

Universidad de las Ciencias Informáticas
Facultad 6



Título: “Sistema de Información de Gobierno. Mercado de datos Comercio Interior”

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

Autoras:

Delmis Arianna Pérez Leyva

Maryin García Devesa

Tutores:

Ing. Yanisbel González Hernández

Msc. Asnioby Hernández López

Ciudad de La Habana, Junio de 2011
“Año 53 de la Revolución Cubana”



*“La inteligencia da bondad, justicia y hermosura;
Como un ala, levanta el espíritu;
Como una corona, hace monarca al que la ostenta”*

José Martí

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

Autoras: _____

Maryin García Devesa

Delmis Arianna Pérez Leyva

Tutores: _____

MSc. Asnioby Hernández López

Ing. Yanisbel González Hernández

DATOS DE CONTACTO

Msc. Asnioby Hernández López: Graduado de Ingeniería en Ciencias Informáticas en la Universidad de las Ciencias Informáticas (UCI) en el año 2007. Graduado de Máster en Informática Aplicada en junio del 2009. Actualmente se desempeña como J' Dpto. de Almacenes de Datos. DATEC. Facultad 6.

Ing. Yanisbel González Hernández: Graduada de Ingeniera en Ciencias Informáticas en el 2008 en la Universidad de Ciencias Informáticas (UCI). Analista de Sistemas del Dpto. de Almacenes de Datos. DATEC. Facultad 6. Fundadora de la Universidad de Ciencias Informáticas. Ha tutorado 4 tesis de pregrado con resultados satisfactorios. Se encuentra matriculada en la maestría de Gestión de Proyectos Informáticos, donde ha cursado el 100% de los cursos, está dedicado a la realización de su trabajo de diploma final.

Le agradezco a todas las persona que siempre han estado conmigo en todo momento, a mi mamá que es el amor de mi vida, mi razón de ser ya que gracias a ella he llegado a ser quien soy, me ha enseñado las cosas buenas y malas de la vida, a saberme levantar cuando he caído y seguir adelante.

A mis hermanos por la gran responsabilidad que tengo de ser mejor para guiarlos y vean en mi un gran ejemplo para ellos.

A mi papa por ser mi gran ejemplo en la vida como persona y motivarme a forjarme como una profesional y por todo el amor que me ha dado.

A mi 2do papá, que me ha ayudado y apoyado en todos los momentos importantes de mi vida; por darme mucho amor y cariño como a una hija.

A mis abuelos por brindarme su amor y por estar conmigo desde los primeros pasos de mi vida.

A mi tío el negro que aunque hoy no se encuentra físicamente en este momento tan importante para mí, se que se sentiría muy orgulloso y le agradezco por toda la ayuda que siempre me brindó.

A mi tía Nery por ser una gran amiga para mí y ayudarme siempre que la he necesitado.

A mi novio por soportar todas mis malcriadeces y por todo el apoyo y amor que me ha dado.

A todas mis amistades con las que he compartido esta etapa tan maravillosa de la vida.

A dos personitas traviesas que son el ejemplo de que no importa en qué momento de la vida de alguien llegues, para ser importante Lay y Susy que han sido capaz de soportarme y quererme como soy, a Lisi que también la he aprendido a querer mucho en el poco tiempo que la conozco.

A Fide que más que todo ha sido mi amigo, que me ha apoyado y ayudado durante estos 5 años.

A mi familia en general y a todas las personas que de una forma u otra han aportado su granito de arena para que hoy este aquí.

A mis tutores que se han portado muy bien y me han sabido guiar en todo el desarrollo de mi trabajo.

A todos los profesores que me dieron su apoyo durante mi carrera.

Maryin García Devesa

A mis padres y mi hermano. Ana Lourdes, Jorge y Liobis, por todos estos años de amor, paciencia y comprensión, por educarme tan bien e inculcarme el amor al estudio, a ser una profesional. A ustedes les agradezco todo lo que soy... Esto es una pequeña parte del fruto de la semillita que una vez sembraron.

A mi familia por todo el cariño que me dan, porque son las personas más especiales que he conocido en la vida. Por esperar siempre lo mejor de mí y confiar en mis decisiones aunque no siempre fueran las mejores.

A mis tutores Yanisbel y Asnioby por su paciencia y esfuerzo de estar siempre dispuestos a ayudarme.

A mi abuelita Petra que aunque no está físicamente conmigo, siempre estará en mi corazón, se que estaría orgullosa de mi.

A mi novio por su cariño y comprensión.

A todos los profesores que contribuyeron a mi formación como ingeniera y me ayudaron en los momentos más difíciles a seguir adelante especialmente a Yoendris, Delvis, Ponce, Yetel.

A todos mis amigos, los de mi casa y los de la universidad, les agradezco los momentos juntos que me permitieron aprender de cada uno y quererlos como son...

Delmis Arianna Pérez Leyva

A mi mamá por ser la persona más importante de mi vida, a mis hermanos que son las personitas que me dan la inspiración para ser mejor cada día, a mi papá y a Chande por ser mis grades ejemplos a seguir en la vida, a mi ahijada para que un día se sienta orgullosa de mi y a mi novio que me ha apoyado en todos los momentos difíciles.

Maryin García Devesa

A mis padres, por estar con los brazos abiertos siempre que los necesito, por todas sus horas de sacrificio, amor, por luchar conmigo en los buenos y malos momentos. Por la confianza que depositaron en mi cuando creyeron que esta no era la carrera correcta. Por todo el cariño que me regalan, por guiarme en la vida, por todos sus esfuerzos para lograr mi desarrollo profesional. Sé que están muy orgullosos de mi... Los quiero mucho...

Delmis Arianna Pérez Leyva

El presente trabajo de diploma enmarca su desarrollo en los mercados de datos, para realizar el análisis de información perteneciente al Comercio Interior en la Oficina Nacional de Estadísticas (ONE). Para guiar este proceso de desarrollo se utilizó la metodología propuesta por el Centro Tecnología de Gestión de Datos (DATEC). Como resultados se muestran las estructuras dimensionales para el modelo propuesto que comprende las jerarquías, dimensiones, tablas de hechos y medidas necesarias para garantizar una buena solución del mercado de datos. Se definen e implementan los mecanismos de extracción, transformación y carga de los datos correspondientes al modelo, además del desarrollo de la capa de Inteligencia del Negocio (BI) para la visualización y análisis de toda la información contenida en el mercado de datos. Se realizaron pruebas a la solución para adquirir la aceptación del cliente y verificar su calidad.

Palabras Claves: mercado de dato, estructura dimensional, ONE, BI, DATEC, Comercio Interior.

Tabla de Contenidos

Introducción	1
CAPÍTULO I: FUNDAMENTACIÓN TEÓRICA	5
1.1. Comercio Interior	5
1.2. Aplicaciones existentes para el Comercio Interior	6
1.2.1. Algunas de las aplicaciones existentes en el mundo.....	6
1.2.2. Algunas de las aplicaciones existentes en Cuba.....	6
1.3. Almacenes de Datos	6
1.3.1. Definición de almacenes de Datos.....	6
1.3.2. Características de un almacén de datos.....	7
1.3.3. Componentes de un almacén de datos	8
1.3.4. Ventajas de un Almacén de Datos.....	9
1.3.5. Desventajas de un Almacén de Datos	10
1.3.6. Modelo Multidimensional.....	10
1.3.7. Modo de Almacenamiento de Datos OLAP.....	11
1.3.8. Mercado de Datos	13
1.4. Metodologías	14
1.4.1. Metodologías para el desarrollo del Almacén de Datos (AD).....	14
1.5. Herramientas	15
1.5.1. Herramientas del proceso de ingeniería.....	15
Conclusiones	18
CAPITULO II: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS	19
2.1. Análisis	19
2.1.1. Descripción del negocio	19
2.1.2. Clasificadores estadísticos	21
2.1.3. Tema de análisis identificado	22
2.1.4. Reglas del negocio.....	22
2.1.5. Necesidades de los usuarios	22
2.1.6. Levantamiento de requisitos	23
2.1.7. Reportes candidatos	27
2.1.8. Casos de Uso del Sistema (CUS)	28
2.1.9. Perfilado de datos	32
2.2. Diseño	32

2.2.1.	Matriz bus	32
2.2.2.	Modelo de datos	33
2.2.3.	Mapa de navegación	38
2.2.4.	Diseño de los cubos OLAP	39
2.2.5.	Diseño de los procesos de integración de datos.....	42
2.2.6.	Política de respaldo y recuperación	43
	Conclusiones	44
	CAPÍTULO III: IMPLEMENTACIÓN DE LA SOLUCIÓN	45
3.1.	Implementación de la Base de Datos.....	45
3.1.1.	Estructura de los datos.....	45
3.1.2.	Usuarios y privilegios en la Base de Datos.....	46
3.2.	Implementación del subsistema de integración de datos (ETL)	46
3.3.	Implementación del subsistema de visualización de datos(BI)	47
3.3.1.	Implementación de los reportes candidatos	48
3.3.2.	Seguridad en la aplicación	48
	Conclusiones	49
	CAPÍTULO IV: VALIDACIÓN Y PRUEBAS DEL MERCADO DE DATOS	50
4.1.	Calidad de software	50
4.1.1.	Importancia de la calidad del software	50
4.1.2.	Cualidades que debe tener un producto de software para que tenga calidad.....	50
4.2.	Pruebas de calidad de software	51
4.2.1.	Modelo V	51
4.2.2.	Tipos de pruebas en el Modelo V	52
4.2.3.	Herramientas de prueba	53
4.3.	Resultados de las pruebas	54
	Conclusiones	55
	Conclusiones generales.....	56
	Recomendaciones	57
	Bibliografía	58
	Referencias Bibliográficas.....	61
	Anexos	64
	Glosario	67

Índice de Figuras

Figura 1: Relaciones entre un componente de un Almacén de Datos	9
Figura 2: Estructura de un Cubo OLAP	10
Figura 3: Arquitectura del sistema ROLAP	13
Figura 4: Flujo de información	21
Figura 5: Diagrama de Casos de Uso del Sistema	28
Figura 6: Modelo de Datos.....	34
Figura 7: Hecho Ingresos por ventas	38
Figura 8: Mapa de navegación	39
Figura 9: Cubo Ingresos por ventas y servicios a la población	40
Figura 10: Cubo Ingresos por ventas y servicios a la población por tipo de mercado	40
Figura 11: Cubo Series mercantil, establecimientos y alimentación	41
Figura 12: Cubo Servicios comerciales.....	41
Figura 13: Diseño de los procesos de integración de datos	42
Figura 14: Diseño de la transformación de la dimensión dpa	43
Figura 15: Transformación del hecho ingreso por venta y servicios a la población	47
Figura 16: Reporte Ingresos por ventas.....	48
Figura 17: Rol y Usuario de Administración	49
Figura 18: Permisos asignados al usuario “admin”	49
Figura 19: Rol y Usuario de Análisis	49
Figura 20: Permisos asignados al usuario “analista”	49
Figura 21: Modelo V	52
Figura 22: Comportamiento de los Indicadores	54

Índice de Tablas

Tabla 1: Descripción de los actores y CUS	29
Tabla 2: Descripción del CU Consultar ingresos	30
Tabla 3: Dimensiones.....	32
Tabla 4: Hechos	33
Tabla 5: Matriz bus.....	33
Tabla 6: Caso de prueba Consultar ingresos por tipos de mercado	64
Tabla 7: Lista de chequeo	65

Introducción

El comercio es la actividad socioeconómica, que consiste en la prestación del servicio de distribución, compra y venta de bienes tanto, a nivel nacional como internacional. El comercio se divide en dos ramas fundamentales, el Comercio Exterior el cual genera las actividades de exportación e importación entre los países y el Comercio Interior que es la actividad que realizan varias personas o empresas dentro de un mismo país.

Cuba, país subdesarrollado, presenta una economía planificada en la cual el Comercio Interior juega un papel importante mediante el intercambio existente entre los diferentes organismos y entidad estatales.

La Oficina Nacional de Estadística (ONE), órgano rector de la estadística en Cuba, tiene como propósito obtener información estadística de calidad adquirida por el Sistema de Información Estadístico Nacional (SIEN). La ONE cuenta con una serie de modelos estadísticos en los cuales se recoge información de todos los sectores de la economía y la sociedad, entre los que se encuentra el Comercio Interior. La información relacionada con este sector, se recoge en diferentes modelos por las entidades existentes en el territorio nacional, posteriormente son enviados a las entidades municipales de la ONE correspondientes a cada territorio y luego se envían a las sedes provinciales. Por último, la información es digitalizada en ficheros con formato 'dbf' y remitidos a la oficina central que se sitúa en Ciudad Habana.

La Universidad de las Ciencias Informáticas (UCI), además de ser una institución educacional, es un centro productor de software; posee varios centros de desarrollo, como el Centro de Tecnologías y Gestión de Datos (DATEC) que en colaboración con la Oficina Nacional de Estadística (ONE), deciden realizar el proyecto Sistema de Información de Gobierno (SIGob), con el objetivo de mejorar el proceso de toma de decisiones en el SIEN.

Anteriormente se desarrolló el trabajo de Diploma titulado "Sistema de Información de Gobierno. Mercado de Datos Comercio", el cual tuvo como alcance el análisis y diseño del mercado de datos culminando con resultados satisfactorios. En la ONE después de finalizado el desarrollo de dicho trabajo se efectuaron cambios en algunos nomencladores como son la División Política Administrativa (DPA) y el Clasificador de Actividades Económicas (CAE). Además las nomenclaturas utilizadas en el diseño de dicha solución en el nombrado de los hechos y dimensiones no cumplen con los estándares de codificación definidos por el centro para lograr una adecuada integración de los datos. Por tal

motivo se decide llevar a cabo un rediseño del mercado de datos, además de realizar los procesos de Extracción, Transformación y Carga (ETL) y la Capa de Visualización (BI).

Actualmente continúan las deficiencias que afectan la gestión de la información del área de Comercio Interior como la no existencia de una forma adecuada de almacenar, recuperar y presentar la información proveniente de estos organismos regionales, se dificulta el análisis de los principales reportes al presentar datos dispersos en ficheros ´dbf´ y ´excel´, la información no está integrada y es inconsistente, las fuentes solo pueden ser consultadas por un especialista de la informática y de la información con alto conocimiento del negocio, los ficheros son generados anualmente dificultando la obtención de la información, el proceso de recuperación y elaboración de informes es costoso en tiempo y esfuerzo; obstaculizando la disponibilidad de información sobre el Comercio Interior afectando el proceso de toma de decisiones.

Por lo anteriormente expuesto se plantea el siguiente **problema de la investigación**: ¿Cómo contribuir a la toma de decisiones en el área Comercio Interior del Sistema de Información de Gobierno?

Se define como **objeto de estudio** los almacenes de datos y como **campo de acción** el mercado de datos para el área Comercio Interior del Sistema de Información de Gobierno.

El **objetivo general** de la presente investigación es: Desarrollar el mercado de datos en el Comercio Interior del Sistema de Información de Gobierno que contribuya a la toma de decisiones.

Como objetivos específicos:

- Realizar análisis y diseño el mercado de datos del área de Comercio Interior.
- Implementar el mercado de datos del área de Comercio Interior.
- Validar el mercado de datos del área de Comercio Interior.

Para el cumplimiento de los objetivos de la investigación se plantean las **tareas de investigación** que a continuación se enumeran:

1. Caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.
2. Levantamiento de los requisitos.

3. Descripción de los casos de uso del mercado de datos.
4. Definición de los hechos, las medidas y las dimensiones del mercado de datos.
5. Diseño del modelo de datos.
6. Definir la arquitectura del mercado de datos.
7. Diseño del subsistema de integración.
8. Diseño del subsistema de visualización.
9. Diseño de los casos de pruebas.
10. Implementación del subsistema de integración.
11. Implementación del subsistema de visualización.
12. Aplicación de las listas de chequeo.
13. Aplicación de los casos de pruebas.

Estructura de la Tesis

El presente trabajo está estructurado de la siguiente forma: introducción, cuatro capítulos, conclusiones, recomendaciones, referencias bibliográficas, bibliografía, anexos y glosario de términos.

Capítulo 1: Fundamentación Teórica

En el presente capítulo se profundizará en la fundamentación teórica que soporta la investigación. Además, se realizará un estudio de las herramientas y metodologías de desarrollo utilizadas, así como de los procesos de integración de los datos e inteligencia de negocio que forman parte del desarrollo de un almacén de datos.

