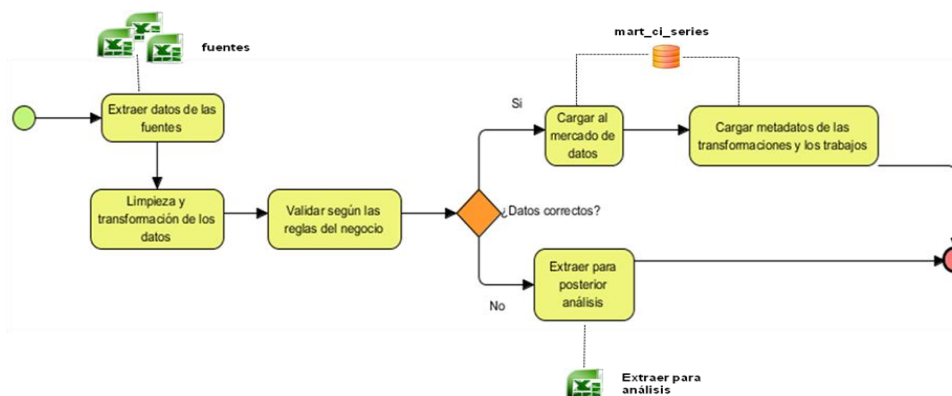


Figura 9: Diseño de los procesos de integración de datos.

A continuación se muestra un ejemplo de cómo poblar una dimensión.

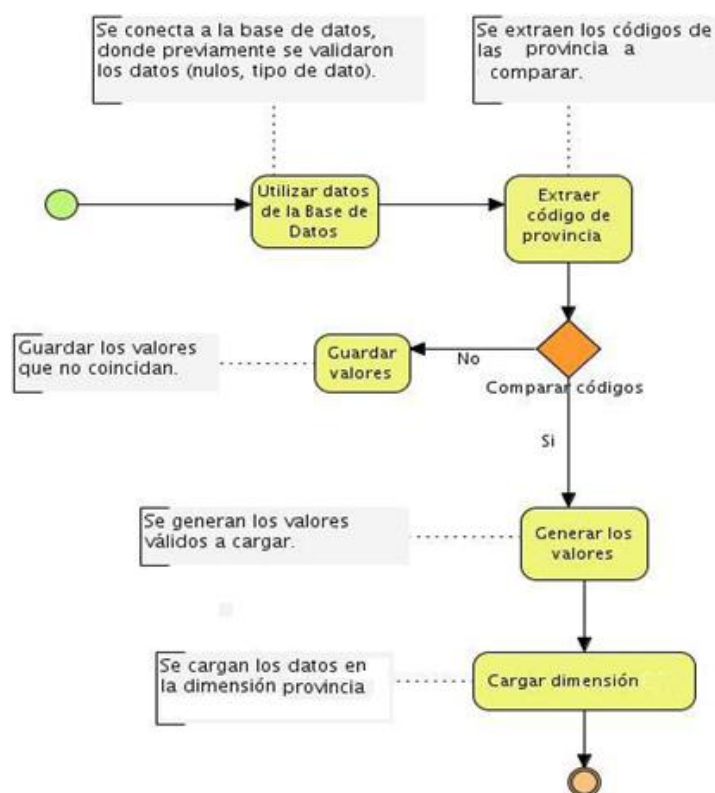


Figura 10: Diseño de la transformación de la dimensión provincia

El diseño visto anteriormente describe los pasos a seguir en la construcción de las transformaciones necesarias para realizar la carga de los hechos y dimensiones.

### 2.2.6 Mapa de navegación

Con el objetivo de que los usuarios accedan a una información con mayor organización y que esta contribuya a un mejoramiento en el apoyo de la toma de decisiones, se llevó a cabo la realización del mapa de navegación de la solución, donde se abordó cada uno de los elementos que componen la estructura de navegación de la información que será mostrada (para un mayor entendimiento de esta estructura, consultar el artefacto Especificación de requisitos de software).

**Área de análisis general (A.A.G):** es el área general donde van a estar contenidas todas las áreas de análisis del almacén de datos. En este caso es el Sistema de Información de Gobierno (SIGOB).

**Área de Análisis (A.A):** constituye las secciones en las que está dividido el almacén de datos, donde se incluyen los Libros de Trabajo y los reportes. El Área de Análisis (AA) identificado es la de Series comercio interno. En esta área se agrupan los libros de trabajo relacionados con el mercado.

**Libro de Trabajo (L.T):** constituye las categorías en los que se agrupan los reportes, donde cada A.A posee varios libros de trabajo. En este caso está compuesta por tres libros de trabajo.

L.T 1 Circulación mercantil mayorista: Este libro de trabajo agrupa dos fuentes las cuales contienen información sobre la circulación mercantil por productos y por unidades físicas, y así mismo contiene 2 reportes candidatos que responden a cada fuente.

L.T 2 Circulación mercantil minorista: En este libro se agrupan 6 fuentes las cuales contienen información de las ventas, en la circulación mercantil, comercio minorista y alimentación pública, con sus respectivos reportes candidatos que en este caso se reducen a cuatro.

L.T 3 Servicios comerciales a población: En este libro se agrupan dos fuentes, las cuales contienen información sobre los servicios comerciales prestados a la población y en valor, los cuales responden a dos reportes candidatos.

**Reportes:** constituye un informe en el que se relacionan de forma que permita tabular los indicadores

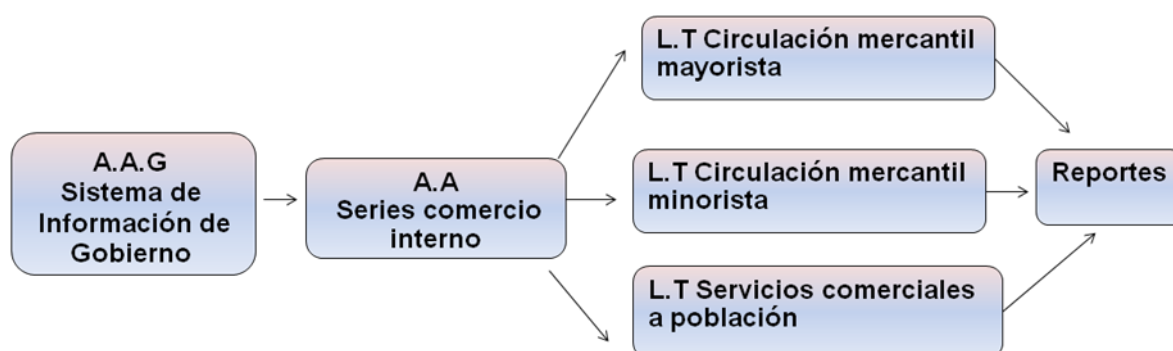


Figura 11: Mapa de Navegación

### 2.2.7 Diseño de los cubos OLAP

En la investigación se diseñaron cinco cubos multidimensionales:

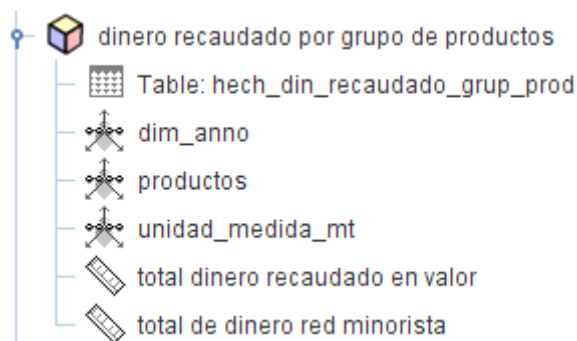
- ✓ Circulación mercantil por unidades físicas.
- ✓ Dinero recaudado por grupo productos.
- ✓ Ventas.



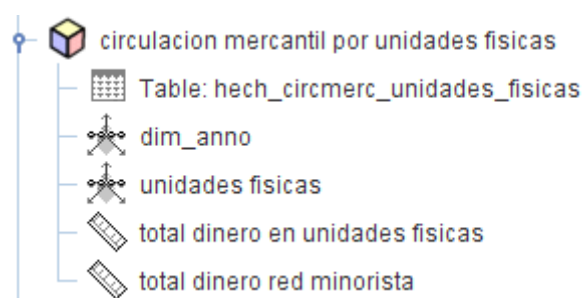
- ✓ Dinero recaudado por tipo establecimiento.
- ✓ Dinero recaudado por servicios comerciales prestados.

El diseño de los cubos se realizó con la herramienta Pentaho Schema Workbench presentada con anterioridad en el Capítulo 1.