Capítulo 2: Análisis y diseño del mercado de datos

En este capítulo se realizará un estudio preliminar del negocio, la necesidades de información, las reglas del negocio, descripción de los casos de uso, identificación de dimensiones, hechos y medidas, desarrollo de la matriz BUS, modelo de datos, arquitectura de información, así como definir los reportes candidatos y el diseño de los procesos de integración.

Capítulo 3: Implementación de la solución

Se implementa el proceso ETL y la capa de inteligencia del negocio para el área de Comercio interior de la ONE, teniendo en cuenta las necesidades del cliente.

Capítulo 4: Validación y pruebas del mercado de datos

Se realizan pruebas para validar la implementación.

CAPÍTULO I: FUNDAMENTACIÓN TEÓRICA

Introducción

En el presente capítulo se profundizará en la fundamentación teórica que soporta la investigación. Además, se realizará un estudio de las herramientas y metodologías de desarrollo utilizadas, así como de los procesos de integración de los datos e inteligencia de negocio que forman parte del desarrollo de un almacén de datos.

1.1. Comercio Interior

El Comercio Interior es el negocio que se hace al vender, comprar o intercambiar géneros o productos para obtener beneficios dentro del mismo país, como es la fabricación de productos que luego se venden entre comerciantes, lo que es posible si existe un vendedor, que realiza la oferta; un comprador, que demanda un servicio y un mercado, lugar donde concurren.

Este comercio se descompone en local y nacional, según se refiera a las transacciones hechas entre los moradores de una localidad o de una nación. Se pueden realizar dos tipos de ventas:

- Al por mayor: se compran grandes cantidades (por lo que salen a un menor costo) y se venden a otras empresas.
- Al por menor: se venden directamente al público en pequeñas cantidades.

El Comercio Interior versa indistintamente sobre mercancías nacionales o extranjeras, y puede hacerse en grande o en pequeña escala. El comercio en gran escala, o al por mayor, hace pedidos considerables a las fábricas del país, recibe importantes cargamentos del extranjero, deposita el conjunto de las expediciones en sus almacenes y libera luego los géneros a los tenderos por pequeñas partidas. El pequeño comercio, o al por menor, conserva las mercancías en las tiendas para ir las vendiendo al público en porciones reducidas.

El comercio se caracteriza por la enorme variedad y cantidad de mercancías que mueve; realizando los intercambios comerciales que pueden compararse a “corrientes” o “flujos”. Diferenciándose los flujos de materias primas, el de productos elaborados y el de capitales e información. Para alcanzar un mejor nivel de organización en el Comercio Interior se han desarrollado varias aplicaciones.(1)(2)(3)

1.2. Aplicaciones existentes para el Comercio Interior

1.2.1. Algunas de las aplicaciones existentes en el mundo

Las aplicaciones del Comercio Interior se encuentran estrechamente relacionadas con las Tecnologías de la Información (TI) contando actualmente a nivel mundial con los Sistemas de Información (SI) automatizados que sirven para el registro de las transacciones diarias y la generación de reportes que presentan información con características de importancia, relevancia, claridad, sencillez y oportunidad de tal forma que sea útil para las personas a quienes se les entrega. Un ejemplo de estas aplicaciones son los registros de clientes y cuentas por cobrar, en la que se les actualiza la apertura de crédito a los clientes y se da seguimiento a sus abonos a la cuenta hasta su cancelación definitiva, generando movimientos en los datos financieros por lo que se utilizan también aplicaciones para caja chica y cuentas bancarias.(4)

1.2.2. Algunas de las aplicaciones existentes en Cuba

La Oficina Nacional de Estadísticas (ONE) conjuntamente con el Sistema de Información Estadística Nacional (SIEN) está llevando a cabo el desarrollo de mercados de datos con el objetivo de almacenar toda la información proveniente de todos los organismos regionales. En la actualidad no existe una aplicación estadística encargada de gestionar la información del Comercio Interior en el país, por lo que surge el presente trabajo de diploma. Para un mejor desarrollo de la investigación se realizó un estudio de la tecnología de almacenamiento de datos utilizada.

1.3. Almacenes de Datos

1.3.1. Definición de almacenes de Datos

Debido al gran cúmulo de información que es manejada actualmente tanto en los organismos como en las empresas del mundo, surge la necesidad de digitalizarla para facilitar una manipulación eficiente de la misma. Con el objetivo de dar solución a este problema surge el término almacén de datos (en inglés Datawarehouse, DWH).

Según el Dr. Eugenio Santos Menéndez “*Un almacén de datos es una colección de datos orientados al tema, integrados, no volátiles e historizados, organizados para el apoyo de un proceso de ayuda a la toma de decisiones*”.(5)

Por su parte Roberto Hernando Velazco lo define como “*Una vez reunidos los datos de los sistemas fuentes se guardan durante mucho tiempo, lo que permite el acceso a datos históricos; proporcionando*

al usuario una interfaz consolidada única para los datos, lo que facilita el trabajo a la hora de realizar las consultas en función de la toma de decisiones".(6)

En su libro "Using the Data Warehouse" William H. Inmon define a un almacén de datos como "*una colección de datos, orientados a hechos relevantes del negocio, integrados, que incluyen el tiempo como característica importante de referencia y no volátiles para el proceso de toma de decisiones".(7)*

Otro reconocido autor como Ralph Kimball, define a un almacén de datos como: "*...los Data Warehouse son una copia de los datos de la transacción estructurados específicamente para la pregunta y el análisis."* (8)

Como se ha planteado anteriormente son disímiles las definiciones dadas por autores y estudiosos de forma muy particular de lo que para ellos constituye un almacén de datos. En la presente investigación se asume como almacén de datos un sistema para la colección de datos con respecto a temas en específico, en los que la información va a prevalecer de una manera efectiva a lo largo del tiempo, y su principal razón de existir es la de brindar ayuda a una organización o empresa con respecto a la toma de decisiones.

1.3.2. Características de un almacén de datos

Temático

Los datos están almacenados por materias o temas. Estos se organizan desde la perspectiva del usuario final, mientras que en las bases de datos operacionales se organizan desde la perspectiva de la aplicación, con vistas a lograr una mayor eficiencia en el acceso a los datos.

Integrado

Todos los datos almacenados en el DWH están integrados, las bases de datos operacionales orientadas hacia las aplicaciones fueron creadas sin pensar en su integración, por lo que un mismo tipo de dato puede ser expresado de distintas maneras en dos bases de datos operacionales diferentes (Por ejemplo, para representar el sexo: 'Femenino' y 'Masculino', 'F' y 'M' o '0' y '1').

No volátil

En el almacén de datos únicamente hay dos tipos de operaciones: la carga de los datos procedentes de los entornos operacionales (carga inicial y carga periódica) y la consulta de los mismos. La actualización de datos no forma parte de la operativa normal de un almacén de datos, ya que no son

muy frecuentes los cambios en ellos y por ende se puede mantener por largos períodos de tiempo la información.

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. Mientras que en los almacenes de datos no son más que una serie de copias en el tiempo tomadas periódicamente, que permiten mantener y referenciar información.(9)

1.3.3. Componentes de un almacén de datos

Los almacenes de datos están integrados por una serie de procesos que en su conjunto definen el ambiente que estos poseen. Aunque cada desarrollo de un almacén de datos se diferencia en cuanto a la especificidad de las organizaciones, por lo general cumplen con la realización de los siguientes componentes:

Área de Procesamiento (staging)

Es el área que se encarga del almacenamiento de los datos temporalmente y de realizar un conjunto de procesos comúnmente llamados de Extracción, Transformación y Carga (ETL, por sus siglas en inglés). Realiza la función de interfaz entre las fuentes operacionales y el área de presentación. En este componente es donde se invierte la mayor cantidad de tiempo y esfuerzo durante el desarrollo del almacén. Al realizar el proceso de extracción de los datos de las distintas fuentes operacionales que se deseen integrar con el objetivo de almacenar toda la información en Base de Datos (BD) relacionales, generalmente utilizadas para realizar el análisis y procesamiento de los datos. Cuando los datos ya han sido almacenados en BD temporales se lleva a cabo su limpieza donde se detectan inconsistencias, duplicaciones, errores de formato e inexistencias, estandarizándose la información almacenada en diferentes fuentes. Estas transformaciones son las que sirven de sostén para realizar la carga de los datos, hacia el almacén de datos, en el área de presentación.

Área de Presentación

Es el componente donde los datos se encuentran organizados, almacenados y disponibles para ser consultados, reportados o analizados por parte de los usuarios finales. La información, diseñada mediante esquemas dimensionales, que ha sido definida por los usuarios como útil para la toma de

decisiones. Generalmente esta área es referenciada como una serie de mercados de datos integrados donde cada uno se encuentra representando a un proceso específico del negocio.

Herramientas de Acceso a Datos

En este componente es utilizada la palabra herramientas para referirse a la variedad de capacidades que pueden ser suministradas a los usuarios del negocio para dar soporte a la toma de decisiones, teniendo como actividad principal la posibilidad de consultar el área de presentación del almacén de datos.

En la figura 1 se muestran los componentes citados anteriormente y la relación existente en el diseño entre ellos:(10)

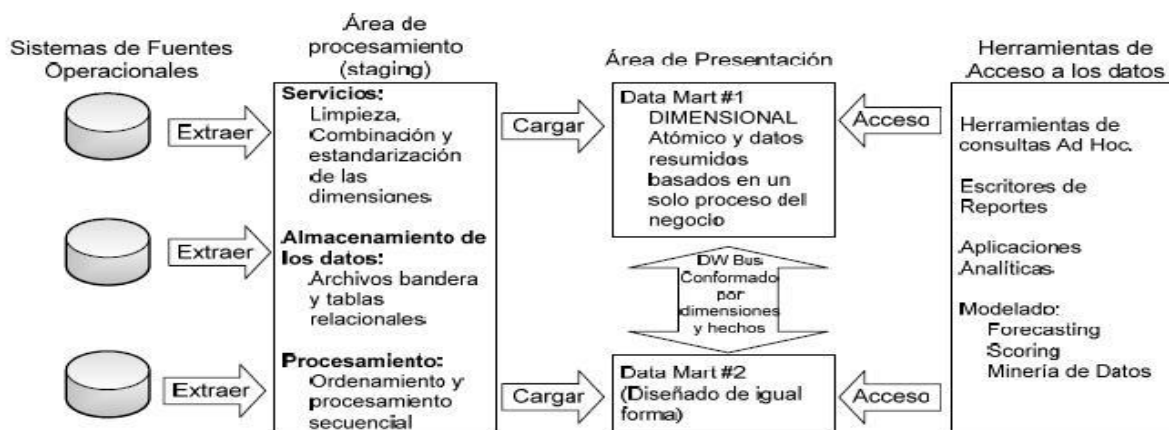


Figura 1: Relaciones entre un componente de un Almacén de Datos

1.3.4. Ventajas de un Almacén de Datos

- La inversión que realiza una organización para una correcta implantación de un sistema de almacén de datos conlleva un costo muy elevado, sin embargo el retorno de la inversión es garantizado en gran medida.
- Como consecuencia de la ventaja anterior se puede conseguir ventaja competitiva, debido a una buena toma de decisiones gracias al almacén de datos implantado.
- Mejoran la productividad de los responsables en la toma de decisiones de la organización debido a que:

- ✓ Los almacenes de datos hacen más fácil el acceso a una gran variedad de datos.
- ✓ Se obtiene una BD clasificada por temas e histórica.
- Integración de información procedente de múltiples sistemas eternos.(11)

1.3.5. Desventajas de un Almacén de Datos

- Supone altos gastos en las inversiones al desarrollar este tipo de proyecto.(11)

1.3.6. Modelo Multidimensional

El modelo multidimensional almacena la misma información que el Modelo Entidad Relación (MER) lo que 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. Algunas de sus principales ventajas son:

- Adición de dimensiones y hechos que no se habían previsto, sin que esto implique volver a cargar los datos ya almacenados.
- Agregar nuevos atributos a las dimensiones. Esta característica de gran adaptabilidad es muy deseable, pues a medida que los analistas añaden nuevos requerimientos al sistema, se pueden ir incorporando las modificaciones sin que esto implique demasiados cambios.

En la siguiente figura se observa el modelo multidimensional de la información (cubo OLAP), dividido en dos grandes grupos: las dimensiones y las tablas de hechos.

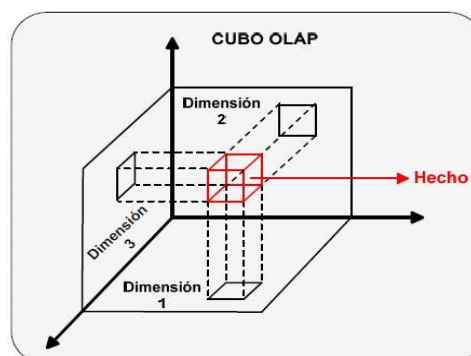


Figura 2: Estructura de un Cubo OLAP

Tablas de hechos

Los hechos son los indicadores del negocio que dan sentido al análisis de las dimensiones. Es la tabla central en un esquema dimensional, donde se almacenan las medidas numéricas del negocio, incluyendo los indicadores asociados a un proceso de negocio en concreto, además de las claves de las dimensiones que intervienen en dicho proceso. Estas medidas se hacen sobre el grano o unidad básica de la tabla, que se determina por el nivel de detalle que se almacenará en la misma.

Tablas de Dimensiones

Las dimensiones son los diferentes puntos de vista por los que queremos analizar la información. Estas incluyen los diferentes atributos que queremos analizar, que además se estructuran de forma jerárquica, conforme a diferentes niveles de detalle. Se construyen incluyendo todos los atributos de una forma desnormalizada, con una clave que identifica el mínimo nivel de detalle. La calidad de todo almacén de datos se mide por la definición de los atributos de las dimensiones. Su poder es directamente proporcional a la calidad y profundidad de estos atributos.(12)

Implementación del Modelo Multidimensional

El modelo multidimensional es modelado de acuerdo a la complejidad del sistema por tres variantes:

1. Esquema estrella: es la técnica más común, en este esquema existe un único elemento central (tabla de hechos) conectado radialmente con las tablas de dimensiones.
2. Esquema copo de nieve (Snowflake): es un esquema derivado de la estrella, donde las tablas de dimensiones se encuentran normalizadas.
3. Esquema constelación: está compuesto por una serie de esquemas de estrella o de copo de nieve, es decir, una tabla de hechos central con otras auxiliares y sus respectivas tablas de dimensiones.(13) (14) (15)

1.3.7. Modo de Almacenamiento de Datos OLAP

Se decidió utilizar la tecnología de Procesamiento Analítico en Línea (OLAP) ya que permite que los datos sean clasificados en diferentes dimensiones y pueden ser vistos unos con otros en diferentes combinaciones para obtener diferentes 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 les permiten formular consultas complejas, arreglar datos de un reporte, cambiar los datos resumidos o detallados etc. Las principales características de OLAP son:

- **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.(16)

Existen tres modelos de información, el Procesamiento Analítico Relacional en Línea (ROLAP, por sus siglas en inglés), el Procesamiento Analítico Multidimensional en Línea (MOLAP, por sus siglas en inglés) y el Procesamiento Analítico el Línea Híbrido (HOLAP, por sus siglas en inglés). En los cuales 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. Específicamente 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.

El sistema ROLAP utiliza una arquitectura de tres niveles:

1. El nivel de BD utiliza bases de datos relacionales para el manejo, acceso y obtención del dato.
2. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.
3. El motor ROLAP se integra con el nivel de presentación, a través de este nivel los usuarios realizan los análisis OLAP.



Figura 3: Arquitectura del sistema ROLAP

La arquitectura ROLAP es capaz de usar datos pre-calculados si estos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del DWH, y soporta técnicas de optimización de accesos para acelerar las consultas.(17)

1.3.8. Mercado de Datos

Es un subconjunto de datos de un almacén relativos a los requisitos de un departamento o área de negocio concreto. Este subconjunto de datos puede funcionar de forma autónoma, o bien enlazados al almacén de datos. El motivo por el cual se crean los mercados de datos es por el crecimiento que tiene el almacén, lo que facilita su construcción y utilización. Las características de los mercados de datos son:

- Se centran en los requisitos de los usuarios asociados a un departamento o área de negocio concretos.
- Como diferencia con los almacenes de datos, los mercados no contienen datos operacionales detallados.
- Son más sencillos a la hora de utilizarlos y comprender sus datos, debido a que la cantidad de información que contienen es mucho menor que en los almacenes de datos.(18)

1.4. Metodologías

- Las metodologías definen y guían el ciclo de vida del desarrollo de los almacenes de datos. Existen dos criterios bien identificados y que han marcado claramente su tendencia sirviéndole de guía a la comunidad mundial en cuanto a este tema. Estas tendencias son conocidas como Metodología de Kimball y Metodología de Inmon, en honor a sus creadores Ralph Kimball y William H. Inmon, personalidades reconocidas en el área de los almacenes de datos.
- Según Inmon propone primeramente construir el almacén de datos y a partir de este los mercados de datos, plantea la creación de un repositorio de datos corporativo como fuente de información consolidada, persistente, histórica y de calidad, basándose en un enfoque descendente (top-down).

Mientras que Kimball se basa en dividir el mundo de Inteligencia de Negocio (BI, por sus siglas en inglés) entre los hechos y las dimensiones, esta es eficaz y conduce a una solución completa en un corto período de tiempo, posee abundante documentación y se puede encontrar una respuesta a casi todas las preguntas que se puedan tener. Entre sus características principales es que su arquitectura es ascendente (bottom-up), plantea que se debe crear por cada departamento un conjunto de mercados de datos independientes orientados a los temas que estén relacionados con él.(22)

1.4.1. Metodologías para el desarrollo del Almacén de Datos (AD)

Se decidió utilizar el modelo propuesto para el desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio (BI) en la Línea de Almacenes de Datos de DATEC, tomando como base la metodología de Ralph Kimball y como complemento, para fortalecer la etapa de levantamiento de requisitos, orientando el trabajo a los Casos de Uso (CU) para así lograr alinearse a las tendencias y normas de la universidad. (23)

Durante el ciclo de vida de este modelo existen una serie de flujos de trabajo que tienen objetivos específicos. Estos se describen a continuación:

- **Estudio Preliminar o Planeación:** se realiza un estudio de lo qué se desea construir y qué condiciones existen para su desarrollo y montaje.
- **Requisitos:** se definen las reglas del negocio y necesidades de la información, además de un levantamiento detallado de cada una de las distintas fuentes de datos a integrar.

- **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 del repositorio de datos, se ejecutan las reglas de extracción, transformación y carga, haciendo los ajustes para integrar la información necesaria y se configuran e implementan las herramientas de BI para obtener los reportes, gráficos, mapas y otros que cubran los requerimientos.
- **Prueba:** se realizan varias pruebas comenzando por pruebas de unidad, pruebas de integración y sistema, hasta pruebas de aceptación con el cliente.
- **Despliegue:** se despliega la solución en el entorno real y en correcto funcionamiento.
- **Soporte y Mantenimiento:** son los servicios que se brindan ya una vez implantada la solución.
- **Gestión y Administración del Proyecto:** 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.

Con el fin de brindar soporte al diseño, construcción, implementación y mantenimiento del almacén de datos, se han utilizado diversas herramientas.

1.5. Herramientas

Las herramientas utilizadas para la solución fueron definidas por el departamento de almacenes de datos con el fin de lograr una integración entre todos los mercado de datos desarrollados para la ONE que en su conjunto forman un almacén de datos.

A continuación se explicarán las herramientas utilizadas por el centro con el fin de dar un acercamiento a la automatización del diseño, construcción, implementación y mantenimiento de los almacenes de datos.

1.5.1. Herramientas del proceso de ingeniería

Herramienta de modelado a utilizar

- **Visual Paradigm para UML 6.4:** 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 UML ayuda a una más rápida construcción de aplicaciones de calidad, y con un menor costo. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. Utiliza un lenguaje estándar, común, lo cual facilita a todo el equipo de desarrollo la comunicación, además de ser disponible en múltiples plataformas. (24)

Herramientas para la solución

Sistemas Gestores de Base de Datos (SGBD)

- **PostgreSQL 8.4.0:** es 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.

Características:

- Tienen la función de mantener la integridad y consistencia en la BD.
- Soporta operadores, funciones, métodos de acceso y tipos de datos definidos por el usuario.
- Soporta distintos tipos de datos: además del soporte para los tipos base, también soporta datos de tipo fecha, monetarios, elementos gráficos.
- Es multiplataforma, está disponible en las plataformas UNIX, Windows y Linux.
- Hace a los almacenes de datos flexibles, posee excelente rendimiento y escalabilidad.
- Incorpora funciones de diversa índole: manejo de fechas, geométricas, orientadas a operaciones con redes.(25)

PgadminIII 1.10: es una 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.(26)

Herramientas para ETL

- **Pentaho Data Integration 4.0.1:** 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.(27)
- **Data Cleaner 1.5.3:** 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 DHW, la investigación estadística, la preparación para el proceso de ETL y otras actividades.(28)

Herramientas para BI:

- **Pentaho Schema Workbench 3.2.0:** es una herramienta 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.(29)
- **Pentaho BI Server 3.6.0:** 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.(30)

- **Mondrian OLAP Server 3.0.4:** permite realizar consultas al DWH 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 brinda 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 MDX¹. 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.(31)
- **Apache Tomcat 5.5.5:** es un servidor web y de aplicaciones que gestiona solicitudes y respuestas Http (incluye el servidor Apache) gracias a sus conectores Http. Es un servidor de aplicaciones o contenedor de Servlets/JSP. 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.(32)

Conclusiones

En el presente capítulo se trataron los principales conceptos de almacén de datos y mercados de datos así como sus componentes, las metodologías y herramientas que se utilizarán.

Para el desarrollo de la presente investigación se decide utilizar el modelo para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio propuesta por DATEC la cual toma como base la metodología de Ralph Kimball, como Sistema Gestor de Base de Datos (SGBD) PostgreSQL y como herramienta de modelado Visual Paradigm. Para desarrollar el proceso de ETL se seleccionaron las herramientas Pentaho Data Integration y Data Cleaner, mientras que para el desarrollo de la capa de visualización el Pentaho BI Server, Pentaho Schema Workbench, Mondrian OLAP Server y Apache Tomcat.

¹ MDX (expresiones multidimensionales) para consultar datos multidimensionales o para crear expresiones MDX para su uso en un cubo.

CAPITULO II: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS

Introducción

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

2.1 Análisis

2.1.1 Descripción del negocio

La ONE, entidad rectora de la estadística en Cuba, es la encargada de gestionar y guardar la información de todo el país actuando como un almacén. Ella cuenta con una serie de modelos estadísticos en los cuales es recogida la información según su tipo. Estos modelos difieren entre sí en cuanto a la frecuencia que son recogidos (anual, trimestral, mensual), por los datos que son medidos y la estructura presente en cada una de las planillas donde es recogida la información. En los modelos 0758, 0768 y en las series (Valor de la circulación mercantil minorista, Ventas totales por tipo de establecimientos en el comercio minorista, Ventas por conceptos en la alimentación pública, Servicios comerciales prestados a la población y Servicios comerciales prestados a la población en valor) se recoge la información referente al comercio minorista. En estos modelos la información es recogida por la Oficina Municipal de Estadísticas de las provincias donde se ajusta el modelo con los datos que son remitidos por los centros informantes que pertenecen a un organismo y es enviado a la Oficina Provincial de Estadísticas. Posteriormente se realiza un resumen por municipios de esta información y es almacenada en un fichero con extensión 'dbf' y luego es colocada en un servidor, que posee una carpeta específica por cada provincia. En la ONE al llegar este fichero los especialistas realizan la limpieza de los datos y es creado un archivo general con toda la información. Por último, esta información se encuentra disponible mediante las publicaciones con que cuenta la ONE.

- Modelo 0758 o Ingresos por ventas de bienes de consumo y servicios a la población: la información es recogida cada tres meses mientras que, en el Modelo 0768 o Ingresos por ventas de bienes de consumo y servicios a la población por tipos de mercados del Comercio Interior se recoge mensual. En ambos modelos la información es recogida a través de una serie de indicadores como:

- ✓ Los Productos alimenticios y bebidas no alcohólicas: donde es anotado el valor de las ventas de alimentos incluyendo pan y dulces, productos cárnicos, productos pesqueros, leches, quesos y huevos, aceites y grasas, productos agropecuarios, azúcar, compotas, confituras y otros productos alimenticios y las bebidas no alcohólicas como refrescos, aguas, té y café.
- ✓ Bebidas alcohólicas, tabacos y cigarros: es donde se anota el valor total de las ventas minoristas de bebidas alcohólicas, tabacos y cigarros.
- ✓ Prendas de vestir y calzado: es anotado el valor total de las ventas minoristas de prendas de vestir y calzado, así como también el servicio prestado por estos conceptos.
- ✓ Vivienda y otros servicios relacionados: es anotado el valor total de los servicios prestados a la población por estos conceptos.
- ✓ Muebles, accesorios, productos textiles y otros artículos: es anotado el valor total de las ventas de muebles, productos textiles y accesorios para el hogar, así como el servicio de reparación de los mismos.
- ✓ Salud: es anotado el valor total de las ventas a la población de productos farmacéuticos, artículos ópticos, auditivos y otros. Incluye los servicios médicos prestados a través de consultas externas, internos en hospitales, policlínicas y a pacientes residentes en el extranjero que vienen al país a recibir este servicio.
- ✓ Mantenimiento y reparación de equipos de transporte: es anotado el valor por el servicio de mantenimiento y reparación de equipos de transporte personal, tales como instalación de partes y accesorios, balanceo de ruedas, revisión técnica, servicios de avería, cambio de aceite, engrasado y lavado (incluye el costo de mano de obra y el costo de materiales); neumáticos (nuevos, usados o recauchados), cámaras para neumáticos, bujías, baterías, amortiguadores, filtros, bombas y otras piezas de repuesto o accesorios para equipos de transporte personal; incluye extintores de fuego para equipos de transporte; productos específicamente para limpieza y mantenimiento de equipos de transporte, tales como pintura, limpiadores de cromo, selladores y pulidores para carrocerías; cubiertas de automotores, motocicletas. Excluye cascos para motociclistas, productos no específicos para la limpieza y mantenimiento, tales como agua destilada, esponjas, gamuzas, detergentes y radio para automóvil.

- ✓ Recreación y cultura: es anotado el total de las ventas y los servicios que por estos conceptos correspondan.
- ✓ Educación: es anotado el valor total tanto en moneda nacional como en moneda libremente convertible por el servicio de enseñanza prestado según corresponda.
- ✓ Restaurantes y hoteles: se anota el valor total por las ventas y los servicios según corresponda.
- ✓ Bienes y servicios diversos: comprende el total de las ventas de bienes y servicios prestados a la población según corresponda.

Flujo de información

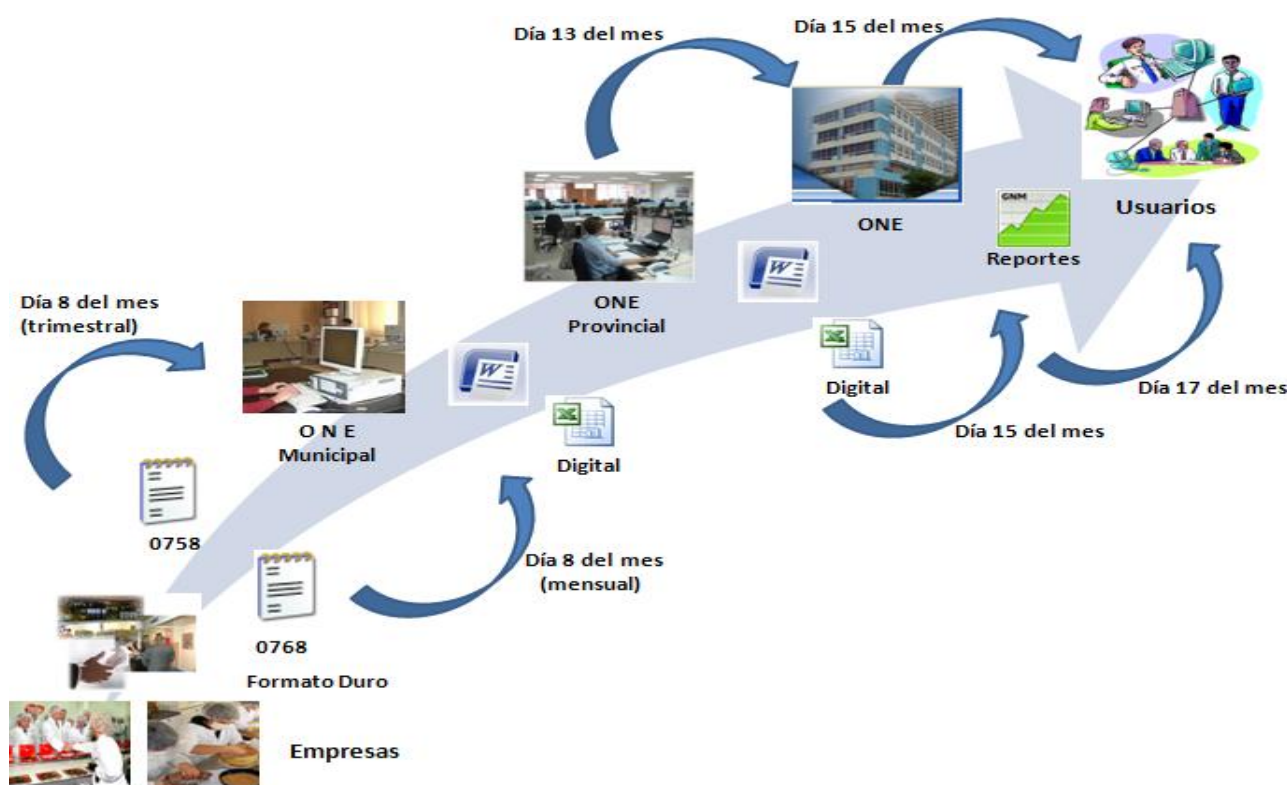


Figura 4: Flujo de información

2.1.2 Clasificadores estadísticos

Los clasificadores son elementos que tienen como objetivo identificar el origen de la información obtenida, para el caso específico de la ONE, ellos constituyen los diversos tipos de agrupaciones que

se ordenan por los centros informantes que proceden de los organismos. En los modelos planteados se utilizan cuatro clasificadores:

- Entidades: son las entidades que integran los organismos.
- Codificador de Organismo (ORG): es la relación de los diferentes organismos existentes en el país.
- División Político-Administrativa (DPA): división político-administrativa por la que está compuesta el país, conformada por 14 provincias y 168 municipios.
- Nomenclador de Actividades Económicas (NAE): está compuesto por sección, división y clase.

2.1.3 Tema de análisis identificado

Un tema de análisis es la forma en que se divide o clasifica la información referente a una determinada organización con respecto a los distintos temas que se incluyen, estos se escogen en correspondencia con los objetivos que se quieren lograr por las áreas de la organización. En la presente investigación se define como tema de análisis el Comercio Interior.

2.1.4 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.

Se definieron para la realización de los cálculos estadísticos en la esfera del Comercio Interior en la ONE 33 reglas del negocio a las cuales se hace referencia en el artefacto Reglas de negocio y transformación, donde se realiza una explicación detalla y concisa de los mismos.

2.1.5 Necesidades de los usuarios

Después de haber realizado un análisis del negocio que tiene lugar en la ONE se define como necesidad de los usuarios el análisis y la difusión de los indicadores relacionados con la actividad de Comercio Interior que es recopilada a través de los siguientes modelos:

- Modelo 0758 referente a los ingresos por la venta de bienes de consumo y servicios a la población que son recaudados en el año actual así como el año anterior en ambas monedas.

Se recoge información de todas las entidades que tengan como objeto social las actividades referentes al comercio minorista, gastronomía y/o servicios a la población, tanto como actividad fundamental como secundaria.