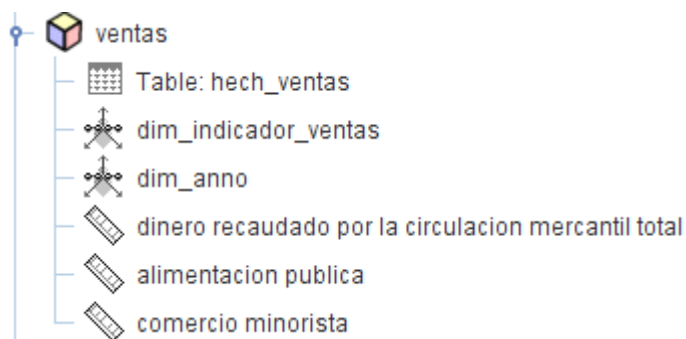
A continuación se muestran los cubos multidimensionales diseñados:



**Figura 12: Cubo dinero recaudado grupo productos.**



**Figura 13: Cubo circulación mercantil por unidades físicas.**



**Figura 14: Cubo ventas.**

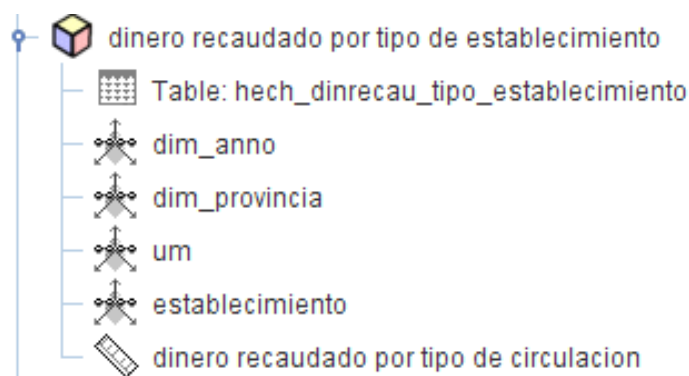


Figura 15: Cubo dinero recaudado por tipo establecimientos.

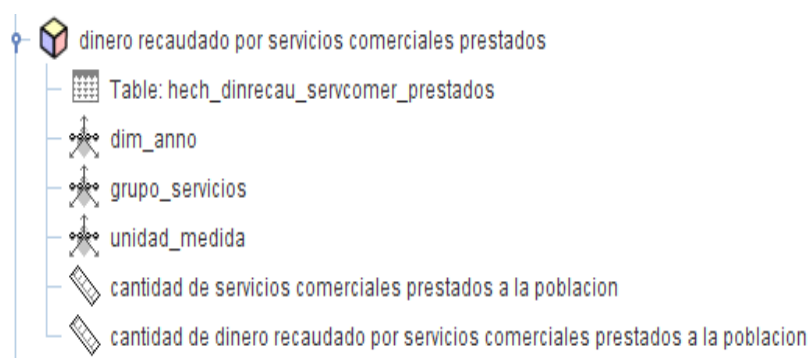


Figura 16: Cubo dinero recaudado por servicios comerciales prestados.

Como se muestra en las figuras anteriores, los cubos OLAP, fueron definidos de acuerdo a la cantidad de tablas de hechos, debido a que estas son la principal fuente de información.

Además se pudo observar la relación existente entre cada hecho con sus respectivas dimensiones y como resultado de esta unión las medidas con que cuenta.

### 2.2.8 Roles y privilegios en la base de datos

El gestor de base de datos Postgres SQL, ofrece a los usuarios la posibilidad según sus necesidades de dar permisos y accesos a cada rol que sea identificado, en función de obtener un mejor control sobre los usuarios, y la información almacenada. A continuación se describirán los roles establecidos:

**Administrador:** tiene acceso a la BD en toda su totalidad, se puede decir que tiene la potestad de administrar y configurar la base de datos, así como el control de los usuarios restantes.

**Analista:** el que ejerce este rol se limita a consultar solamente la información de la BD.

**Administrador de ETL:** su función es la de realizar los procesos de ETL en la interacción con la base de datos.

### 2.2.9 Políticas de respaldo y recuperación

Las políticas de seguridad y respaldo que se utilizarán en el mercado de datos Series de comercio interno están divididas en tres puntos fundamentalmente.

**Periodicidad de las salvas:** en la base de datos de la ONEI se guarda anualmente una salva de este mercado y de la información contenida en él, y en todo momento se verifica que exista una copia escrita de la información almacenada en el servidor.

**Tablas involucradas:** dentro del proceso de desarrollo del mercado se ha visto la necesidad de utilizar cinco tablas de hechos identificadas en el proceso de análisis, al igual que siete tablas de dimensiones.

**Backups existentes:** cada año se realiza una salva en el mes de diciembre con los datos de todo el año, además se tiene un servidor de respaldo con todos los datos para en caso de que ocurra un incidente que atente contra la seguridad de la entidad estos queden a salvo.

### Conclusiones del capítulo

Después de realizado el proceso de análisis y diseño del mercado de datos Series de comercio interno, se concluyó que:

- Se identificaron eficientemente 10 requisitos informativos, 25 requisitos funcionales, 10 requisitos no funcionales y 6 reglas de negocio, de acuerdo a las necesidades de los usuarios en correspondencia con el área de comercio interno en el trabajo con series históricas.
- Se diseñaron eficientemente 5 hechos y 7 dimensiones, pertenecientes al mercado de datos Series de comercio interno.
- Se diseño el modelo de datos perteneciente al mercado de datos Series de comercio interno, identificando nuevas dimensiones y refinando algunas existentes, garantizando el correcto funcionamiento del sistema.

**CAPÍTULO 3: Implementación y Prueba del mercado de datos Series de comercio interno****Introducción**

En el presente capítulo se muestra como está estructurado el modelo físico, se realizará la implementación de los subsistemas de integración y visualización. Se llevará a cabo la fabricación de los reportes candidatos, describirán los procesos de extracción, transformación y carga de los datos al almacén. Además se realizarán diferentes pruebas que parten, de la aplicación de listas de chequeo y los casos de pruebas, con el objetivo de que el mercado de datos cuente con la calidad que merece el cliente.

**3.1 Implementación del subsistema de almacenamiento**

Al implementar el subsistema de almacenamiento quedará estructurado el repositorio de datos, donde se almacenarán los datos de interés para la empresa. Todos estos datos serán cargados en el repositorio después que hayan pasado por los procesos de ETL.

El subsistema de almacenamiento del mercado Series de comercio interno, cuenta con cuatro esquemas, el de las dimensiones compartidas (dimensiones), el que contiene las seis tablas temporales (sgt\_etl), el que contiene los metadatos (Metadatos) en este caso solo se van a utilizar dos tablas la md\_trabajo y md\_transformacion y finalmente el del propio mercado (mart\_ci\_series) que cuenta con cinco tablas de dimensiones y cinco tablas de hechos.

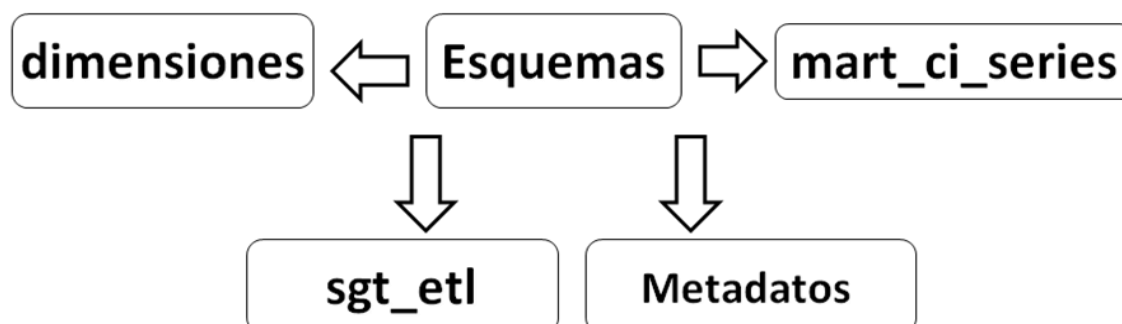


Figura 17: Subsistema de almacenamiento

### 3.2 Implementación del modelo físico

El modelo físico es una representación del diseño de los datos, el cual proporciona conceptos que describen los detalles de cómo se pueden almacenar los datos en un computador, permitiendo de alguna forma describir las estructuras de los datos en la base de datos, y ver la forma en que se relacionan. Parte importante del modelo físico son los esquemas y las tablas que lo componen.

#### Esquemas y tablas

Los esquemas son una forma de tener organizada la información contenida en la base de datos, estos pueden contener tablas, operadores, tipos de datos y funciones. Permiten que el usuario siempre que tenga permiso pueda acceder a ellos.

Actualmente la solución cuenta con siete tablas de dimensiones, cinco tablas de hechos y seis tablas temporales.

<b>dimensiones</b>	-dim_provincia -dim_temporal_anno
<b>mart_ci_series</b>	-dim_indicador_productos -dim_indicador_unidades_fisicas -dim_indicador_ventas -dim_servicios_comerciales -dim_tipo_establecimiento -hech_circmerc_unidades_fisicas -hech_din_recaudado_grup_prod -hech_dinrecau_servcomer_prestados -hech_dinrecau_tipo_establecimiento -hech_ventas -meta_ci_trabajo -meta_ci_transformacion
<b>sgt_etl</b>	-temp_servicios_comerciales_prestados -temp_tipo_establecimiento_primero -temp_tipo_establecimiento_segundo -temp_unidades_fisicas -temp_unidades_fisicas_nuevo

	-temp_ventas
metadatos	-md_trabajo -md_transformacion

Tabla 4 Hechos, dimensiones, tablas temporales y metadatos.