- Modelo 0768 referente a los ingresos por la venta de bienes de consumo y servicios a la población recaudados por tipo de comercio, el total recaudado tanto en moneda nacional como en moneda libremente convertible. Esta información es recogida por las entidades estatales subordinadas administrativa o metodológicamente al Ministerio del Comercio Interior (MINCIN), pertenecientes al Sistema del Comercio Interior en Cuba, concebidas y autorizadas a ejecutar el mismo, (Resolución 333/2004 del MINCIN).
- Series: valor de la circulación mercantil minorista donde se recoge información sobre comercio minorista, alimentación pública y como total de estas la circulación mercantil minorista; ventas totales por tipo de establecimientos en el comercio minorista, se recoge información de diferentes tipos de establecimientos; ventas por conceptos en la alimentación pública recoge la información sobre algunos productos asignados a la población. La información referente a las series antes mencionadas es registrada por años y provincias. Además se analizan otras series como: Servicios comerciales prestados a la población, recoge la información de diferentes servicios en cantidad y Servicios comerciales prestados a la población en valor, recoge la información de diferentes servicios por su valor. En estas últimas series la información se almacena por año.

2.1.6 Levantamiento de requisitos

A partir de las necesidades de los usuarios o clientes, se definen a través de un proceso de captura de requisitos, los requisitos de información que representan toda la información que debe estar disponible en la aplicación y los requisitos multidimensionales que son las variables de entrada y de salida de dicha aplicación, además los requisitos funcionales que no son más que capacidades o condiciones que el sistema debe cumplir para dar respuesta a todos los requisitos de información mencionados anteriormente y no funcionales que son las propiedades o cualidades que el producto debe cumplir.

Requisitos de Información

Se identificaron 104 requisitos de información dentro del mercado de datos Comercio Interior. A continuación se mencionan algunos de ellos y los restantes se referencian específicamente en el

artefacto Especificación de Requisitos, donde se realiza una explicación detallada y concisa de los mismos.

Modelo 0758 o Ingresos por venta de bienes de consumo y servicios a la población

RI1. Obtener ingresos de los productos alimenticios y bebidas no alcohólicas en la venta real en Moneda Nacional (CUP) por el dpa, nae, organismo, entidad, comercio, productos y servicios, en un tiempo dado.

RI2. Obtener ingresos de los productos alimenticios y bebidas no alcohólicas real en Monedas Convertibles (CUC) por el dpa, nae, organismo, entidad, comercios, productos y servicios, en un tiempo dado.

RI3. Obtener ingresos de los productos alimenticios y bebidas no alcohólicas del año anterior en CUP por el dpa, nae, organismo, entidad, comercios, productos y servicios, en un tiempo dado.

RI4. Obtener ingresos de los productos alimenticios y bebidas no alcohólicas del año anterior en CUC por el dpa, nae, organismo, entidad, comercios, productos y servicios, en un tiempo dado.

RI5. Obtener total de ingresos en ambas monedas real de los productos alimenticios y bebidas no alcohólicas por el dpa, nae, organismo, entidad, comercios, productos y servicios, en un tiempo dado.

Modelo 0768 o Ingresos por venta de bienes de consumo y servicios a la población por tipos de mercados de Comercio Interior

RI6. Obtener ingresos según comercio de los productos alimenticios y bebidas no alcohólicas por el dpa, nae, entidad, comercios y clasificación, productos y servicios, en un tiempo dado.

RI7. Obtener total de ingresos de los productos alimenticios y bebidas no alcohólicas por el dpa, nae, entidad, comercios y clasificación, productos y servicios, en un tiempo dado.

RI8. Obtener total de ingresos de los productos alimenticios y bebidas no alcohólicas en CUC por el dpa, nae, entidad, comercios y clasificación, productos y servicios, en un tiempo dado.

Series históricas

Valor de la circulación mercantil minorista

RI9. Obtener el valor de la circulación mercantil minorista dado las provincias y los años.

RI10. Obtener el valor del comercio minorista dado las provincias y los años.

RI11. Obtener el valor de la alimentación pública dado las provincias y los años.

Ventas totales por tipos de establecimientos en el comercio minorista

RI12. Obtener ventas en la tienda de productos alimenticios dado las provincias y los años.

RI13. Obtener ventas en la tienda de productos industriales dado las provincias y los años.

RI14. Obtener ventas en las farmacias y ópticas dado las provincias y los años.

RI15. Obtener total de ventas por tipos de establecimientos dado las provincias y los años.

Ventas por conceptos en la alimentación pública

RI16. Obtener ventas por conceptos en la alimentación pública de comestibles dados las provincias y los años.

RI17. Obtener ventas por conceptos en la alimentación pública de bebidas alcohólicas dado las provincias y los años.

RI18. Obtener ventas por conceptos en la alimentación pública de cervezas dado las provincias y los años.

RI19. Obtener ventas por conceptos en la alimentación pública de tabacos y cigarrillos dados las provincias y los años.

RI20. Obtener total de ventas por conceptos en la alimentación pública dado las provincias y los años.

Servicios comerciales prestados a la población

RI21. Obtener la cantidad de servicios prestados a la población dado el año.

RI22. Obtener el valor de servicios prestados a la población dado el año.

Requisitos Funcionales

Requisitos Funcionales del Subsistema Integración de Datos

RF1. Almacenar la información histórica

RF2. Actualizar la información almacenada

RF3. Realizar salvadas de información de resguardo

RF4. Integrar de forma automática los datos

RF5. Integrar de forma manual los datos

RF6. Permitir nuevos procesos de integración de datos

Requisitos Funcionales del Subsistema de visualización de datos

RF7. Autenticar Usuario

RF8. Insertar usuario

RF9. Eliminar usuario

RF10. Insertar rol

RF11. Eliminar rol

RF12. Insertar reporte

RF13. Modificar reporte

RF14. Eliminar reporte

RF15. Configurar elementos del cubo.

RF16. Mostrar consulta MDX.

RF17. Configurar tabal OLAP

RF18. Mostrar padres.

RF19. Mostrar propiedades.

RF20. Suprimir filas y columnas vacías.

RF21. Intercambiar eje.

RF22. Detallar miembros.

RF23. Entrar en detalles.

RF24. Mostrar datos de origen.

RF25. Mostrar gráfico.

RF26. Configurar gráfico.

RF27. Configurar impresión.

RF28. Exportar reporte a .pdf.

RF29. Exportar reporte a .xls.

Requisitos no Funcionales

Los requerimientos no funcionales son propiedades o cualidades que el producto debe tener. Se identificaron 27 requisitos no funcionales dentro del mercado de datos Comercio Interior, a los cuales se hace referencia específicamente en el artefacto Especificación de Requisitos, donde se realiza una explicación detalla y concisa de los mismos.

2.1.7 Reportes candidatos

Teniendo en cuenta los pedidos de información anteriormente mencionados, se identificaron ocho reportes, los cuales se mencionan a continuación, haciendo referencia de una manera más detallada en el artefacto reportes candidatos:

- Ingresos por ventas

- Ingresos por ventas, tipos de mercados del comercio interior
- Ingresos por ventas, tipos de mercados por el total en ambas monedas
- Valor de la circulación mercantil minorista
- Ventas por conceptos en la alimentación pública
- Ventas totales por tipos de establecimientos en el comercio minorista
- Servicios comerciales prestados a la población
- Servicios comerciales prestados a la población, en valor

2.1.8 Casos de Uso del Sistema (CUS)

El diagrama de CU es la representación de la relación entre los CU y los actores del negocio, que no son más que la captura de los requisitos y los actores que son las personas o programas que interactúa con el sistema. Después de realizar un análisis del levantamiento de requisitos se define el siguiente diagrama de CU de sistema, así como una breve descripción de todos sus componentes.

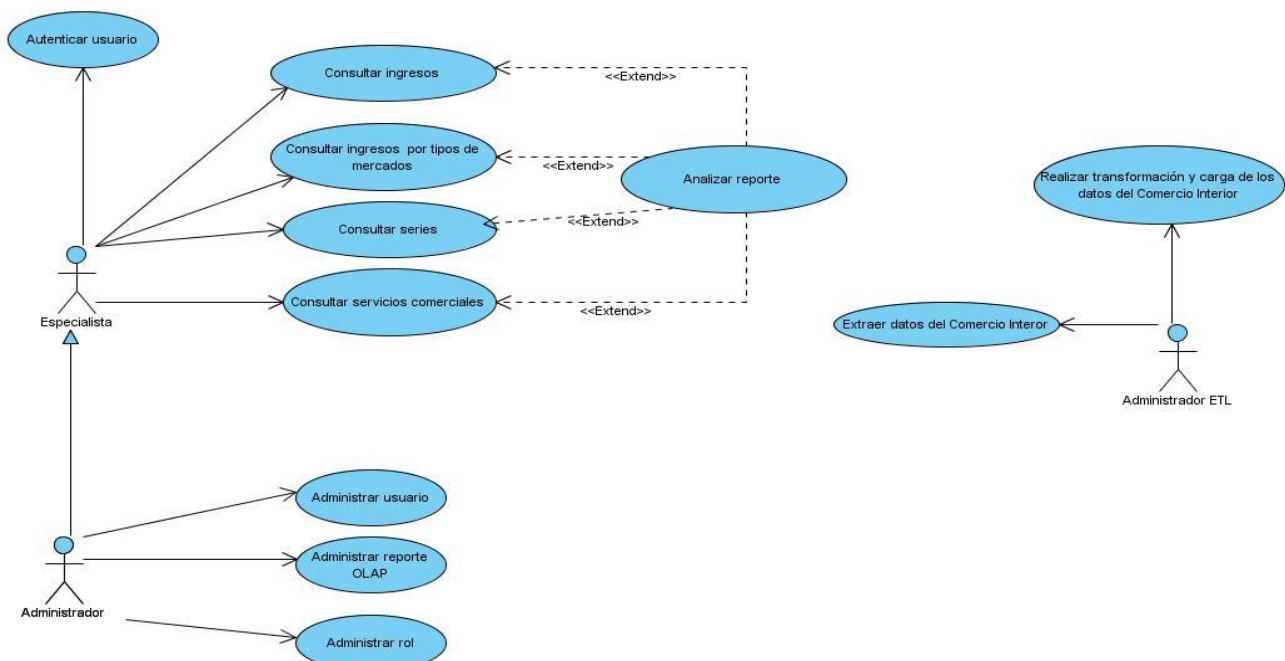


Figura 5: Diagrama de Casos de Uso del Sistema

Tabla 1: Descripción de los actores y CUS

Actor	Descripción
Especialista	Se encarga de analizar y consultar la información.
Administrador	Se encarga de realizar todas las operaciones de administración de usuarios, reportes y roles.
Administrador ETL	Es el encargado de monitorizar los procesos de extracción, transformación y carga de los datos.
Caso de Uso	Descripción
Consultar ingresos	Visualiza todos los reportes de ingresos por ventas de bienes y consumo y servicios a la población.
Consultar ingresos por tipos de mercado	Visualiza todos los reportes de ingresos por ventas de bienes y consumo y servicios a la población por tipos de mercado del comercio minorista.
Consultar series	Visualiza todos los reportes de series circulación mercantil, tipo de establecimiento y alimentación pública.
Consultar servicios comerciales	Visualiza todos los reportes de servicios comerciales prestados a la población en cantidad y en valor.
Extraer datos del Comercio Interior	Realiza la extracción de los datos.
Realizar transformación y carga de los datos del Comercio Interior	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.

A continuación se describe el Caso de Uso (CU) Consultar ingresos, los restantes se referencian en el artefacto Modelo de CUS.

Tabla 2: Descripción del CU Consultar ingresos

Caso de Uso:	Consultar ingresos	
Tipo:	Información	
Actores:	Especialista, Administrador	
Resumen	El caso de uso inicia cuando el actor entra al sistema, selecciona el reporte de ingresos por venta de bienes de consumo y servicios a la población que desea consultar. El caso de uso termina cuando se muestra el reporte correspondiente.	
Precondiciones:	El usuario tiene que estar autenticado.	
Referencias	RI1,..., RI66,CU7	
Prioridad	Crítico	
Flujo Normal de Eventos		
Acción del Actor	Respuesta del Sistema	
1. El actor entra al sistema	2 El sistema muestra las áreas de análisis existentes.	
3. El actor selecciona el AA Comercio interior.	4 El sistema muestra los libros de trabajos que contiene el AA Comercio interior.	

5. El actor selecciona el LT Ingresos por venta de bienes de consumo y servicios a la población.	6. El sistema muestra los reportes pertenecientes al LT Ingresos por venta de bienes de consumo y servicios a la población.
7. El actor selecciona el reporte deseado.	8. El sistema muestra la información contenida en el reporte seleccionado y brinda la posibilidad de analizar el reporte desde diferentes perspectivas. Ir al CU: Analizar reporte.

Sección: “Ingresos por venta”

Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad
Variables de entrada relacionadas con el CU Consultar Ingresos: <ul style="list-style-type: none"> • Nivel Territorial • NAE • Entidades • Organismo • Comercios • Indicadores de servicios • Trimestres • Unidad de medida 	Variables de salida disponibles en el caso de uso Consultar Ingresos: <ul style="list-style-type: none"> • Ingresos por venta real en CUP • Ingresos por venta real en CUC • Ingresos por venta en el año anterior en CUP • Ingresos por venta en el año anterior en CUC • Total de ingresos de ambas monedas real • Total de ingresos de ambas moneda del año anterior 	Rango de tiempo en que se solicitan las variables de salida: <ul style="list-style-type: none"> • Trimestral

Prototipo de interfaz

Poscondiciones Los reportes de ingresos han sido analizados por el especialista.

2.1.9 Perfilado de datos

El perfilado de los datos permite lograr un mejor entendimiento de los datos y así definir nuevas reglas del negocio, que pasarían a ser las reglas de transformación para llevar a cabo los procesos de ETL. Como resultado se obtuvieron:

- Que no existen valores nulos
- El tipo de dato identificado es varchar
- Existen valores duplicados, distintos y únicos.

2.2 Diseño

2.2.1 Matriz bus

La Matriz Bus no es más que la representación de la relación que existe entre los hechos y las dimensiones, que han sido diseñadas para el sistema. Las tablas de hechos se relacionan directamente con las dimensiones algunas son comunes para los dos modelos y otras específicas (ver tabla 5).

Tabla 3: Dimensiones

D1: dim_organismo	D8: dim_comercio_int
D2: dim_nae	D9: dim_comercios_clasificacion
D3: dim_dpa	D10: dim_provincia
D4: dim_entidad	D11: dim_servicios_prestados

D5: dim_temporal_mes	D12: dim_temporal_anno
D6: dim_temporal_trimestre	D13: dim_um
D7: dim_indicadores_servicios	

Tabla 4: Hechos

H1: hech_ing_vta_serv_poblacion	H3:hech_series_mercantil_establecimiento _alimentacion
H2:hech_ing_vta_serv_poblacion_tipo_merc_mincin	H4: hech_servicios_comerciales

Tabla 5: Matriz bus

Hechos/Dimensiones	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13
H1	X	X	X	X		X	X	X					X
H2		X	X	X	X		X		X				X
H3										X		X	X
H4			X								X	X	

2.2.2 Modelo de datos

Seguidamente se muestra el modelo de datos, correspondiente a las dimensiones y hechos seleccionados, así como las medidas especificadas según los modelos 0758-09 ,0768-04 y las series valor de la circulación mercantil minorista, ventas totales por tipos de establecimientos en el comercio minorista, ventas por conceptos en la alimentación pública y servicios comerciales prestados a la población.

- **dim_dpa:** dimensión referente a la variable de entrada dpa.

Jerarquía:

Provincia ->Municipio

- **dim_temporal_mes:** dimensión referente a la variable de entrada temporal mes.

Jerarquía:

Año ->Semestre->Trimestre->Mes

- **dim_temporal_trimestre:** dimensión referente a la variable de entrada temporal trimestre.

Jerarquía:

Año ->Semestre->Trimestre

- **dim_comercios_clasificacion:** dimensión referente a la variable de entrada comercios clasificación.

Jerarquía:

Tipo de comercio ->Clasificación -> Código de columna

- **dim_servicios_prestados:** dimensión referente a la variable de entrada servicios prestados.

Jerarquía:

Temática ->Subtemática->Nombre

Hechos y medidas

Al realizar el análisis de los modelos propuestos se definen cuatro hechos, debido a que la información recogida en cada una de estos es distinta.

hech_ing_vta_serv_poblacion: en el hecho ingresos por ventas y servicios a la población, se recoge información trimestral de los ingresos de años anteriores y real del año actual en ambas monedas y de ello el total en CUC y CUP.

Medidas

- **vta_real_cuc:** medida referente a la variable de salida ingresos por venta real en CUC.
- **vta_real_cup:** medida referente a la variable de salida ingresos por venta real en CUC.
- **anno_anterior_cuc:** medida referente a la variable de salida ingresos por venta real en el año anterior en CUC.
- **anno_anterior_cup:** medida referente a la variable de salida ingresos por venta real en el año anterior en CUP.

hech_ing_vta_serv_poblacion_tipo_merc_mincin: en el hecho ingresos por venta y servicios a la población por tipos de mercado, se recoge información mensual por tipo de mercado, el total de los ingresos y de ello el total en CUC y el ingreso según comercio.

Medidas

- **total:** medida referente a la variable de salida total de ingresos.
- **total_cuc:** medida referente a la variable de salida total de ingresos en CUC.
- **ingreso_segun_comercio:** medida referente a la variable de salida ingreso según comercio.

hech_series_mercantil_establecimientos_alimentacion: en este hecho se agrupa información de cinco series pertenecientes al comercio minorista. Se recoge información de la alimentación pública, el comercio minorista, tiendas de productos industriales, tiendas de productos alimenticios, farmacias y ópticas, comestibles, bebidas alcohólicas, cervezas, tabacos y cigarros, así como el total de las ventas por tipos de establecimientos y el total de las ventas en la alimentación pública.

Medidas

- **circulacion_comercio_minorista:** medida referente a la variable de salida circulación comercio minorista.
- **circulacion_alimentacion_publica:** medida referente a la variable de salida circulación en la alimentación pública.

- **cant_tiendas_productos_alimenticios:** medida referente a la variable de salida cantidad de ventas de las tiendas de productos alimenticios.
- **cant_tiendas_productos_industriales:** medida referente a la variable de salida cantidad de ventas de las tiendas de productos industriales.
- **cant_farmacias_opticas:** medida referente a la variable de salida cantidad de ventas en las farmacias y óptica.
- **cant_comestibles:** medida referente a la variable de salida cantidad de ventas de comestibles.
- **cant_bebidas_alcoholicas:** medida referente a la variable de salida cantidad de ventas de bebidas alcohólicas.
- **cant_cervezas:** medida referente a la variable de salida cantidad de ventas de cervezas.
- **cant_tabacos_cigarros:** medida referente a la variable de salida cantidad de ventas de tabacos y cigarros.
- **total_vta_tipo_establecimiento:** medida referente a la variable de salida total de ventas por tipos de establecimientos.
- **total_vta_alimentacion_publica:** medida referente a la variable de salida total de ventas por conceptos de alimentación pública.

hech_servicios_comerciales: en el hecho servicios comerciales se recoge información sobre el total de los servicios prestados a la población así como su valor.

Medidas

- **total_cant_serv_comercial_prestados_pobl:** medida referente a la variable de salida total de la cantidad de servicios comerciales prestados a la población.
- **total_valor_serv_comercial_prestados_pobl:** medida referente a la variable de salida total del valor de servicios comerciales prestados a la población.

A continuación se muestra un ejemplo del hecho Ingresos por ventas y servicios a la población, los restantes se referencian en el artefacto Especificaciones del Modelo de Datos Dimensional, donde se realiza una explicación detalla y concisa de los mismos.

hech_ing_vta_serv_poblacion		
+#dim_indicadores_servicios_id	int4	Nullable = false
+#dim_dpa_id	int4	Nullable = false
+#dim_entidad_id	int4	Nullable = false
+#dim_organismo_id	int4	Nullable = false
+#dim_nae_id	int4	Nullable = false
+#dim_temporal_trimestre_id	int4	Nullable = false
+#dim_comercio_int_id	int4	Nullable = false
+#dim_um_id	int4	Nullable = false
vta_real_cuc	float8	Nullable = false
vta_real_cup	float8	Nullable = false
anno_anterior_cuc	float8	Nullable = false
anno_anterior_cup	float8	Nullable = false

Figura 7: Hecho Ingresos por ventas

Después de diseñar como se va almacenar la información se pasa al diseño de la forma en que debe visualizarse la misma.

2.2.3 Mapa de navegación

El desarrollo de un almacén de datos, como solución de la investigación, permitirá almacenar la información concerniente al Comercio Interior en las oficinas de la ONE para transformarlos en datos útiles a las personas encargadas de tomar decisiones. Partiendo de esto se especifica el tema de análisis identificado, así como los libros de trabajo en las que se divide este tema, que no son más que las Áreas de Análisis (AA) las cuales agrupan los reportes construidos. A continuación se abordan cada uno de los elementos que componen la estructura de navegación de la información que será presentada. Para un mayor entendimiento de esta estructura, consultar el artefacto Arquitectura de la información.

- **Tema de análisis:** constituye las secciones en las que está dividido el almacén de datos, donde incluye los Libros de Trabajo y los reportes. El Área de Análisis (AA) identificado es el Comercio Interior.
- **Libro de Trabajo (LT):** constituye las categorías en los que se agrupan los reportes, donde cada AA posee varios libros de trabajo.
- **Reportes:** constituye un informe en el que se relacionan de forma que permita tabular los

indicadores.



Figura 8: Mapa de navegación

2.2.4 Diseño de los cubos OLAP

En la presente investigación se desarrollaron cuatro cubos multidimensionales: Ingreso por venta y servicios a la población; Ingresos por venta y servicios a la población por tipos de mercados del Comercio Interior; Servicios comerciales y Series mercantil, establecimientos y alimentación, para diseñar los cubos OLAP se utilizó la herramienta Pentaho Schema Workbench presentada en el Capítulo 1, la cual facilita la creación de estos cubos. A continuación se muestran los cubos multidimensionales creados:

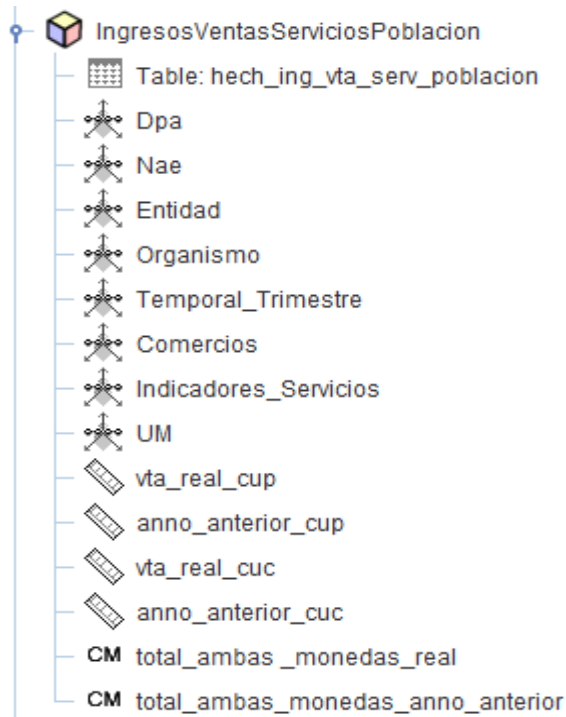


Figura 9: Cubo Ingresos por ventas y servicios a la población

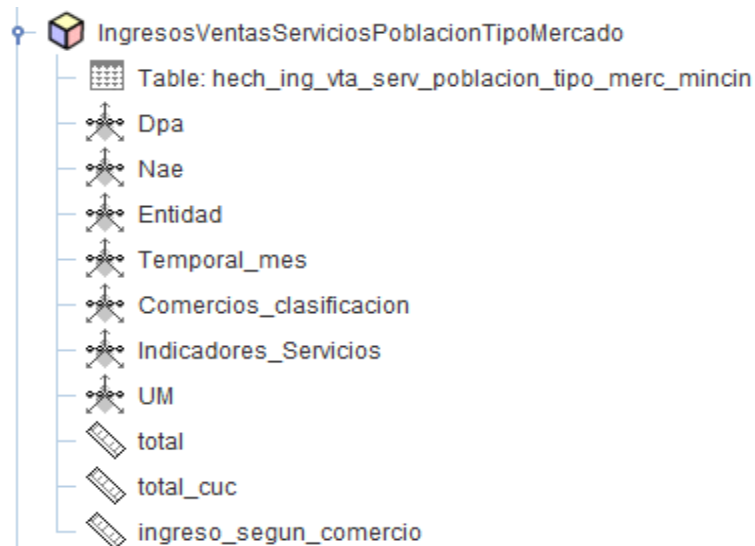


Figura 10: Cubo Ingresos por ventas y servicios a la población por tipo de mercado



Figura 11: Cubo Series mercantil, establecimientos y alimentación

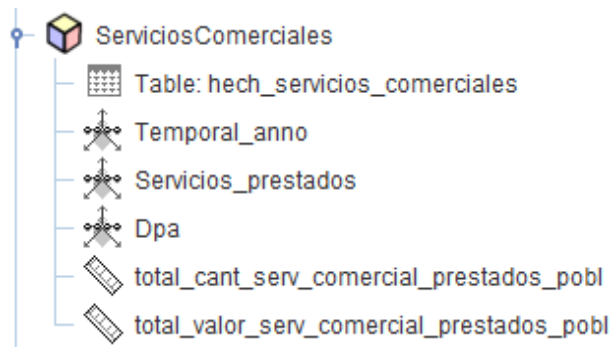


Figura 12: Cubo Servicios comerciales

Como bien se observa en las figuras anteriores, estos cubos OLAP se definieron por cada tabla de hecho, ya que estas tablas son las fuentes de información principal de la organización estructural de los datos, además es posible observar las tablas de dimensiones relacionadas con el hecho así como las medidas del mismo.

Para lograr una adecuada integración de los datos proveniente de las fuentes se realiza el diseño de los procesos de integración.

2.2.5 Diseño de los procesos de integración de datos

Para organizar los procesos de integración de datos se realizó un diseño de cómo podrían ejecutarse los mismos, a continuación se muestra una figura con el flujo general del proceso.

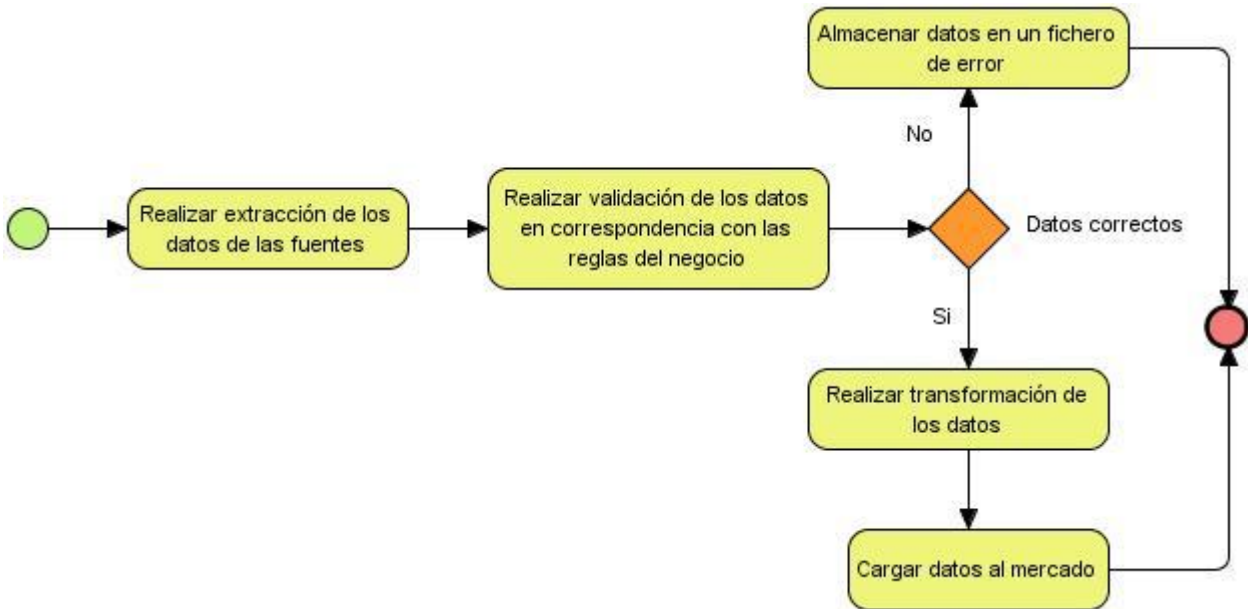


Figura 13: Diseño de los procesos de integración de datos

El diseño de las transformaciones describen los pasos a seguir para realizar la carga de los hechos y las dimensiones. A continuación se muestra un ejemplo del diseño de como poblar una dimensión:

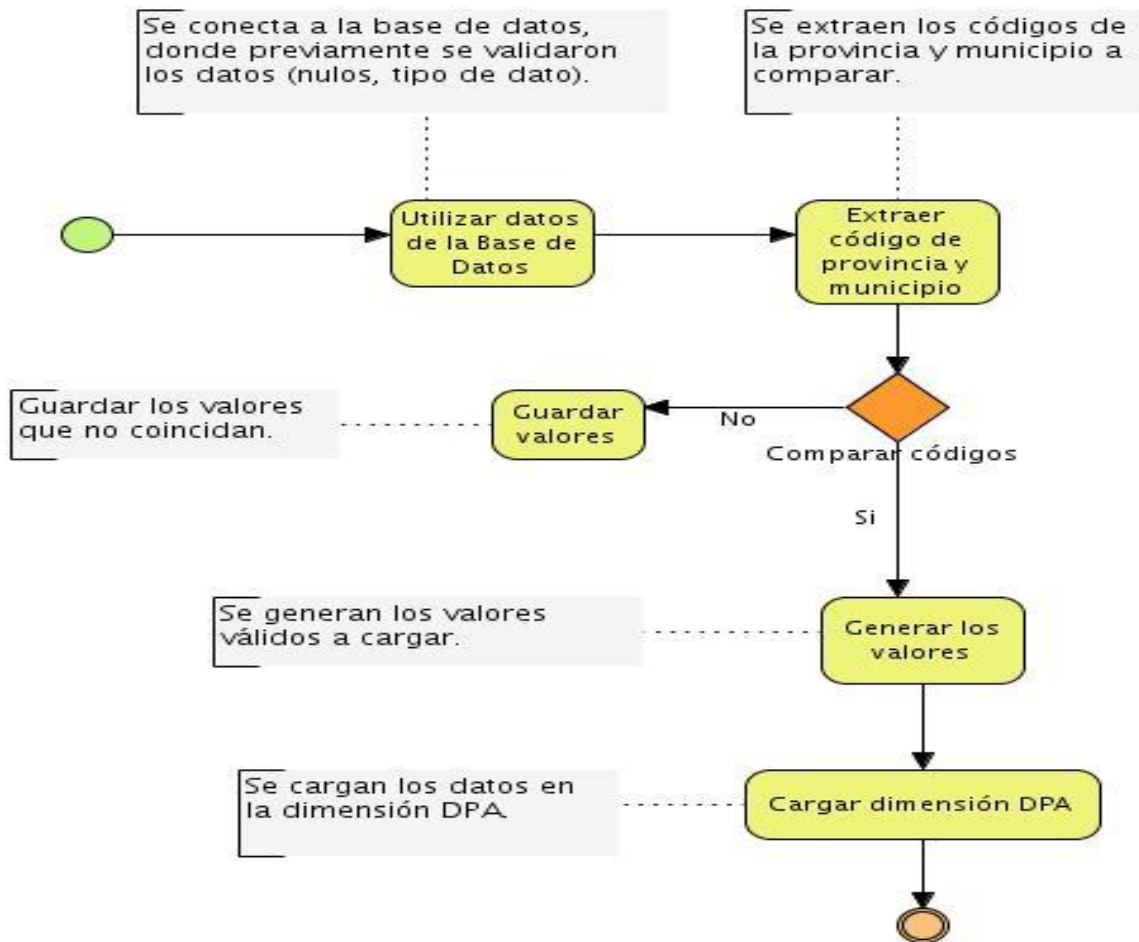


Figura 14: Diseño de la transformación de la dimensión dpa

2.2.6 Política de respaldo y recuperación

Las políticas de respaldo y recuperación en el mercado de datos Comercio Interior están convenidas por tres puntos fundamentales:

- ✓ Periodicidad de salvadas del sistema: en la ONE las salvadas del sistema se realizan en dependencia de como sea capturada la información. Para el área de Comercio Interior estas se realizarán con una frecuencia trimestral para el modelo 0758, mensual para el modelo 0768 y anual para las series históricas.
- ✓ Tablas involucradas: se involucran en la realización de las salvadas las tablas hech_ing_vta_serv_poblacion, hech_ing_vta_serv_poblacion_tipo_merc_mincin, hech_servicios_comerciales, hech_series_mercantil_establecimientos_alimentacion. En caso

de que cambie alguna dimensión en un determinado período del tiempo, se realizará una salva de la misma.