### 3.3 Implementación del subsistema de integración de datos (ETL)

Como parte del proceso de integración de datos, se debe tener en cuenta lo expuesto por Ralph Kimball, acerca de las pautas que se deben cumplir para la correcta integración de los datos al almacén, el mismo identificó cuatro subsistemas de ETL clasificados en los siguientes grupos:

**Extracción:** extraer la información de la fuente de origen.

**Limpieza y conformación:** consiste en acciones que permiten validar y aumentar la calidad de la información.

**Entrega:** consiste en la preparación de la información para su posterior entrega.

**Gestión:** este proceso es similar al proceso de carga, ya que posibilita mediante los subsistemas que lo integran, añadir nuevos datos al repositorio, y realizar cambios en los procesos de ETL, para obtener una correcta consolidación de los datos en el almacén (24).

#### 3.3.1 Arquitectura de integración

En el proceso de desarrollo de un software, la arquitectura es el diseño más importante en la estructura del sistema, debido a que conjuga un conjunto de patrones que permite guiar la construcción del mismo. Para un correcto desarrollo de la aplicación es recomendable que se analice rigurosamente todo el proceso de integración de los datos.

En este proceso la arquitectura está fundamentada por algunos elementos necesarios para la correcta implementación del sistema que se desea desarrollar, por tal motivo hay que tener en cuenta las fuentes de datos, el área temporal y el mercado de datos en sí.

**Fuente de datos:** son los ficheros que guardan información histórica de los sistemas, los cuales se encuentran en formato excel.

**Área temporal (Staging Area):** es la línea divisoria entre la fuente de origen y el mercado de datos. Permite realizar la integración de los datos y su transformación, además de poder tener una copia temporal de seguridad de los datos contenidos en las fuentes.

**Mercado de datos:** constituye el destino a donde se integrarán los datos a cargar.

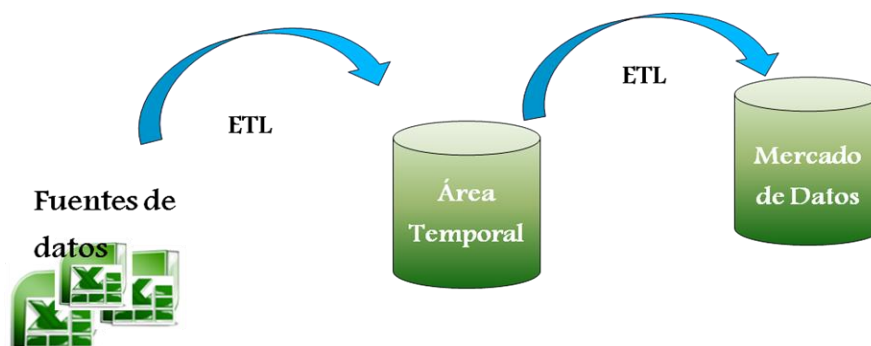


Figura 18: Arquitectura de integración.

### 3.3.2 Implementación de los trabajos

En el mercado de datos Series de comercio interno fueron implementados tres trabajos o Jobs (en inglés), los cuales siguen un orden en la jerarquía, primeramente tenemos al trabajo principal el cual es el encargado de llamar a los dos restantes trabajos el de las dimensiones y posteriormente el de los hechos, con sus respectivas aéreas temporales.

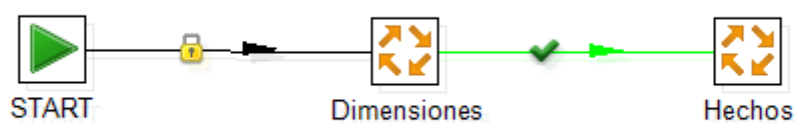


Figura 19.Trabajo principal.

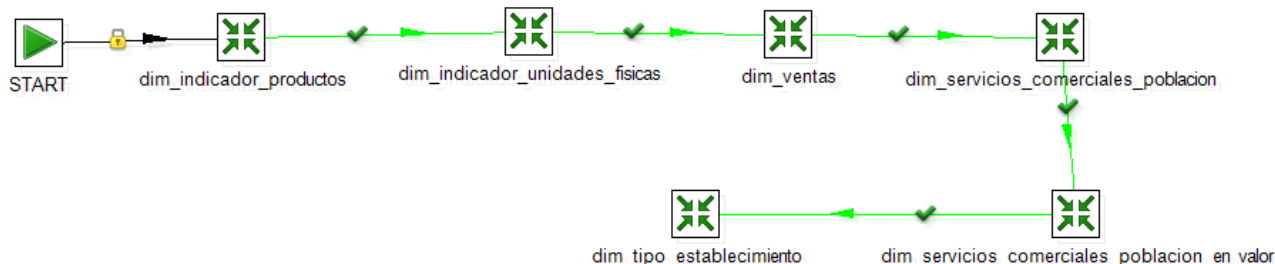


Figura 20: Trabajo de las dimensiones.

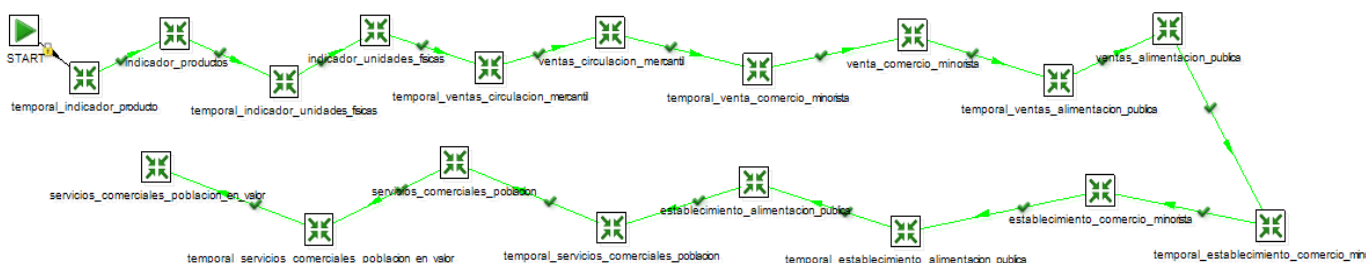


Figura 21: Trabajos de los hechos.

En los trabajos, se pueden encontrar agrupados las transformaciones o los hechos, los cuales siguen un orden de ejecución de tal forma que hasta que no haya terminado la ejecución completa de la primera transformación este no pasa a la siguiente y así sucesivamente, hasta completar en su totalidad la ejecución de todo el trabajo.

### 3.3.3 Implementación de las transformaciones

La información almacenada en las fuentes, es extraída para proceder a realizar el proceso de transformación y limpieza, ajustándola a las necesidades de información del cliente. Posteriormente se cargan los datos en las tablas del área temporal.

La transformación de los datos se hará acorde a las reglas de negocio definidas. Para su ejecución se realizaron las transformaciones de las dimensiones y de los hechos correspondientes.

A continuación se muestra un ejemplo de la transformación realizada para la carga de la dimensión indicador\_productos y del hecho din\_recau\_grup\_prod, las demás se pueden encontrar en la carpeta de fuente en la solución en formato ktr.



En esta transformación se tratan los datos relacionados con los indicadores de los productos, relacionados con el comercio interno. La información se carga de la tabla temp\_unidades\_fisicas y posteriormente se realizan las validaciones pertinentes a los datos para que al poblar el hecho los datos almacenados sean los deseados. Finalmente se carga el hecho con la transformación correspondiente.

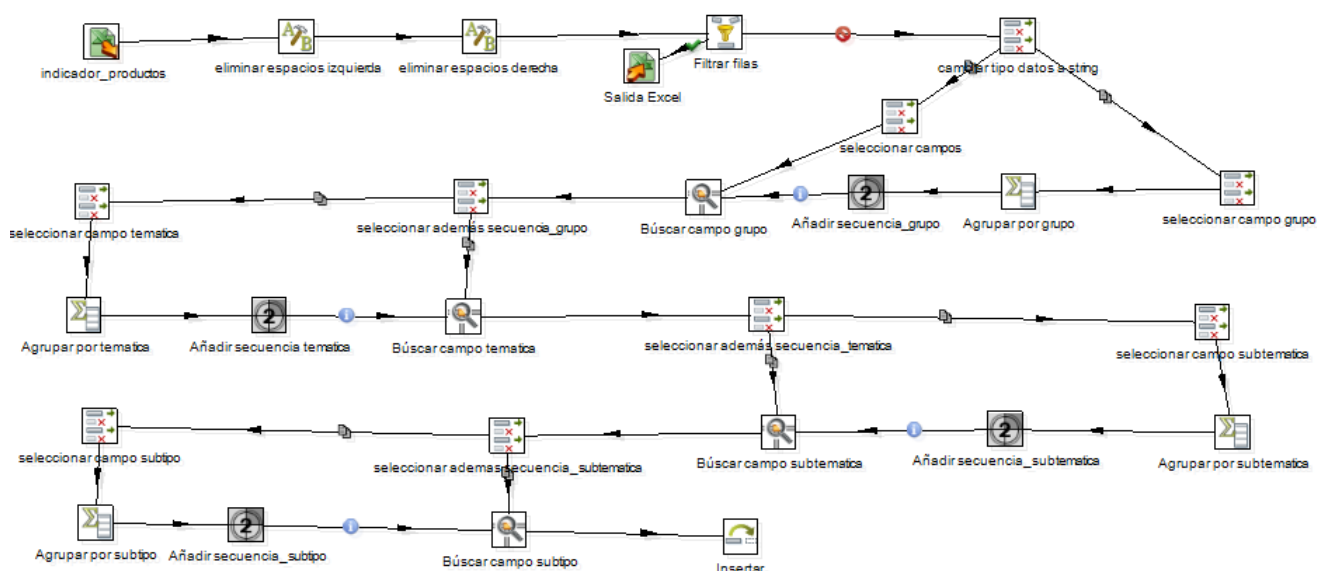


Figura 22: Transformación de la dimensión indicador\_productos.