- ✓ Salvas: las salvas totales se realizan en los meses de junio y diciembre, además se realizan salvas parciales de acuerdo a la periodicidad.

Conclusiones

En el presente capítulo fueron definidos los principales procesos que se llevan a cabo en el desarrollo del análisis y diseño del mercado de datos de Comercio Interior. El proceso del negocio fue realizado para un mejor entendimiento de los usuarios que no se encuentran dentro de este y así comprenda todo el proceso realizado. Se determinaron las reglas del negocio a medir, para lograr una estandarización del envío de la información, quedaron definidos los requisitos funcionales, no funcionales, de información, multidimensionales, para dar respuesta a las necesidades de los usuarios y conocer las condiciones, propiedades que tanto el sistema como el producto deben cumplir. Se especificaron y detallaron los CU del sistema, para lograr un mejor entendimiento de las necesidades de los usuarios. También se especificaron 20 tablas que componen el modelo multidimensional y por último se definieron las políticas de respaldo y recuperación para una correcta seguridad de la información.

CAPÍTULO III: IMPLEMENTACIÓN DE LA SOLUCIÓN

Introducción

En el presente capítulo se realiza la implementación para el proceso de ETL en el área de Comercio Interior de la ONE al igual que la capa de visualización de los datos, teniendo en cuenta las necesidades del cliente.

3.1 Implementación de la Base de Datos

3.1.1 Estructura de los datos

Es una forma de organizar los datos. Estas estructuras de datos se caracterizan por las funciones de acceso que utilizan para almacenar y acceder a elementos individuales. Se pueden definir también como una forma de organizar un conjunto de datos elementales con el objetivo de facilitar su manipulación. Por debajo de estas estructuras hay un nivel, en el cual se sitúan los archivos, discos, particiones y espacios de tablas. La buena práctica con dichos elementos y el dominio pleno de los mismos inciden grandemente en el éxito de la solución.

Esquemas

Los esquemas representan una manera de tener organizada toda la información contenida en las BD estos pueden contener funciones, operadores y tipos de datos. A diferencia de las BD los esquemas no se encuentran separados entre sí, esto permite que el usuario pueda tener acceso a ellos siempre que posea los permisos adecuados. La solución cuenta con 20 tablas en total, de ellas 13 son dimensiones, cuatro hechos y 3 auxiliares, distribuidas por los 3 esquemas.

Para el desarrollo del sistema propuesto en la presente investigación se definieron cuatro esquemas:

- dimensiones: contiene las tablas de las dimensiones propuestas que son comunes con la BD de la ONE
- mart_comercio_int: contiene todas las tablas de hechos propuestas y las dimensiones propias del mercado de datos.
- stg_etl: contiene las tablas correlacionadoras de los modelos propuestos.

3.1.2 Usuarios y privilegios en la Base de Datos

Postgres ofrece la posibilidad de agrupar a los usuarios según las necesidades de permisos y accesos que necesitará cada rol para realizar su función como trabajador del sistema.

Usuarios y Roles:

Los usuarios y roles definidos en la BD contribuyen a garantizar la seguridad de la misma.

A continuación se describen los roles establecidos:

- **Administrador:** tiene acceso a la BD en su totalidad, dígase administración y configuración tanto de la BD como de los usuarios restantes.
- **Analista:** su rol se basa en consultar la información de la BD.
- **Administrador de ETL:** su función se basa en la realización de los procesos de ETL en la interacción con la BD.

Privilegios:

Los privilegios que se les asigna a los usuarios del sistema son basados en el rol que desempeñan. En caso del:

- **Administrador:** este usuario se le asigna los derechos de *Owner, Select, Update, Insert, Delete, Refresh y Trigger* sobre la estructura de la BD.
- **Analista:** solo tiene derechos de *Select* los datos almacenados en el Almacén.
- **Administrador de ETL:** por el rol que este usuario desempeña a la hora de interactuar con la BD, se le asigna el privilegio de *Select, Update, Insert, Delete, Refresh y Trigger*, sobre los datos en el Almacén.

3.2 Implementación del subsistema de integración de datos (ETL)

El proceso de ETL se basa en extraer los datos de las fuentes, realizar las transformaciones y limpieza de los mismos y posteriormente ser cargados en el sistema. La transformación de los datos se hará de acuerdo a las reglas que se definieron en el negocio. Para su realización se ejecutaron las transformaciones de las dimensiones y de los cuatro hechos correspondientes, a continuación se

muestra un ejemplo de la transformación realizada para la carga del hecho ingresos por ventas y servicios a la población, las demás se pueden encontrar en la carpeta de código fuente en el expediente de proyecto de la solución en formato ktr.

Ing_vta_poblacion: en la transformación del hecho ingreso por venta de bienes y servicios a la población se realiza la carga de los registros fuentes ('dbf') correspondiente al modelo 0758. Posteriormente se valida que los datos cargados de la fuente no contengan valores nulos y el tipo de dato, además de otras validaciones necesarias para que al poblar el hecho los datos almacenados sean los correctos. Finalmente se carga el hecho con la transformación correspondiente.

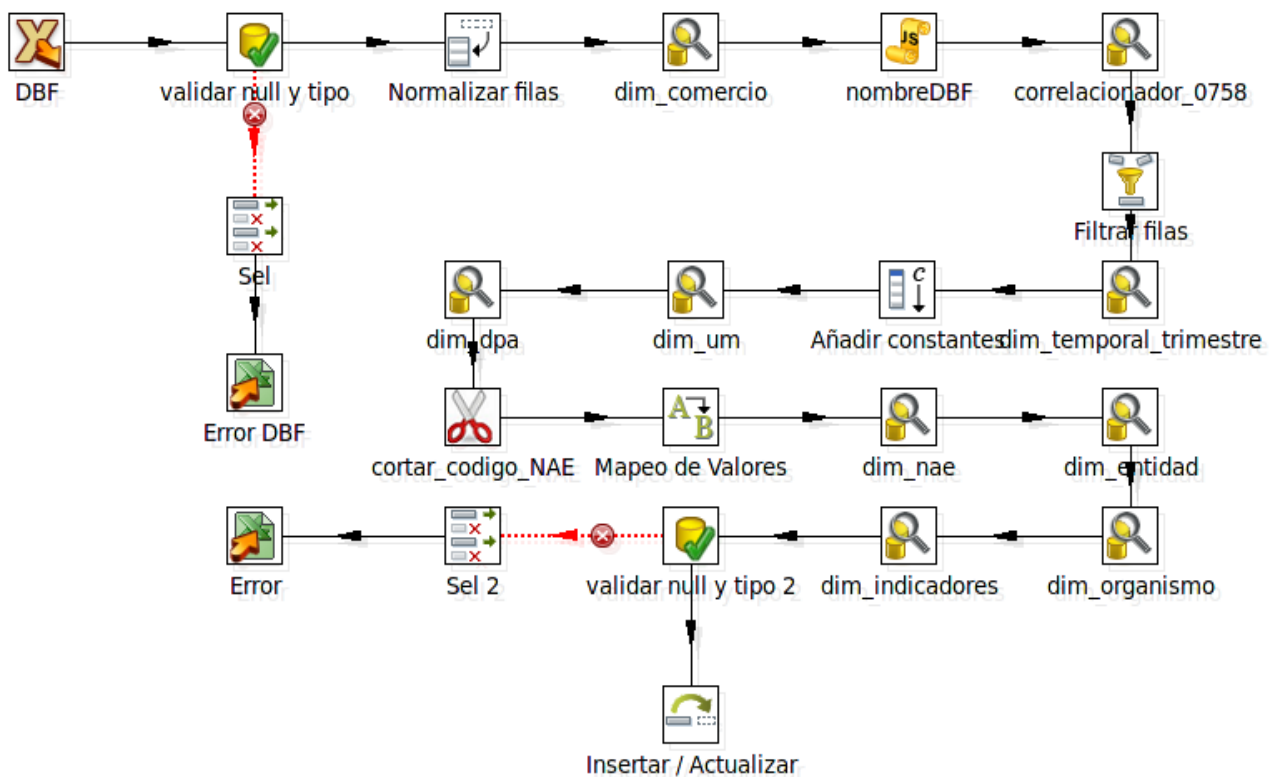


Figura 15: Transformación del hecho ingreso por venta y servicios a la población

3.3 Implementación del subsistema de visualización de datos(BI)

La Inteligencia del Negocio, al igual que otros términos o conceptos no escapa a la diversidad de interpretaciones. Sin embargo queda esencialmente claro que: “no es una metodología, software, sistema o herramienta específica, es más bien una colección de tecnologías que van desde arquitecturas para almacenar datos, metodologías, técnicas para analizar información y software, entre otros, con un fin común para el apoyo a la toma de decisiones”. (34)

3.3.1 Implementación de los reportes candidatos

Las expresiones multidimensionales (MDX) constituyen un lenguaje de consulta para BD multidimensionales, permiten consultar objetos como los cubos OLAP y son utilizadas en Inteligencia de Negocios para generar reportes que facilitan la toma de decisiones basados en datos históricos, con la posibilidad de cambiar la estructura o rotar el cubo. Las consultas MDX devuelven un conjunto de celdas que contienen los datos del cubo. (34)

Para la realización de cada reporte se define una consulta MDX, por lo que se tienen 8 consultas, que coinciden con los reportes candidatos. A continuación se muestra un ejemplo de cómo queda visualizado el reporte **Ingresos por ventas**, una vez ejecutada la consulta MDX en el Pentaho BI Server.

Indicadores de servicios	Comercio minorista				Gastronomía			
	● Ingresos por venta real en cup	● Ingresos por venta real en cuc	● Ingresos por venta en el año anterior en cup	● Ingresos por venta en el año anterior en cuc	● Ingresos por venta real en cup	● Ingresos por venta real en cuc	● Ingresos por venta en el año anterior en cup	● Ingresos por venta en el año anterior en cuc
Total de ingresos	684310,0	4817,9	514820,0	2138,6	2259480,0	838,5	184344,0	935,3
Productos alimenticios y bebidas no alcohólicas	422736,0	1249,7	296072,0	407,9	2094950,0	320,4	49082,1	312,8
Bebidas alcohólicas, tabacos y cigarros	45670,4	2640,9	23291,8	23,9	120962,0	238,1	86272,3	372,5
Prendas de vestir y calzado	30444,1	116,0	11521,7	102,0	0,0	0,0	0,0	0,0
Vivienda y otros servicios relacionados	5336,1	0,7	6674,8	0,0	0,0	0,0	0,0	0,0
Muebles, accesorios, productos textiles y otros artículos	9045,2	674,7	6676,8	1585,0	0,0	0,0	0,0	0,0
Salud	165484,0	1,4	164393,0	0,2	0,0	0,0	0,0	0,0
Mantenimiento y reparación de equipos de transporte	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Recreación y cultura	4903,7	44,0	5774,8	19,6	0,0	0,0	0,0	0,0
Educación	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

Figura 16: Reporte Ingresos por ventas

Un factor importante en la soluciones de almacenes de datos es la seguridad, para ello se definen roles, usuarios y permisos.

3.3.2 Seguridad en la aplicación

El acceso a los servicios del sistema deberá ser a través de un estándar Web seguro (https). El sistema permitirá la transmisión de la información por canales cifrados usando el protocolo HTTPS.

Las contraseñas de los usuarios no deben ser almacenadas en texto plano, esta información debe ser privada y específica de cada uno, de manera que nadie pueda reemplazar la identidad de un usuario en el sistema.

Las aplicaciones desplegadas en el Pentaho BI Server muestran un continuo incremento, así como los usuarios que tiene acceso a estas. Lo que trae como resultado que se creen y configuren dos roles y dos usuarios que poseen diferentes niveles de acceso al sistema.

El rol de administrador posee todos los permisos sobre el sistema y tiene asignado el usuario admin.

Details	
Role Name:	administrador
Description:	Administrador
Assigned Users	
	admin

Figura 17: Rol y Usuario de Administración

Permissions for admin:	
<input checked="" type="checkbox"/>	All Permissions
<input checked="" type="checkbox"/>	Create
<input checked="" type="checkbox"/>	Update
<input checked="" type="checkbox"/>	Execute
<input checked="" type="checkbox"/>	Delete
<input checked="" type="checkbox"/>	Grant Permissions
<input checked="" type="checkbox"/>	Schedule

Figura 18: Permisos asignados al usuario "admin"

El rol de Análisis posee única y exclusivamente permisos de ejecución sobre el sistema y tiene asignado el usuario analista.

Details	
Role Name:	analisis
Description:	Analisis
Assigned Users	
	analista

Figura 19: Rol y Usuario de Análisis

Permissions for analista:	
<input type="checkbox"/>	All Permissions
<input type="checkbox"/>	Create
<input type="checkbox"/>	Update
<input checked="" type="checkbox"/>	Execute
<input type="checkbox"/>	Delete
<input type="checkbox"/>	Grant Permissions
<input type="checkbox"/>	Schedule

Figura 20: Permisos asignados al usuario "analista"

Conclusiones

En este capítulo se abordaron los temas que dieron paso a la implementación del mercado de datos. Se dio a conocer como quedó físicamente el mercado, logrando una estructura robusta entre las relaciones de las tablas multidimensionales. La forma en que se agrupan las tablas en la BD es por esquemas, para así alcanzar una mejor organización. Se logró una buena sincronización entre la fuente de los datos y el destino, al extraer y transformar la información para ser cargada en la BD. Al tener los datos cargados en el mercado de datos se logran transformar en información valiosa para realizar acciones y análisis. Esto se realiza en el proceso de BI.

CAPÍTULO IV: VALIDACIÓN Y PRUEBAS DEL MERCADO DE DATOS

INTRODUCCIÓN

En este capítulo se realizan pruebas al mercado de datos Comercio Interior, con el objetivo de verificar su correcto funcionamiento y que cuente con la calidad requerida por el cliente. Esta validación se realiza a partir de la aplicación de listas de chequeo y los casos de pruebas diseñados.

4.1 Calidad de software

La calidad es un elemento esencial en todos los aspectos de la vida incluyendo el desarrollo de software. El interés por la calidad crece de forma continua, a medida que los clientes se vuelven más selectivos y comienzan a rechazar productos poco fiables o que realmente no dan respuesta a sus necesidades.

4.1.1 Importancia de la calidad del software

La importancia de la calidad del software radica en que el software es un producto mental, no es restringido por las leyes de la física o por los límites de los procesos de fabricación, es algo abstracto y su calidad también lo es. Sirve de plataforma para desarrollar en la organización, una serie de actividades, procesos y procedimientos, encaminados a lograr que las características del producto cumplan con los requisitos del cliente u organización, en pocas palabras que sean de calidad. (35)

4.1.2 Cualidades que debe tener un producto de software para que tenga calidad

La calidad del software está determinada por un conjunto de cualidades que lo caracterizan, estas determinan su utilidad y existencia. Para que un producto de software obtenga una buena calidad debe cumplir con las siguientes cualidades:

- **Mantenibilidad:** capacidad del producto de software de ser modificado. Las modificaciones pueden incluir las correcciones, mejoras o adaptaciones del software a cambios en el ambiente, así como en los requisitos y las especificaciones funcionales.
- **Funcionalidad:** es la capacidad del software para proporcionar funciones que satisfacen las necesidades declaradas e implícitas cuando el software se usa bajo las condiciones especificadas.
- **Portabilidad:** capacidad de producto de software de ser transferido de un ambiente a otro.

- **Confiabilidad:** la capacidad del producto de software para mantener un nivel de ejecución especificado cuando se usa bajo las condiciones especificadas.
- **Eficiencia:** capacidad del producto de software para proporcionar una ejecución o desempeño apropiado, en relación con la cantidad de recursos utilizados usados, bajo condiciones establecidas.
- **Usabilidad:** capacidad del producto de software de ser comprendido, aprendido, utilizado y de ser atractivo para el usuario, cuando se utilice bajo las condiciones especificadas. (36)

La Calidad del Software debe aplicarse a medida que el producto vaya siendo construido, paralelo al proceso de desarrollo, se realiza la etapa de pruebas las cuales constituyen un pilar indispensable para evaluar y determinar la calidad de un software.

4.2 Pruebas de calidad de software

Las pruebas de software permiten depurar un software desde sus inicios, para que el mismo llegue a manos de los clientes cumpliendo todas sus funcionalidades y libre de defectos.

Al igual que para el proceso de desarrollo, existen modelos de mejora del proceso de prueba, en la presente investigación se utilizó el Modelo V el cual fue definido por el Centro de Tecnologías de Gestión de Datos (DATEC) con el fin de lograr que el producto posea calidad y así mejorar el proceso de pruebas.

4.2.1 Modelo V

El Modelo V demuestra cómo se relacionan las actividades de pruebas con las de análisis y diseño. Este modelo presenta como gran ventaja que se puedan realizar pruebas en todo el ciclo de vida de desarrollo de un software. En la Figura 21 se observa que la punta de la V es la codificación, con el análisis y el diseño a la izquierda y la prueba y el mantenimiento a la derecha. El modelo V sugiere que la prueba unitaria y de integración sea hecha también para verificar el diseño, verificando que todos los aspectos del diseño y del programa se han implementado correctamente en el código, la prueba del sistema debe verificar el diseño del sistema, asegurando que todos los aspectos del diseño están correctamente implementados. La prueba de aceptación que es dirigida por el cliente, valida los requerimientos. La vinculación entre la parte izquierda y derecha del Modelo V implica que si se encuentran problemas durante la verificación y la validación entonces el lado izquierdo de la V puede ser ejecutado nuevamente para solucionar el problema y mejorar los requerimientos, el diseño y el código antes de retomar las pruebas del lado derecho. (35)

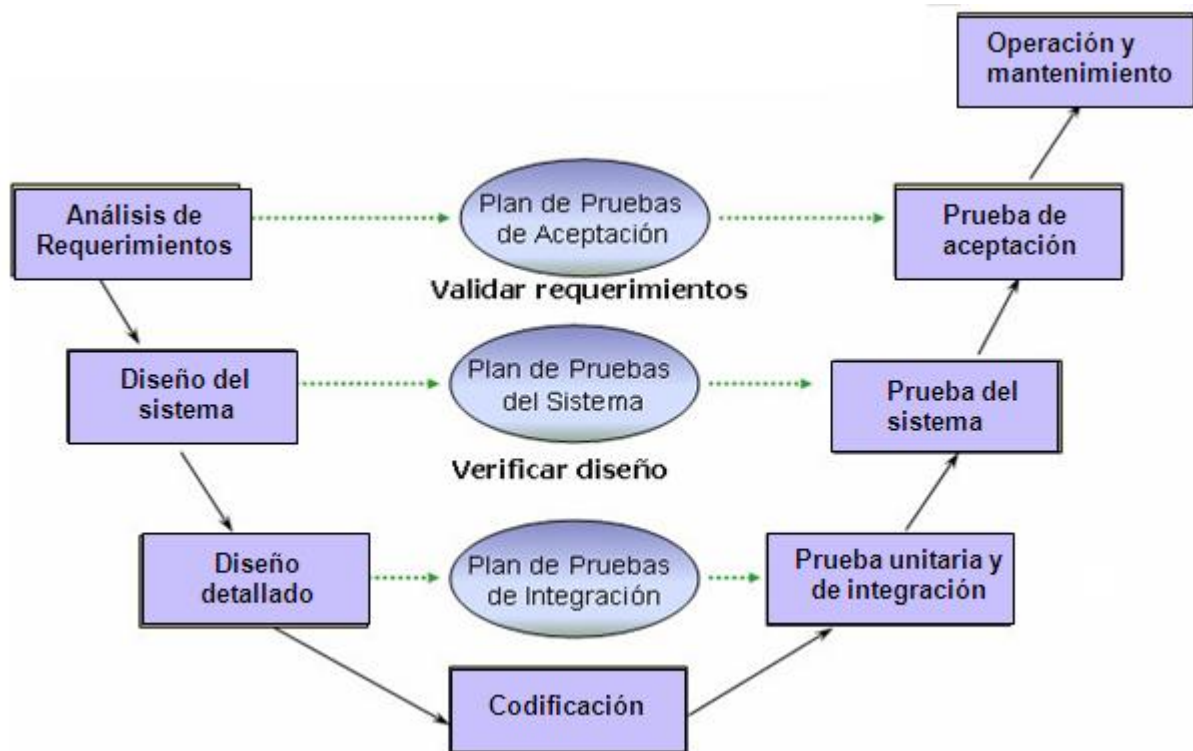


Figura 21: Modelo V

4.2.2 Tipos de pruebas en el Modelo V

Las pruebas que a continuación se mencionan son aplicadas en el centro DATEC utilizando el Modelo V:

- **La prueba unitaria:** es el proceso de probar los componentes individuales (subprogramas o procedimientos) de un programa. El propósito es descubrir discrepancias entre la especificación de la interfaz de los módulos y su comportamiento real. Estas pruebas son diseñadas y ejecutadas por el desarrollador una vez terminado el desarrollo de cada componente. (35)
- **La prueba de integración:** son las pruebas que se realizan para determinar la integración de los componentes dentro de un sistema y evaluar su correcta interfaz, funcionalidad y desempeño. Estas pruebas son diseñadas y ejecutadas por el desarrollador cuando la solución está completa junto a los especialistas del centro.
- **La prueba de sistema:** son las pruebas que se realizan para determinar el correcto funcionamiento de un sistema y su cumplimiento contra las especificaciones del producto.

Estas pruebas son ejecutadas por el desarrollador o equipo de pruebas en un ambiente controlado. (37)

Independientemente de estos tipos de prueba que se realizan, el centro efectúa un grupo de pruebas encaminadas a la terminación del ciclo de desarrollo del producto. Las pruebas realizadas son:

- **Pruebas internas:** es realizada por el equipo de desarrollo y especialistas de los departamentos que participan en la solución.
- **Pruebas de liberación:** se realiza por los especialistas de calidad, donde en un primer momento es desarrollada por los especialistas de calidad de DATEC y luego por los especialistas del Centro de Calidad para Aplicaciones Tecnológicas (CALISOFT).
- **Pruebas de aceptación:** pruebas que se realizan directamente con el cliente para validar su conformidad con el producto.

4.2.3 Herramientas de prueba

Las herramientas utilizadas para aplicar las pruebas de la solución definidas por el centro DATEC son las siguientes:

Casos de prueba: un caso de prueba no es más que un conjunto de entradas de pruebas, condiciones de ejecución y resultados esperados que se diseñan con el propósito de cumplir un objetivo en particular o una función esperada (38). Un caso de prueba debe garantizar que el producto se comporta como se describe en las especificaciones funcionales del diseño. En la presente investigación se utilizan los casos de pruebas basados en CU para probar la solución. Se realizaron cuatro casos de prueba, un ejemplo es la tabla ocho de los anexos del caso de prueba Consultar ingresos por tipo de mercado. En el expediente de proyecto, además del caso de prueba mencionado anteriormente, se encuentran los restantes casos de prueba realizados.

Lista de chequeo: la lista de chequeo es un documento que contiene un conjunto de parámetros a medir sobre un aspecto determinado con el objetivo de verificar el grado de cumplimiento de determinadas reglas. Es un instrumento de medición y evaluación que consiste básicamente en un formulario de preguntas referentes al software y a los artefactos que se generan.

La lista de chequeo definida para la evaluación del mercado de datos de Comercio Interior, analiza 14 indicadores, de los cuales ocho son críticos. En la tabla nueve de los anexos se muestran la lista de chequeo aplicada para evaluar el diseño realizado en la presente investigación.

4.3 Resultados de las pruebas

Los resultados obtenidos en las pruebas realizadas a la solución son los siguientes:

- **Pruebas internas:** realizadas por los especialistas del departamento y miembros del equipo de desarrollo arrojando los siguientes resultados.
 - ✓ Aplicación: 1 No Conformidades (NC)
 - ✓ Reportes candidatos: 9 NC
 - ✓ Casos de prueba: 7 NCPara un total de 17 NC.
- **Lista de chequeo:** después de aplicada, se genera el siguiente gráfico de barra, donde se visualiza el comportamiento de los 14 indicadores identificados, de los cuales 8 son críticos. Ver figura 22.

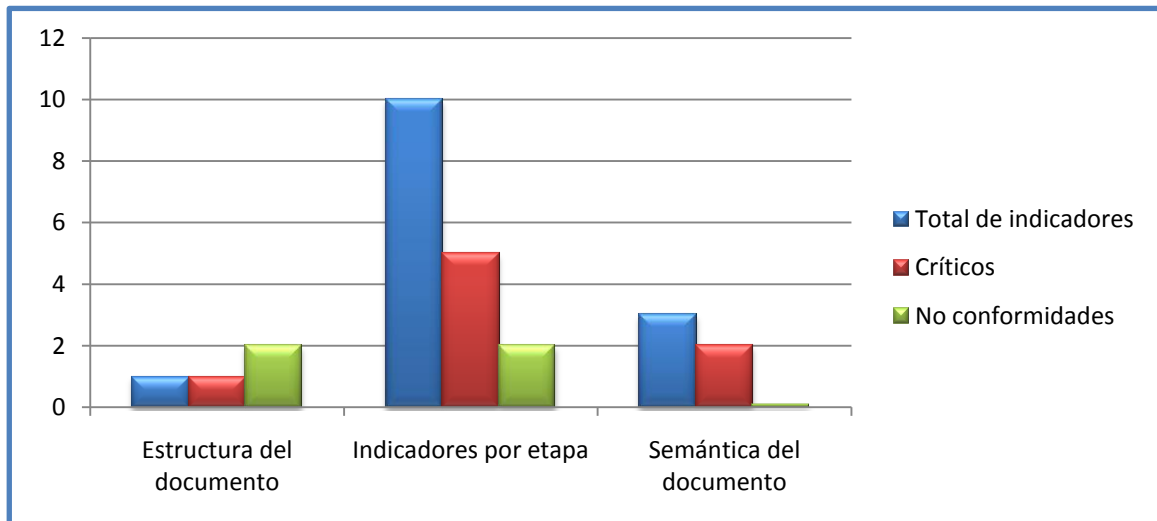


Figura 22: Comportamiento de los Indicadores

- **Prueba de liberación:** el producto fue liberado por Calisoft con 0 no conformidades, lo cual demuestra la calidad del mercado de datos Comercio Interior.
 - **Pruebas de aceptación:** se obtuvo la carta de aceptación del producto por parte del cliente.
- Todas las NC anteriormente mencionadas fueron resueltas en el período de tiempo establecido.

Conclusiones

En este capítulo se realizaron las pruebas necesarias al mercado de datos, se utilizaron las herramientas definidas para la aplicación de las pruebas como la lista de chequeo, donde se cubren todos los aspectos aplicables en el tema de análisis y necesarios para cumplir con un buen diseño, además de los casos de prueba garantizando la calidad de la solución y el cumplimiento de los requerimientos del cliente.

Conclusiones generales

Después de desarrollada la solución se arribaron a las siguientes conclusiones:

- El análisis y diseño del mercado de datos cumplió con todo lo requerido en el negocio.
- Las estructuras dimensionales que fueron implementadas para el control de la información cumplen las condiciones necesarias para la integración de los datos.
- La implementación de los reportes permitió la disponibilidad de la información para su consulta por parte de los usuarios, apoyando el proceso de toma de decisiones.
- El Sistema Gestor de Base de Datos PostgreSQL garantizó la administración de los volúmenes de información que necesita este tipo de solución.
- Los resultados satisfactorios de las pruebas realizadas permitieron validar la calidad del producto, obteniéndose la carta de aceptación del producto por parte del cliente.

Recomendaciones

- Analizar una posible integración con el Sistema Informático de Gestión Estadística (SIGE), tomando como fuente de datos los modelos de control de indicadores que establece dicho sistema.
- Implementar para el mercado de datos de Comercio interior los procesos de auditoría de metadatos como parte del subsistema de integración de datos.

Bibliografía

[En línea] [Citado el: 10 de 12 de 2010.] econlink.com.a.

[En línea] [Citado el: 2 de 12 de 2010.]

<http://bdigital.eafit.edu.co/bdigital/PROYECTO/P005.74CDA779/marcoTeorico.pdf>.

[En línea] [Citado el: 2 de 12 de 2010.]

http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/rosette_u_jm/capitulo3.pdf.

[En línea] [Citado el: 2 de 12 de 2010.] http://etl-tools.info/es/bi/almacendedatos_esquema-constelacion.htm.

[En línea] [Citado el: 21 de 11 de 2010.]

<http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/OLAP.pdf..>

[En línea] [Citado el: 24 de 5 de 2011.]

<file:///C:/Documents%20and%20Settings/Ary/Escritorio/Documentacion%20Cap%204/Cap%204/Calidad%20de%20Software.htm>.

[En línea] [Citado el: 25 de 11 de 2010.] <http://antares.inegi.org.mx/metadatos/metadat1.htm>.

[En línea] [Citado el: 25 de 11 de 2010.]

http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.

[En línea] [Citado el: 27 de 11 de 2010.] <http://www.ibiblio.org/pub/linux/docs/LuCaS/Postgresql-es/web/navegable/todopostgresql/app-pgadmin.html..>

[En línea] [Citado el: 28 de 1 de 2011.] <http://www.wordreference.com/definicion/comercio%20interior>.

[En línea] 2009. [Citado el: 19 de Junio de 2011.] <http://msdn.microsoft.com/es-es/library/ms145514.aspx>.

[En línea] 27 de Septiembre de 2008. <http://gerenciainnovacion.blogspot.com/2008/09/conceptos-inteligencia-de-negocios.html>.

[En línea] http://www.google.com/cu/#hl=es-419&biw=1024&bih=575&q=etapas+de+un+almacen+de+datos&aq=f&aqi=&aql=&oq=&gs_rfai=&fp=666b36a6e513e2ed .

[En línea] http://www.google.com/cu/#hl=es-419&biw=1024&bih=575&q=etapas+de+un+almacen+de+datos&aq=f&aqi=&aql=&oq=&gs_rfai=&fp=666b36a6e513e2ed .

[En línea] Pentaho Corporation, 2005. [Citado el: 15 de 02 de 2011.] <http://Pentaho.com>.

2005, "Pentaho BI Suite Enterprise Edition." [En línea] Pentaho Corporation, 2005. [Citado el: 29 de 11 de 2010.] http://www.pentaho.com/products/data_services/?hp=y..

Álvarez Marañón, Gonzalo. ¿Qué es Java? *Sitio oficial de la IEC.* [En línea] 1999.

<http://www.iec.csic.es/CRIPTONOMICON/java/quesjava.html>.

Apache TomCat. [En línea] <http://tomcat.apache.org/>..

BI, El rincón del. [En línea] [Citado el: 20 de 11 de 2010.]

<http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimensional/>..

BIBLIOTECA VIRTUAL de Derecho, Economía y Ciencias Sociales_Álvaro de la Helguera y García.

eumed.net. [En línea] [Citado el: 28 de 1 de 2011.] <http://www.eumed.net/libros/2006a/ah-circ/1e.htm>.

Breidenbach, Craig Walls y Ryan. *Spring in Action. Second Edition.* s.l. : Greenwich: Manning Publications Co., 2008.

Browne, Paul. O'Reilly on Java.com. *The independent Source for Enterprise Java.* [En línea] 2007.

http://www.oreillynet.com/onjava/blog/2007/09/red_hat_developer_studio_good.html.

Cabrera, María Evelia Casales. [En línea] [Citado el: 18 de 11 de 2010.]