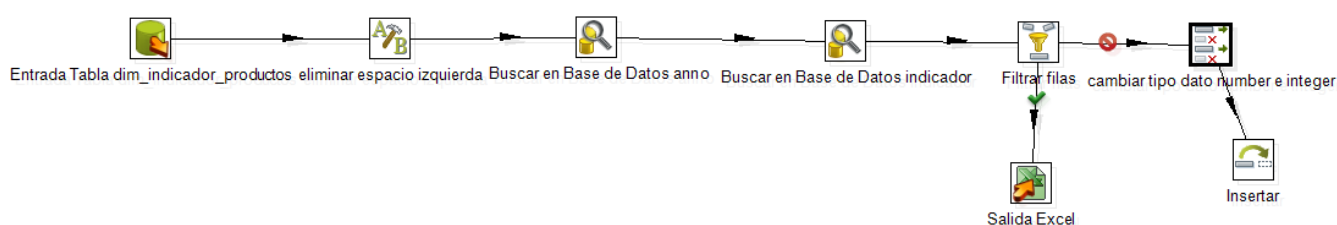


Figura 23: Transformación del hecho din\_recaud\_grupo\_prod.

### 3.4 Implementación del subsistema de visualización

Para desarrollar el sistema es preciso conocer como están integradas las fuentes, por lo que se hace necesario la implementación de los reportes candidatos.

#### 3.4.1 Implementación de los reportes candidatos

Los reportes candidatos o tablas de salida como también son conocidos, son los que van a contener los valores de los datos que son de interés para los clientes, estos son implementados a través de



unidades físicas	Anno 2002		Anno 2003		Anno 2004		Anno 2005		Anno 2006		Anno 2007		Anno 2008		Anno 2009		Anno 2010		
	Total	Red minorista	Total	Red minorista	Total	Red minorista	Total	Red minorista	Total	Red minorista	Total	Red minorista	Total	Red minorista	Total	Red minorista	Total	Red minorista	
Alimentos	4,122	2,615	4,086	2,323.8	4,912.43	2,905.4	5,794.567	3,168.228	5,432.6	3,218.1	5,145.1	3,316	6,277.8	3,372.1	6,336.1	3,757.3	7,262.8	3,734.1	
Aceite vegetal refinado Mt	24.6	18.6	70.7	47.3	44.7	24.6	63.4	27.8	68.7	36	70	33.7	67.3	31.8	102.5	51.7	70.4	34.2	
Amoz consumo Mt	475.9	335.8	508.9	314.5	687.1	483.7	732	528.2	669.6	601.3	671.2	644.9	622	472.1	520.6	411.8	756.7	643.8	
Azúcar Mt	395.8	276.9	398.5	277	532.7	378.5	652.3	364.8	634.3	281.9	427.3	316.3	1,575.3	498.9	1,680.4	791.5	1,681.9	657.9	
Bebidas alcohólicas (excluye vinos)	360.2	141.7	156.5	80.2	617.43	96.7	157.8	67.3	158.7	61.2	194.8	-0.7	181.6	90.8	136.5	64.9	296.9	56.7	
Carne de cerdo en bandas Mt	15.2	3.2	18.4	2.6	17.6	1.1	25.3	0.7	31.8	2.3	50.6	2.9	56.8	5.2	45.5	5.4	59.1	14.4	
Carne desh. de res (exc.hig.) Mt	21.6	8.9	18.6	6.8	14.7	3.5	20.9	5.3	20.1	6.1	20.1	5.6	22.8	7.5	23.4	8	24.9	9.5	
Carne fresca de aves (incluye ave nacional) Mt	22.2	10.5	44.8	28	109.5	29.7	60.9	29.9	52	25.8	101.3	40.6	64.3	32.6	87.6	53.4	90.7	53.1	
Conservas de frutas y vegetales	23.6	15.8	27.5	6.2	35.7	17.2	50.7	21.9	53.3	25	57.4	25.8	59.3	26.6	64.5	23.8	56.9	23.5	
Frijoles	129.4	67.7	181.2	75.7	2.3	1.5	4.9	3	1.9	1.1	5.2	3.7	5	3.8	13.2	6.9	21.8	5.8	
Frutas	132.1	112.3	194	112.3	27.7	157.4	88.5	128.429	80.63	139	47.2	108.3	58.7	122.8	36.4	119.5	25.9	312.3	85
Galleta Mt	17.8	7.9	14.8	7.3	17.4	9.1	34.3	22.6	26.6	10.8	27.4	12.1	33.6	29.3	37.7	19.8	35.5	21.5	
Granos Mt	129.5	67.7	181.2	75.7	291.7	133.1	278.1	188.6	334.2	139.3	231.4	155.9	221.9	143.4	166.7	110.5	139.2	89.3	
Helados Mmgal	8.4	1	8.1	0.3	13.4	0.3	13.9	0.4	10.6	0.7	12	0.5	14.6	0.8	12.9	0.2	17	1.5	
Hortalizas y legumbres	133.4	89.1	91.8	31.8	112.5	99.3	171.908	72.403	113.1	38	122.2	65.3	121.8	50.3	151	80.3	459.6	78.7	
Huevos para consumo MMU	1,037.7	623.6	949.5	472.8	905.8	480.2	1,677.5	888	1,734.9	1,083.9	1,703.3	1,087.1	1,778.5	1,093.8	1,849	1,253.5	1,728.8	1,122.9	
Hígado, vísceras y otros comestibles Mt	12.6	1.3	10.3	1	11.2	0.6	13.2	1.2	15.9	1.7	18.7	2	23.5	3.1	19.1	2.9	19.5	3.2	
Lacteosy Mt	8.2	5.8	8	6.6	8	6.8	17.2	8.5	13.9	4.9	12	7.3	7	3.5	5.1	2.9	6.3	3.7	
Leche en polvo Mt	14.7	11.1	15.5	11.6	27.7	24.8	36.6	30.4	40.3	36.9	38.2	35.4	24.8	20.6	32.6	26.8	39.2	32.9	
Leche evaporada Mt	8.7	7.8	17.1	15.6	20.4	17.3	42.6	19.7	44.6	43.9	55.6	53.9	59.7	57.8	44.7	43.6	0.9	0.2	
Leche pasteurizada y cruda Mt	217.5	200.6	236.6	210.8	250	202.9	242.2	113.1	153.7	135	139.7	119.8	104.8	88	119.5	103.8	114.3	97.5	
Pan Mt	509.3	333.1	508.9	333.5	507.5	332.1	453	313.9	437.6	309.1	469.5	322.7	468	303.1	475.5	351.5	474.9	315.1	
Pasta alimenticias Mt	19.1	15.8	21.3	17.2	31.3	23.7	26	20.1	22.8	16.8	32.8	23.3	48.1	33.7	55.1	29.1	58.8	35.8	
Pescados y mariscos frescos Mt	30.2	16	26	15.7	66.9	31.3	63.2	53.1	40.6	17.9	54.3	20	60.3	27.3	62.4	28	53.6	21.4	
Picadillo de res extendido Mt	18.2	16.6	19.9	16.8	26.7	22.2	23.9	13.3	22.4	19	24.2	20	26	18.8	28.6	23.2	28.3	23.5	
Quesos Mt	7	0.6	8.8	0.5	8.9	0.2	9.6	0.2	13	1.2	13.2	1	13.4	13.4	18.5	1.3	15.2	1	
Sal Mt	70.2	48.9	69.9	46.7	72	50.4	80.7	54.2	159.3	45.4	161.5	46.7	150.5	40.5	154.3	35.6	43.7	26.8	

Figura 25: Reporte 14.2 Circulación mercantil mayorista total y a la red minorista en unidades físicas.

La vista del reporte puede ser cambiada de acuerdo a las exigencias del usuario, esto se puede llevar a cabo gracias a la utilización de los componentes que se muestran en la barra de herramientas (Figura25).

### 3.5 Importancia de la calidad del software

Hoy en día en el mercado actual competitivo no es suficiente con desarrollar un producto o servicio y distribuirlo, lo más importante es lograr la aceptación del cliente, y esta se logra, ofreciendo un producto con calidad. Sin embargo obtener un software con calidad no es un estado que se alcanza fácilmente, debido a que, al decir calidad estamos diciendo que nuestro producto ofrece confiabilidad, usabilidad, seguridad, portabilidad e integridad.

Por lo que para garantizar que la solución final sea la más idónea, se deben realizar diferentes pruebas de software como las pruebas unitarias, las pruebas de integración, las pruebas de sistema y las pruebas de aceptación que permitan uniformar la filosofía de trabajo, en aras de lograr una mayor confiabilidad, mantenibilidad y facilidad, a la vez que eleven la productividad, tanto para la labor de desarrollo como para el control de la calidad del software.

### 3.6 Pruebas de software

#### 3.6.1 Modelo V

A la hora de seleccionar el modelo que servirá como guía en las actividades de prueba, se decidió utilizar el Modelo V, y no el modelo en Cascada aunque estos dos se relacionan mucho, debido a que el Modelo V, considera las actividades de prueba como un proceso que corre en paralelo con las actividades de análisis y diseño, donde se realizan pruebas durante todo el ciclo de vida del software, y no se establece como una fase independiente al final del proyecto, como ocurre en el Modelo en Cascada. (25). Por otra parte el Modelo V, es el modelo definido por el Centro de Tecnologías de Gestión de Datos (DATEC), además es uno de los modelos más idóneos para obtener un producto con mayor calidad, logrando alcanzar mejoras en el proceso de pruebas.

Al utilizar el Modelo V (Figura 21), se puede ver que la codificación forma el vértice de la V en conjunto con el análisis y el diseño a la izquierda, y las pruebas y el mantenimiento a la derecha. La unión con líneas discontinuas entre las fases de la parte izquierda y las pruebas de la derecha representa una doble información. Por un lado sirve para indicar en qué fase de desarrollo se deben definir las pruebas correspondientes. Por otro lado, sirve para saber a qué fase de desarrollo hay que volver si se encuentran fallos en las pruebas correspondientes (26).

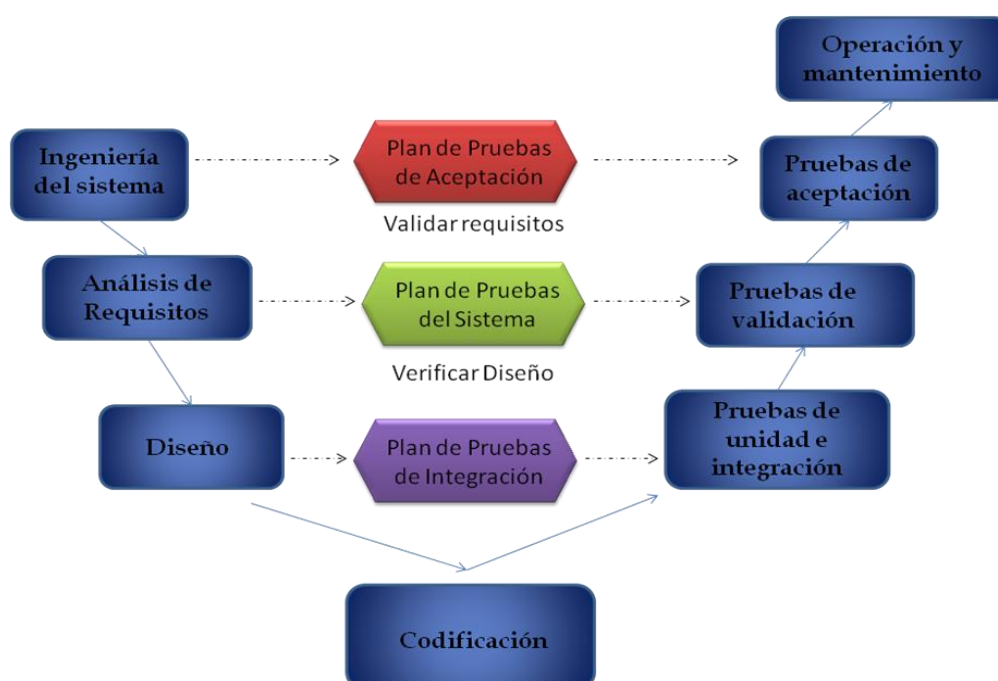


Figura 26: Modelo V.

Entre las pruebas que pueden ser aplicadas al software se pueden encontrar:

**Prueba unitaria:** se enfocan en un programa o un componente que desempeña una función específica que puede ser probada y que se asegura que funcione tal y como lo define la especificación del programa. Los programadores siempre prueban el código durante el desarrollo, por lo que las pruebas unitarias son realizadas solamente después de que el programador considera que el componente está libre de errores.

**Prueba de integración:** su objetivo es identificar errores introducidos por la combinación de programas o componentes probados unitariamente, además, verificar que las especificaciones de diseño sean alcanzadas. Componentes individuales son combinados con otros componentes para asegurar que la comunicación, enlaces y los datos compartidos ocurran apropiadamente. No son verdaderamente pruebas de sistema porque los componentes no están implementados en el ambiente operativo.

**Prueba de sistema:** son usualmente conducidas para asegurar que todos los módulos trabajan como sistema, sin error. Es similar a la prueba de integración pero con un alcance mucho más amplio. Las pruebas del sistema examinan qué tan bien el sistema cumple con los requerimientos de la organización y su utilidad, seguridad y desempeño. También se realizan estas pruebas a la documentación del sistema. **(27)**

Independientemente de estos tipos de pruebas comentados anteriormente, en el centro DATEC se llevan a cabo otro grupo de pruebas encaminadas a la terminación del ciclo de desarrollo del producto. Entre las que podemos encontrar:

**Pruebas internas:** son realizadas por todo el equipo de desarrollo y por los especialistas que participan en la solución.

**Pruebas de liberación:** son realizadas por los especialistas de calidad, en primera instancia el software es evaluado por los especialistas de calidad de DATEC y posteriormente por los especialistas del Centro de Calidad para Aplicaciones Tecnológicas (CALISOFT).

**Pruebas de aceptación:** son las pruebas que se realizan directamente con el cliente, donde el mismo evalúa y valida, de acuerdo a su conformidad con el producto final. **(28)**

### 3.6.2 Herramientas de prueba

#### Aplicación de los Casos de Prueba

Un caso de prueba es diseñado según las funcionalidades definidas en los casos de usos de información. El mismo se realiza para comprobar que la documentación del producto se corresponde con lo que se ha establecido en la aplicación. Para el mercado Series de comercio interno se elaboraron 3 casos de prueba, en donde fueron agrupados los 10 requisitos de información.

A continuación se muestra uno de los diseño de caso de prueba correspondiente al caso de uso “Presentar información de la circulación mercantil mayorista”

Escenario	Descripción	Variables de Entrada	Variables de Salida	Respuesta del sistema	Flujo central
EC 1.1	Reporte que muestra el valor de la circulación mercantil total y a la red minorista por grupos de productos.	-Año -Indicador unidades físicas	-cantidad de dinero recaudado por la circulación mercantil mayorista total y a la red minorista por grupo de productos.	El sistema muestra todas las variables disponibles para los análisis, ubicados en las filas y las columnas que pueden ser visualizadas	Se abre la aplicación. Se autentifica. Se entra al sistema. Se despliega hacia la derecha el componente ubicado en el lateral izquierdo que contiene el
EC 1.2	Reporte que muestra circulación mercantil mayorista total y a la red minorista en unidades físicas.	-Año -Indicador productos	-circulación mercantil mayorista total y a la red minorista en unidades físicas.	para cada reporte.	navegador. Se selecciona el áreas de análisis <b>A.A Series Comercio interno</b> Se selecciona el libro de trabajo L.T 1 Circulación mercantil mayorista En la parte inferior izquierda se selecciona el reporte deseado. En el área de trabajo se visualiza la tabla correspondiente al reporte.

Tabla 5: Presentar información de la circulación mercantil mayorista.

**Aplicación de listas de chequeo**

Otras de las pruebas aplicadas son las relacionadas con las listas de chequeo, las cuales constituyen un componente importante para el control de los riesgos y tienen como funcionalidad principal la de detectar situaciones peligrosas que pueden provocar daños en el software.

La evaluación de las listas de chequeo se rige por un documento que tiene un conjunto de parámetros a medir sobre un aspecto determinado, que puede ser un aspecto de la calidad del software que se esté probando o las características del documento.

Estructura del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
crítico	1. ¿Los entregables contienen las secciones obligatorias de la plantilla estándar definidas para un expediente de proyecto? (portada, control de versiones, reglas de confidencialidad, tabla de contenidos y contenido) (ver expediente de proyecto).	0			
crítico	2. ¿El alcance del proyecto describe correctamente los datos de las dimensiones y hechos del mercado de datos?	0			
crítico	3. ¿El objetivo expresa correctamente el propósito del documento?	0			
	4. ¿Se hace un uso adecuado del control del documento?	0			
	5. ¿En la sección de acrónimos se definen todos los acrónimos utilizados en el documento?	0			
	6. ¿En el entregable, la definición de las variables se hace correctamente?	0			

	7. ¿Existe una adecuada correspondencia entre las variables definidas y las descripciones que tienen estas variables?	0			
Indicadores definidos en el desarrollo					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
	1. ¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?	0			
critico	2. ¿Los reportes son configurables a través de la interfaz del sistema?	0			
	3. ¿La interfaz está orientada a facilitar el uso de las funciones del sistema por parte de los usuarios?	0			
	4. ¿El sistema responde de una forma rápida a la información que le sea solicitada por el usuario?	0			
critico	5. ¿El sistema garantiza la confidencialidad y seguridad de acceso a los datos por rol de los usuarios?	0			
	6. ¿Los datos e información derivados del proceso de análisis realizado mediante la aplicación, apoyan la toma de decisiones en la Institución?	0			
critico	7. ¿Los cambios en los datos se reflejan automáticamente en los reportes de forma instantánea?	0			
Semántica del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios



crítico	¿Se ha identificado errores ortográficos en el entregable?	0			
crítico	¿Se entiende claramente lo que se ha especificado en el documento?	0			
	¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?	0			

Tabla 6: Aplicación de las listas de chequeo.

### 3.6.3 Resultado de Pruebas

#### Pruebas unitarias

Mientras que se iban desarrollando e implementando los subsistemas se le fueron aplicando pruebas unitarias a cada uno por separado, lo que trajo consigo que se identificaran problemas en el orden de los datos por lo que se definió agregar a las transformaciones pertinentes los componentes Agrupar por, Añadir secuencia, Buscar campos, Agrupar campos por, entre otros, lo que permitió que todos los datos quedaran ordenados según las fuentes.

Además se validó que en el subsistema de visualización, algunos de los totales que se mostraban no coincidían con los de la fuente. Pero el problema quedó solucionado, al consultar con el cliente salió a relucir que muchos de los totales que se encontraban en las fuentes eran datos preliminares, quiere decir, que esos totales que estaban publicados en las fuentes, eran datos que estaban arreglados, a conveniencia de la institución, por tal razón al sumalizar los datos que se encontraban en la fuente algunos totales no coincidían. Pero se llegó al acuerdo que se iban a tomar así, ya que no era problema de la aplicación. Por lo que las pruebas unitarias arrojaron 2 No Conformidades (NC), las que quedaron erradicadas.

#### Pruebas de Integración

Durante la aplicación de estas pruebas se pudo constatar que existían algunos errores a la hora de integrar el sistema con los diferentes módulos que lo componen, por lo que salieron a relucir 2 NC, las que con ayuda de los especialistas del departamento quedaron solucionadas.

### Pruebas de sistema

Después de la aplicación de estas pruebas por los especialistas de DATEC, estas arrojaron como resultado 10 NC, las que quedaron anuladas siguiendo las especificaciones de los especialistas.

### Listas de Chequeo

Después que el mercado de datos Series comercio interno ha pasado por la revisión con las listas de chequeo, se generó el grafico de barras (Figura 22), el cual muestra el comportamiento de los 17 indicadores identificados de los cuales 8 son críticos.

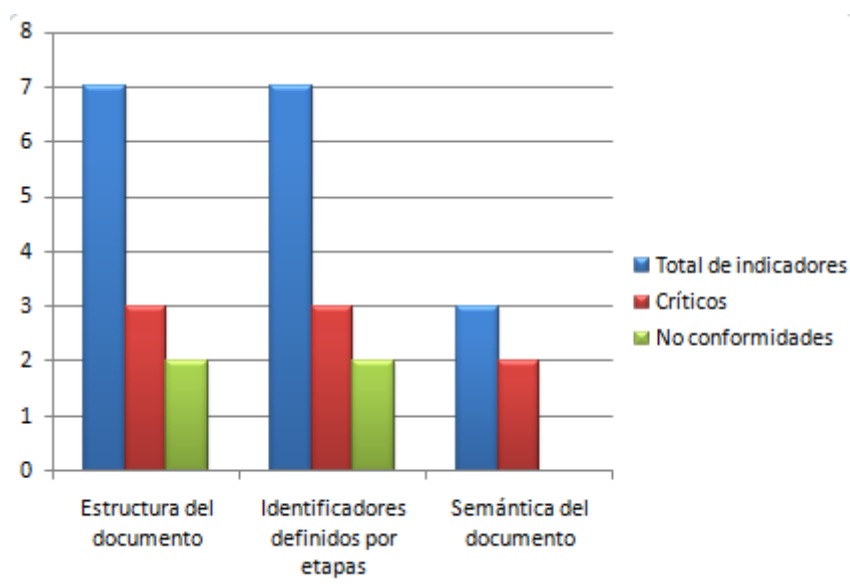


Figura 27: Resultado de las listas de chequeo

### Prueba de liberación

Después de haber pasado por la revisión de todos los especialistas de CALISOFT, el producto fue liberado debido a que, arrojó como resultado 0 no conformidades, lo que demuestra que el mercado de datos Comercio Interno en el trabajo con series históricas fue realizado con calidad.

### Pruebas de aceptación

Dadas las pruebas de aceptación realizadas por el cliente, dejó como resultado una carta de aceptación la cual es asignada por el cliente (**Anexo 2**), en donde se demuestra que el mercado Comercio Interno en el trabajo con series históricas, cumple con todas las especificaciones y está listo para utilizarse.



### **Conclusiones del capítulo**

Al concluir la implementación y validación del mercado de datos Series de comercio interno, se pudo constatar que:

- Se llevó a cabo de forma acertada la implementación del modelo físico logrando una estructura robusta entre las relaciones de las tablas multidimensionales, donde se identificaron 3 esquemas, entre los que están distribuidas las tablas del mercado Series de comercio interno.
- Se realizaron los procesos de ETL a partir del desarrollo de las transformaciones y los trabajos, los que ayudaron al correcto llenado de la base de datos, correspondiente al mercado de datos Series de comercio interno.
- A partir de la implementación del subsistema de visualización quedó especificada el área de análisis, los libros de trabajo y los reportes, contribuyendo así con el proceso de toma de decisiones.
- Se realizaron las pruebas definidas por el centro, al mercado de datos con ayuda de diferentes herramientas como las listas de chequeo, dando como resultado un mercado de datos bien elaborado.

## CONCLUSIONES GENERALES

Después de finalizada la investigación, se arribaron a las siguientes conclusiones:

- Se consultó la metodología, las herramientas y tecnologías a utilizar en el desarrollo del mercado de datos Series de comercio interno.
- Después de realizado el análisis y diseño del mercado, se pudo cumplir con todos los requerimiento del negocio, además de satisfacer las necesidades del cliente.
- Al realizar la implementación de los subsistemas de integración y visualización, se pudo obtener como resultado un mercado de datos poblado y funcional.
- Después de haber realizado distintas pruebas al mercado de datos, arrojó como resultado que quedara listo para el apoyo a la toma de decisiones.

### RECOMENDACIONES

- Poner en práctica el mercado de datos en la ONEI, y explotar todas sus funcionalidades dando pasos a detectar nuevas.
- Garantizar la superación del personal, para la óptima utilización del mercado de datos en la ONEI.
- Darle continuidad al trabajo a partir de la aplicación de técnicas de minería de datos, para potenciar la investigación en cuanto al descubrimiento de patrones e información relevante para la toma de decisiones.
- Elaborar una guía para la migración de las herramientas utilizadas, hacia nuevas versiones, cuando sea requerido.
- Potenciar el uso de semáforos o celdas de colores durante la visualización, para mejorar el proceso de toma de decisiones.

### REFERENCIAS BIBLIOGRÁFICAS

- 1- **Coreas, Lic Edwin Osmil.** GestioPolis. [Online] 2006. [Cited: noviembre 2011, 2011.] <http://www.rena.edu.ve/cuartaEtapa/Informática/Tema10>.
- 2- **R.Rivadiera, Gustavo.** Scribd. [Online] 2010. <http://www.scribd.com/doc/9219011/Kim-Ball>.
- 3- **Cabrera, María Evelia Casales.** Data Warehouse(Almacenes de Datos). [Online] [Cited: diciembre 20, 2011.] <http://hp.fciencias.unam.mx/alg/bd/dwh.pdf>.
- 4- **ORALLO, J.H.** *Análisis y extracción de Conocimiento en Sistemas de Información Datawarehouse y Datamining.* 2003.
- 5- **Sinnexus.** [Online] 2008. [Cited: diciembre 20, 2011.] [http://www.sinenexus.com/business\\_intelligence/datawarehouse.aspx](http://www.sinenexus.com/business_intelligence/datawarehouse.aspx).
- 6- **Duarte, Elizabeth Vidal.** *Duarte, Elizabeth Vidal. Trabajo de investigación.* 2007.
- 7- **MIS RESPUESTAS.COM.** [Online] 2006. [Cited: diciembre 23, 2011.] <http://www.misrespuestas.com/que-es-una-metodologia.html>.
- 8- **Warehouse, W H.INMON. [aut. libro] John Wiley & Sons. Building the Data. Building the Data Warehouse.** 2005.
- 9- **Kimball, Ralph.** *The Data Warehouse ETL Toolkit.* s.l. s.l. : Segunda Edicion. 2002.
- 10- **Hernández, Yanisbel González.** *Propuesta de metodología para el desarrollo de almacenes de datos en DATEC. [Documento] La Habana : s.n.*
- 11- **Nader, Ing.Javier.** "Sistema de apoyo gerencial universitario". Tesis de Magister en Ingeniería de Software. . 2003.
- 12- **Prix Data.** [Online] 2006. [Cited: diciembre 22, 2011.] <http://www.dataprix.com/blogs/respinosamilla/cubos-olap->
- 13- **Visual Paradigm.** [Online] [Cited: diciembre 23, 2011.]
- 14- **Ibiblio,** The Public's Library and Digital Archive. [Online] [Cited: diciembre 23, 2011.] <http://www.ibiblio.org/pub/linux/docs/LuCaS/Postgresql-es/web/navegable/todopostgresql/app-pgadmin.html...>
- 15- **Pentaho Open Source Business Intelligence: Kettle.Project.** [Online] 1 25, 2005. <http://kettle.pentaho.org..> ..
- 16- **DataCleaner.** [Online] <http://datacleaner.eobjects.org/> ...

- 17- **Pentaho Shema Workbench**. [Online] [Cited: diciembre 26, 2011.]  
<http://churriwifi.wordpress.com/2010/07/04/17-3-preparando-el-analisis-dimensional-definicion-de-cubos-utilizando-schema-workbench/>. ...
- 18- **Pentaho BI Suite Enterprise Edition**. [Online] 2005. [Cited: diciembre 26, 2011.]  
[http://www.pentaho.com/products/data\\_services/?hp=y...](http://www.pentaho.com/products/data_services/?hp=y...)
- 19- **Ing.Lilliam Vega Torres, Ing.Luis Rojas Díaz,Lic.Cecilia Placeres Villar**. Ciudad de la Habana, Cuba : s.n. : s.n., Noviembre 2008.
- 20- **Apache TomCat**. [Online] [Cited: diciembre 27, 2011.] <http://tomcat.apache.org/...>
- 21- **Ponniah, P.** (2001). Data Warehousing Fundamentals. E.U.A: Wiley Publishing Inc.
- 22- **Ralph Kimball y otros**. The Data Warehouse Lifecycle Toolkit. E.U.A: Wiley Publishing Inc
- 23- **Ralph Kimball y Margy Ross**. (2002). The Data Warehouse Toolkit. E.U.A: Wiley Publishing Inc.
- 24- **Curto ,J** . Blllage. [En línea]2009,[Citado: abril 25, 2012 julio 17]  
<http://bi.social.uoc.edu/smc/blog/34-subsistemas-etl-de-kimball>
- 25- **Lamancha, Beatriz Pérez**. [Online] 2006. [Cited: abril 10, 2012.]  
<http://www.fing.edu.uy/~bperez/.../Tesis%20-%20Beatriz%20Perez-%202006.pdf>..
- 26- **Vera, Martha E Rojas**. PMO.www.spanishpmo.com. [Online] julio 27, 2010. [Cited: abril 12, 2012.]  
<http://spanishpmo.com/index.php/ciclos-de-vida-modelo-en-v>.
- 27- **García, Fabian López**. Sistema de Información de Gobierno.Mercado de datos para las áreas de Cultura y Deporte, Trabajo de diplomapara optar por el título de Ingeniero en Ciencias Informáticas. 2011.
- 28- **Delmis Arianna Pérez Leyva, Maryin García Devesa**. Sistema de Información de Gobierno. Mercado de datos Comercio Interior, Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas. 2011.



## BIBLIOGRAFÍA

1. **Apache TomCat** [En línea]. - 27 de diciembre de 2011. - <http://tomcat.apache.org/...>
2. **C.C.M Evelia** Data Warehouse(Almacenes de Datos) [En línea]. - 20 de diciembre de 2011. - <http://hp.fciencias.unam.mx/alg/bd/dwh.pdf>.
3. **Coreas Lic Edwin Osmil** GestioPolis [En línea]. - 2006. - 2011 de noviembre de 2011. - <http://www.rena.edu.ve/cuartaEtapa/Informática/Tema10>.
4. **Cuervo M. C. y O. Y. B. Moreno.** Herramientas libres para modelar software. [En línea]. - 2005.
5. **Curto J.** Data Warehousing, Data Warehouse y Datamart. [En línea]. - 2007.
6. DataCleaner [En línea]. - <http://datacleaner.eobjects.org/> ...
7. **Curto ,J .** Blllage. [En línea]2009,[Citado: abril 25, 2012 julio 17] <http://bi.social.uoc.edu/smc/blog/34-subsistemas-etl-de-kimball>
8. **Delmis Arianna Pérez Leyva Maryin García Devesa** Sistema de Información de Gobierno. Mercado de datos Comercio Interior, Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas [Sección del libro]. - 2011.
9. **Duarte Elizabeth Vidal** Duarte, Elizabeth Vidal. Trabajo de investigación [Informe]. - 2007.
10. etl-tools [En línea]. - [http://etl-tools.info/es/bi/almacenedatos\\_esquema-constelacion.htm..](http://etl-tools.info/es/bi/almacenedatos_esquema-constelacion.htm..)
11. **García, Fabian López.** Sistema de Información de Gobierno.Mercado de datos para las áreas de Cultura y Deporte, Trabajo de diplomapara optar por el título de Ingeniero en Ciencias Informáticas. 2011.
12. **Gerencia e Innovacion** [En línea]. - 27 de septiembre de 2009. - <http://gerenciainnovacion.blogspot.com/2008/09/conceptos-inteligencia-de-negocios.html..>
13. **Ing.Lilliam Vega Torres Ing.Luis Rojas Díaz,Lic.Cecilia Placeres Villar.** [Informe]. - Ciudad de la Habana, Cuba : s.n. : [s.n.], Noviembre 2008..
14. **Hernández, Yanisbel González.** *Propuesta de metodología para el desarrollo de almacenes de datos en DATEC.* [Documento] La Habana : s.n.
15. **Kimball Ralph y Ross, Margie.** The Data Warehouse Toolkit: John Wiley and Sons, Inc [Sección del libro]. - [s.l.] : ISBN 0-471-20024-7., 2002. - 2da Edición. New York .
16. **Kimball Ralph.** The Data Warehouse ETL Toolkit. s.l. s.l. : Segunda Edicion [Sección del libro]. - 2002.

17. **Lamancha Beatriz Pérez** [En línea] / prod. Informática. Proceso de Testing funcional independientemente. Tesis de Maestría en. - 2006. - 10 de abril de 2012. - <http://www.fing.edu.uy/~bperez/.../Tesis%20-%20Beatriz%20Perez-%202006.pdf>..
18. **Menéndez Dr. Eugenio Santos.** O.E.I [En línea]. - [http://www.oei.eui.upm.es/Asignaturas/BD/DYOBBD/Ejemplo\\_DW.pdf](http://www.oei.eui.upm.es/Asignaturas/BD/DYOBBD/Ejemplo_DW.pdf)...
19. **MIS RESPUESTAS.COM** [En línea]. - 2006. - 23 de diciembre de 2011. - <http://www.misrespuestas.com/que-es-una-metodologia.html>.
20. **Nader Ing. Javier** Scibd [En línea] = Tesis de Magister en Ingeniería del software."Sistema de Apoyo Gerencial Universitario". - 9 de mayo de 2011. - 5 de marzo de 2012. - <http://www.scribd.com/doc/63954285/30/Definicion-de-Inteligencia-de-Negocios>.
21. **Nader Ing. Javier** "Sistema de apoyo gerencial universitario". Tesis de Magister en Ingeniería de Software. [Sección del libro]. - 2003.
22. **ORALLO J.H.** Análisis y extracción de Conocimiento en Sistemas de Información Datawarehouse y Datamining [Informe]. - 2003.
23. **Pentaho Open Source Business Intelligence: Kettle.Project** [En línea]. - 25 de 1 de 2005. - <http://kettle.pentaho.org..> ..
24. **ibiblio, The Public's Library and Digital Archive** [En línea]. - 23 de diciembre de 2011. - <http://www.ibiblio.org/pub/linux/docs/LuCaS/Postgresql-es/web/navegable/todopostgresql/app-pgadmin.html>...
25. **Pentaho BI Suite Enterprise Edition.** [En línea]. - 2005. - 26 de diciembre de 2011. - [http://www.pentaho.com/products/data\\_services/?hp=y...](http://www.pentaho.com/products/data_services/?hp=y...)
26. **Pentaho Shema Workbench** [En línea]. - 26 de diciembre de 2011. - <http://churriwifi.wordpress.com/2010/07/04/17-3-preparando-el-analisis-dimensional-definicion-de-cubos-utilizando-schema-workbench/>. ..
27. **Prix Data** [En línea]. - 2006. - 22 de diciembre de 2011. - <http://www.dataprix.com/blogs/respinosamilla/cubos-olap->.
28. **R.Rivadera Gustavo** Scibd [En línea] / prod. warehouses) La metodología de Kimball para el diseño de almacenes de datos (Data. - 2010. - <http://www.scribd.com/doc/9219011/Kim-Ball>.
29. **Schmidt K.Ferreira y J.** Sistemas de Información. Pentaho. [En línea]. - 2009.

30. Sinnexus [En línea]. - 2008. - 20 de diciembre de 2011. - [http://www.sinenexus.com/business\\_intelligence/datawarehouse.aspx](http://www.sinenexus.com/business_intelligence/datawarehouse.aspx).
31. **Torres L. P.** Curso Almacenes de Datos: importancia del estándar. [Sección del libro]. - 2006.
32. **Vera Martha E Rojas** PMO. [www.spanishpmo.com](http://www.spanishpmo.com) [En línea]. - 27 de julio de 2010. - 12 de abril de 2012. - <http://spanishpmo.com/index.php/ciclos-de-vida-modelo-en-v>.
33. Visual Paradigm [En línea]. - 23 de diciembre de 2011.
34. **Warehouse W H.INMON.** [aut. libro] **John Wiley & Sons.** **Building the Data** Building the Data Warehouse. [Libro]. - 2005.



Standard measures

	14.3 - Valor de la circula...	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]		
Row count	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17		
Null values	8	11	11	10	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11		
Empty values	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
Highest value	Venta promedio diaria	UM	9133,4	9315,5	86,99	96,01	87,82	9959,9	924	933	980	6,7	8,5	7,3	6,4	6,4	7,2	6,1	6,1	79,4	91,21	91,99	94,98	67,74
Lowest value	De 1995 al 1997 con ...	%	-	-2	-2,3	-3,6	1	1	1	1,4	1042	110,31	104,94	103,6	106,78	111,25	100,02	106	11,7	2007	2008	2009	106,02	

Análisis de las medidas

String analysis

	14.3 - Valor de la circula...	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	[Colu...]	
Char count	250	11	21	27	24	26	26	24	25	22	27	28	28	28	27	26	26	26	25	29	27	27	33
Max chars	76	3	6	6	6	6	6	6	7	7	7	7	7	7	7	6	7	7	7	7	7	7	8
Min chars	5	1	1	1	2	3	3	1	1	1	3	3	3	3	3	3	2	2	3	4	3	3	3
Avg chars	27,78	1,83	3,5	3,86	4	4,33	4,33	4	4,17	3,67	4,5	4,67	4,67	4,67	4,5	4,33	4,33	4,33	4,17	4,83	4,5	4,5	5,17
Max white spaces	14	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Min white spaces	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Avg white spaces	5,56	0	0,14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Uppercase chars	6%	81%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Lowercase chars	56%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Non-letter chars	37%	18%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Word count	45	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
Max words	15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Min words	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Análisis de los string

Fuente 14.3: Valor de la circulación mercantil minorista de bienes.

Standard measures

	14.6 - Valor de la circulación mercantil minorista de bienes por provincias	[Column 2]	[Column 3]	[Column 4]
Row count	111	111	111	111
Null values	23	17	22	21
Empty values	0	0	0	0
Highest value	CONCEPTOS	Circulación	964,6	Millones de pesos
Lowest value	Camagüey	Mercantil	Comercio	Pública

Análisis de las medidas

String analysis

	14.6 - Valor de la circulación mercantil minorista de bienes por provincias	[Column 2]	[Column 3]	[Column 4]
Char count	1053	578	490	513
Max chars	23	11	10	17
Min chars	4	3	3	3
Avg chars	11,97	6,15	5,51	5,7
Max white spaces	7	1	2	2
Min white spaces	0	0	0	0
Avg white spaces	3,45	0,02	0,03	0,04
Uppercase chars	10%	0%	0%	0%
Lowercase chars	50%	4%	3%	6%
Non-letter chars	39%	94%	96%	93%
Word count	152	94	89	92
Max words	4	1	1	3
Min words	1	1	1	1

Análisis de los string

Fuente 14.6: Valor de la circulación mercantil minorista de bienes por provincias.

Standard measures																				
	14.9 - Servicios comerciales pres...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...
Row count	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45
Null values	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	15
Empty values	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Highest value	Total	96,2	8816,9	770,9	8162,4	9833,2	989,3	94955,5	9577,3	912,4	9665,9	9134,6	87617,7	84,4	87	9316,6	97990,5	969,9	8964,6	
Lowest value	Aire Acondicionado	1101,8	-	-	-	-	-	-	-	-	-	-	1130,7	10,6	11,3	-	1168,5	10,2	10333,...	Miles...

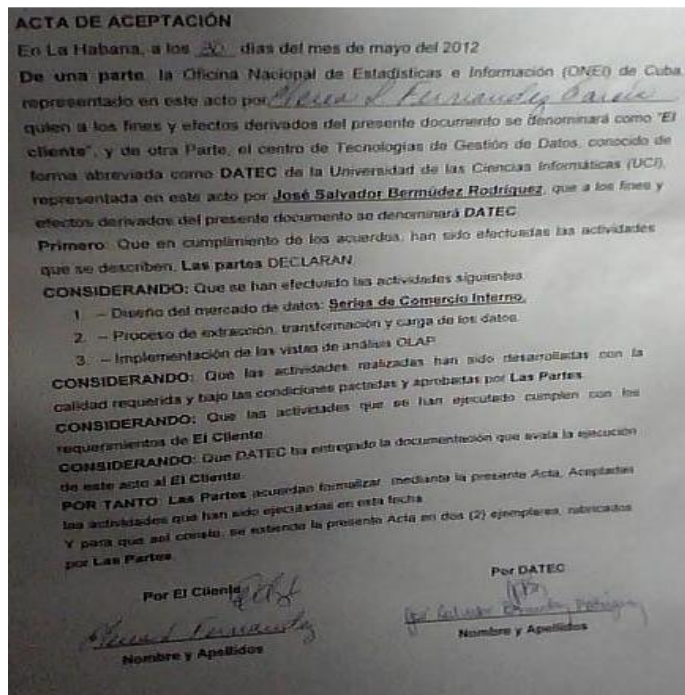
Análisis de las medidas

String analysis																				
	14.9 - Servicios comerciales pre...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...	[Colu...
Char count	638	165	160	164	162	162	160	159	157	171	248	166	176	165	202	168	163	168	167	
Max chars	35	7	7	8	7	8	7	7	7	16	76	8	8	8	36	8	7	8	20	
Min chars	5	1	1	1	1	1	1	1	1	1	1	2	2	3	2	2	2	3	2	
Avg chars	20,58	5,32	5,16	5,29	5,23	5,23	5,16	5,13	5,06	5,52	7,75	5,35	5,68	5,32	6,31	5,42	5,26	5,42	5,57	
Max white spaces	18	0	0	0	0	0	0	0	0	13	76	0	0	0	36	0	0	0	4	
Min white spaces	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Avg white spaces	5,52	0	0	0	0	0	0	0	0	0,42	2,78	0	0	0	1,12	0	0	0	0,13	
Uppercase chars	6%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
Lowercase chars	67%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	8%	
Non-letter chars	26%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	90%	
Word count	60	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	32	
Max words	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	
Min words	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	

Análisis de los string

Fuente 14.9: Servicios comerciales prestados a la población.

Anexo 2: Carta de aceptación



## GLOSARIO

**ONEI:** Oficina Nacional de Estadísticas e Información.

**SIEN:** Sistema de Información Estadístico Nacional.

**SIGOB:** Sistema de Información de Gobierno.

**Data Warehouse (DWH):** Almacén de datos.

**AD:** Almacén de datos.

**Data Mart:** Mercado de datos.

**MD:** Mercado de datos.

**DATEC:** Centro de Tecnologías de Gestión de Datos.

**UML:** Lenguaje de Modelado Unificado. Es el lenguaje visual que permite especificar, construir y documentar un software.

**ETL:** Extracción, transformación y carga.

**BI:** Inteligencia del negocio.

**XML:** Estándar de información cuyas siglas vienen dadas por su nombre en inglés eXtensible Markup Language.

**MDX:** Es el lenguaje que se usa para realizarle consultas a los cubos OLAP. Las siglas MDX son el acrónimo de MultiDimensional expressions, (en español) expresiones multidimensionales.

**OLAP:** Procesamiento Analítico en Línea.

**ROLAP:** Procesamiento Analítico en Línea Relacional.

**HOLAP:** Procesamiento Analítico en Línea Híbrido.

**MOLAP:** Procesamiento Analítico en Línea Multidimensional.

**RN:** Reglas de negocio.

**RI:** Requisitos de información.

**RF:** Requisitos funcionales.

**SQL:** Lenguaje de Consulta Estructurado.

**Jerarquía:** Es la organización de la información por niveles dentro de una dimensión.

**URL:** Localizador Uniforme de Recursos

**Cubo:** Es las colección de dimensiones y medidas en un área temática particular.

**Perfilado de Datos:** Proceso de examen de los datos disponibles en una fuente de datos que facilita la recogida de estadísticas e información acerca de los datos.

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

**No conformidad:** Se entiende como no conformidad a aquellos errores que son detectados por el equipo de desarrollo durante la evaluación del producto.