<http://hp.fcencias.unam.mx/~alg/bd/dwh.pdf..>

- Cáceres Cruz, Yander y Alea Boffill, Adonys.** *Análisis, Diseño e Implementación del Módulo de Investigaciones Internas del SIIPOL.* Ciudad Habana : s.n., 2009.
- Calisoft. [En línea] [Citado el: 24 de 5 de 2011.] <http://calisoft.uci.cu/tmp/documentos/normas/iso/NC-ISO-IEC%209126-1.pdf>.
- Caramaza.** *Tecnologías y Metodologías para la Construcción del Sistema de Gestión del Conocimiento.* Madrid : s.n., 2005.
- Control de Calidad de Software- Fase de Pruebas. [En línea] [Citado el: 24 de 5 de 2011.] <http://www.manuelmorales.net/?p=1426>.
- DataCleaner. [En línea] [Citado el: 25 de 11 de 2010.] <http://datacleaner.eobjects.org/> .
- Díaz, Yohanlena Hartman, Dailen Ramón Zequeira.** *Implementación del proceso de extracción, transformación y carga en un almacén de datos operacional para CIMEX.* 2009-2010. Tesis(Ingeniero en Ciencias Informáticas).
- EcuRed. [En línea] http://www.ecured.cu/index.php/Herramienta_CASE.
- EcuRed. [En línea] http://www.ecured.cu/index.php/Prueba_de_RUP.
- EL RINCÓN DEL VAGO. [En línea] [Citado el: 28 de 1 de 2011.] <http://html.rincondelvago.com/industria-comercio-y-servicios.html>.
- Espinoza, Humberto.** [En línea] 2005. [Citado el: 23 de 11 de 2010.] http://www.lgs.com.ve/pres/PresentacionES_PSQL.pdf.
- Flores., GestioPolis Lic. Edwin Osmil Coreas.** GestioPolis. [En línea] 2006. [Citado el: 15 de 11 de 2010.]
- Free Download Manager. [En línea] [Citado el: 26 de 11 de 2010.] http://www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_%28M%C3%8D%29_14720_p/.
- Georjina, Yadiurvis Aguilar Moya y Ana.** *Sistema de Infomación de Gobierno: Mercado de datos para la actividad del Comercio.* 2010.
- IBM Corporation & Sun Microsystems, Inc.** Eclipse Platform Technical Overview. *Sitio web oficial de la comunidad de desarrollo Eclipse.* [En línea] <http://www.eclipse.org/whitepapers/eclipse-overview.pdf>.
- Ing. Alberto Límia Navarro¹, Ing. Mabel Medina Rodríguez², Ing. Daulemys Rigo Portillo³, Ing. Madelys Cuesta Villa⁴, Ing. Yamila Mateu Romero⁵, Ing. Julio Ernesto Ortiz⁶, Ing. Osniel Hernández Calvo⁷, Ing. Marisleydis Socas Alvez⁸ , Anisley Delfino Rodríg.** *METODOLOGÍA PARA EL DESARROLLO DE SOLUCIONES DE ALMACENES DE DATOS E INTELIGENCIA DE NEGOCIO EN CENTALAD.*
- Ing.Lilliam Vega Torres, Ing.Luis Rojas Díaz,Lic.Cecilia Placeres Villar.** *I3.* Ciudad de la Habana, Cuba : s.n., Noviembre 2008.
- Jacobson, Ivar, Booch, Grady y Rumbaugh, James.** *El proceso unificado de desarrollo de Software.* s.l. : Editorial Felix Varela, 1999.
- JSON. *Sitio oficial.* [En línea] <http://www.json.org/json-es.html>.
- Lago, Ramiro.** Introducción a JSF. [En línea] 2007. <http://www.proactiva-calidad.com/java/jsf/introduccion.html>.
- Lamancha, Beatriz Pérez.** *Tesis de Maestría en Informática: Proceso de Testing Funcional Independiente.* Montevideo, Uruguay : s.n., 2006.
- Larman, Craig.** *UML y Patrones. Introducción al análisis y diseño orientado a objetos.* s.l. : México: Prentice Hall, 1999.

- Mann, Kito D.** *JavaServer Faces in Action*. s.l. : Greenwich: Manning Publications Co., 2005.
- Menéndez, Dr. Eugenio Santos.** [En línea] [Citado el: 16 de 11 de 2010.]
http://www.oei.eui.upm.es/Asignaturas/BD/DYOBDEjemplo_DW.pdf..
- Nader, Ing. Javier.** "Sistema de apoyo gerencial universitario" . 2003. Tesis de Magister en Ingeniería de Software.
- Nader, Ing. Javier.** "Sistema de apoyo gerencial universitario" . 2003. Tesis de Magister en Ingeniería de Software.
- Object Management Group.** UML. [En línea] <http://www.uml.org/>.
- Pentaho Shema Workbench.** [En línea] [Citado el: 21 de 11 de 2010.]
<http://churriwifi.wordpress.com/2010/07/04/17-3-preparando-el-analisis-dimensional-definicion-de-cubos-utilizando-schema-workbench/>.
- Pereira Ojeda, Maikel y Gago Martínez, Yudanis.** *Análisis, Diseño e Implementación del módulo Experticias Criminales del Sistema de Investigación e Información Policial*. Ciudad Habana : s.n., 2008.
- Prix, Data.** [En línea] [Citado el: 19 de 11 de 2010.]
<http://www.dataprix.com/blogs/respinosamilla/cubos-olap-line-analytic-processing..>
- Prix, Data.** [En línea] [Citado el: 23 de 11 de 2010.]
<http://www.dataprix.com/blogs/respinosamilla/cubos-olap-line-analytic-processing..>
- Project., Pentaho Open Source Business Intelligence: Kettle.** [En línea] 25 de Enero de 2010.
[Citado el: 2010 de 11 de 2010.] <http://kettle.pentaho.org..>
- Ralph Kimball y Ross, Margy.** *The Data Warehouse Toolkit*. . EUA : Wiley Publishing Inc, 2002.
- Rivero Guevara, Humberto.** *Análisis, Diseño e Implementación del Módulo Aprehensión del SIIPOL*. Ciudad Habana : s.n., 2008.
- Rodríguez Brito, Dayami.** *Plan de Aseguramiento de la Calidad*. Ciudad de la Habana : s.n., 2011.
- Rumbaugh, James, Jacobson, Ivar y Booch, Grady.** *Lenguaje Unificado de Modelado. Manual de Referencia*. s.l. : Madrid: Pearson Educación, 2000.
- Sánchez, Leopoldo Zenaido Zepeda.** *Metodología para el Diseño Conceptual de Almacenes de Datos*. . Universidad Politécnica de Valencia. Valencia, : s.n., 2008. . pág. 211, Tesis de Doctorado.
- SOA agenda.** [En línea] <http://soaagenda.com/journal/articulos/que-son-los-frameworks/>.
- Torres, Liudmila Padrón.** mailxmailx. [En línea] [Citado el: 17 de 11 de 2010.]
<http://www.mailxmail.com/curso-almacenes-datos-importancia-estandar/caracteristicas-almacen-datos>.
- Vázquez Pérez, Amed y Galvez Alonso, Arlan.** *Análisis, Diseño e Implementación del submódulo Evidencia del módulo Registro y Control del SIIPOL*. Ciudad Habana : s.n., 2009.
- Velasco, Roberto Hernando.** [En línea] 16 de 11 de 2010.
<http://www.rhernando.net/modules/tutorials/doc/bd/dw.pdf..>
- Visual Paradigm.** *Sitio oficial Visual Paradigm*. [En línea] <http://www.visual-paradigm.com>.
- Wilson, Leslie B.** *Comparative Programming Languages, Second Edition*.

Referencias Bibliográficas

1. [En línea] [Citado el: 28 de 1 de 2011.]
<http://www.wordreference.com/definicion/comercio%20interior>.
2. BIBLIOTECA VIRTUAL de Derecho, Economía y Ciencias Sociales_Álvaro de la Helguera y García. *eumed.net*. [En línea] [Citado el: 28 de 1 de 2011.]
<http://www.eumed.net/libros/2006a/ah-circ/1e.htm>.
3. EL RINCÓN DEL VAGO. [En línea] [Citado el: 28 de 1 de 2011.]
<http://html.rincondelvago.com/industria-comercio-y-servicios.html>.
4. **Flores., GestioPolis Lic. Edwin Osmil Coreas.** GestioPolis. [En línea] 2006. [Citado el: 15 de 11 de 2010.]
5. **Menéndez, Dr. Eugenio Santos.** [En línea] [Citado el: 16 de 11 de 2010.]
http://www.oei.eui.upm.es/Asignaturas/BD/DYOBBD/Ejemplo_DW.pdf..
6. **Velasco, Roberto Hernando.** [En línea] 16 de 11 de 2010.
<http://www.rhernando.net/modules/tutorials/doc/bd/dw.pdf..>
7. **Torres, Liudmila Padrón.** mailxmailx. [En línea] [Citado el: 17 de 11 de 2010.]
<http://www.mailxmail.com/curso-almacenes-datos-importancia-estandar/caracteristicas-almacen-datos>.
8. **Cabrera, María Evelia Casales.** [En línea] [Citado el: 18 de 11 de 2010.]
<http://hp.fciencias.unam.mx/~alg/bd/dwh.pdf..>
9. **Nader, Ing. Javier.** "*Sistema de apoyo gerencial universitario*". 2003. Tesis de Magister en Ingeniería de Software.
10. **Ralph Kimball y Ross, Margy.** *The Data Warehouse Toolkit*. . EUA : Wiley Publishing Inc, 2002.
11. **Prix, Data.** [En línea] [Citado el: 19 de 11 de 2010.]
<http://www.dataprix.com/blogs/respinosamilla/cubos-olap-line-analytic-processing..>
12. **BI, El rincón del.** [En línea] [Citado el: 20 de 11 de 2010.]
[http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimensional/..](http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimensional/)
13. [En línea] [Citado el: 2 de 12 de 2010.]
http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/rosette_u_jm/capitulo3.pdf.
14. [En línea] [Citado el: 2 de 12 de 2010.]
<http://bdigital.eafit.edu.co/bdigital/PROYECTO/P005.74CDA779/marcoTeorico.pdf>.
15. [En línea] [Citado el: 2 de 12 de 2010.] http://etl-tools.info/es/bi/almacenedatos_esquema-constelacion.htm.
16. **Nader, Ing. Javier.** "*Sistema de apoyo gerencial universitario*". 2003. Tesis de Magister en Ingeniería de Software.
17. [En línea] [Citado el: 21 de 11 de 2010.]
<http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/OLAP.pdf..>

18. **Prix, Data.** [En línea] [Citado el: 23 de 11 de 2010.]
<http://www.dataprix.com/blogs/respinosamilla/cubos-olap-line-analytic-processing..>
19. [En línea] http://www.google.com/cu/#hl=es-419&biw=1024&bih=575&q=etapas+de+un+almacen+de+datos&aq=f&aql=&aql=&oq=&gs_rfai=&fp=666b36a6e513e2ed .
20. **Díaz, Yohanlena Hartman, Dailen Ramón Zequeira.** *Implementación del proceso de extracción, transformación y carga en un almacén de datos operacional para CIMEX.* 2009-2010. Tesis(Ingeniero en Ciencias Informáticas).
21. [En línea] 27 de Septiembre de 2008.
<http://gerenciainnovacion.blogspot.com/2008/09/conceptos-inteligencia-de-negocios.html>.
22. **Ing. Alberto Límia Navarro¹, Ing. Mabel Medina Rodríguez², Ing. Daulemys Rigo Portillo³, Ing. Madelys Cuesta Villa⁴, Ing. Yamila Mateu Romero⁵, Ing. Julio Ernesto Ortiz⁶, Ing. Osniel Hernandez Calvo⁷, Ing. Marisleydis Socas Alvez⁸ , Anisley Delfino Rodríg.** *METODOLOGÍA PARA EL DESARROLLO DE SOLUCIONES DE ALMACENES DE DATOS E INTELIGENCIA DE NEGOCIO EN CENTALAD.*
23. **Sánchez, Leopoldo Zenaido Zepeda.** *Metodología para el Diseño Conceptual de Almacenes de Datos.* . Universidad Politécnica de Valencia. Valencia, : s.n., 2008. . pág. 211, Tesis de Doctorado.
24. Free Download Manager. [En línea] [Citado el: 26 de 11 de 2010.]
[http://www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_%28M%C3%8D%29_14720_p/.](http://www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_%28M%C3%8D%29_14720_p/)
25. **Espinoza, Humberto.** [En línea] 2005. [Citado el: 23 de 11 de 2010.]
http://www.lgs.com.ve/pres/PresentacionES_PSQL.pdf.
26. [En línea] [Citado el: 27 de 11 de 2010.]
<http://www.ibiblio.org/pub/linux/docs/LuCaS/Postgresql-es/web/navegable/todopostgresql/app-pgadmin.html>..
27. **Project., Pentaho Open Source Business Intelligence: Kettle.** [En línea] 25 de Enero de 2010. [Citado el: 2010 de 11 de 2010.] <http://kettle.pentaho.org>..
28. DataCleaner. [En línea] [Citado el: 25 de 11 de 2010.] <http://datacleaner.eobjects.org/> .
29. Pentaho Shema Worbench. [En línea] [Citado el: 21 de 11 de 2010.]
<http://churriwifi.wordpress.com/2010/07/04/17-3-preparando-el-analisis-dimensional-definicion-de-cubos-utilizando-schema-workbench/>.
30. **2005, "Pentaho BI Suite Enterprise Edition."** [En línea] PentahoCorporation, 2005. [Citado el: 29 de 11 de 2010.] http://www.pentaho.com/products/data_services/?hp=y..
31. **Ing.Lilliam Vega Torres, Ing.Luis Rojas Díaz,Lic.Cecilia Placeres Villar.** *I3.* Ciudad de la Habana, Cuba : s.n., Noviembre 2008.
32. Apache TomCat. [En línea] <http://tomcat.apache.org/>..
33. **Caramaza.** *Tecnologías y Metodologías para la Construcción del Sistema de Gestión del Conocimiento.* Madrid : s.n., 2005.

34. [En línea] 2009. [Citado el: 19 de Junio de 2011.] <http://msdn.microsoft.com/es-es/library/ms145514.aspx>.
35. **Lamancha, Beatriz Pérez.** *Tesis de Maestría en Informática: Proceso de Testing Funcional Independiente*. Montevideo, Uruguay : s.n., 2006.
36. Calisoft. [En línea] [Citado el: 24 de 5 de 2011.] <http://calisoft.uci.cu/tmp/documentos/normas/iso/NC-ISO-IEC%209126-1.pdf>.
37. Control de Calidad de Software- Fase de Pruebas. [En línea] [Citado el: 24 de 5 de 2011.] <http://www.manuelmorales.net/?p=1426>.
38. **Georjina, Yadiurvis Aguilar Moya y Ana.** *Sistema de Infomación de Gobierno: Mercado de datos para la actividad del Comercio*. 2010.

Anexos

Tabla 6: Caso de prueba Consultar ingresos por tipos de mercado

Escenario	Descripción	Perfil de análisis	Indicadores a medir	Respuesta del sistema	Flujo central
EC 1.1 Ingresos por ventas, tipos de mercados del Comercio interior	Permite visualizar el reporte con las variables presentes en el mismo	-Indicadores de Servicios -Comercios -Nivel Territorial -NAE -Meses -Entidad -Unidad de medida	-Ingresos según el comercio	Se muestra la tabla con los valores correspondientes a cada escenario.	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 navegador. Se selecciona las áreas de análisis de AA.Comercio interior.
EC 1.2 Ingresos por ventas, tipos de mercados por el total en ambas monedas		-Indicadores de Servicios -Nivel Territorial -NAE -Meses -Entidad -Unidad de media	-Total de ingresos -Total de ingresos en CUC		Se selecciona el libro de trabajo de LT. Ingresos por ventas de bienes de consumo y servicios a la población por tipos de mercados del comercio interior. En la parte inferior izquierda se selecciona el reporte deseado. En el área de trabajo se visualiza la tabla correspondiente al reporte.

Tabla 7: Lista de chequeo

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)				
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?				
crítico	2. ¿Los reportes son configurables a través de la interfaz del sistema?				
	3. ¿La interfaz está orientada a facilitar el uso de las funciones del sistema por parte de los usuarios?				
crítico	4. ¿No existen restricciones para construir cubos OLAP con dimensiones y niveles de agregación ilimitados?				
crítico	5. ¿Los usuarios son capaces de manipular los resultados de manera que se ajusten a sus necesidades,				

	conformando nuevos reportes?				
	6. ¿El sistema responde de una forma rápida a la información que le sea solicitada por el usuario?				
	7. ¿El sistema refleja cualquier lógica del negocio para poder responder a preguntas específicas?				
crítico	8. ¿El sistema garantiza la confidencialidad y seguridad de acceso a los datos por rol de los usuarios?				
	9. ¿Los datos e información derivados del proceso de análisis realizado mediante la aplicación, apoyan la toma de decisiones en la Institución?				
crítico	10. ¿Los cambios en los datos se reflejan automáticamente en los reportes de forma instantánea?				
Semántica del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
crítico	1. ¿Se han identificado errores ortográficos en los entregables?				
crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?				
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?				

Glosario

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

Jerarquía: implica una organización de niveles dentro de una dimensión, donde cada nivel representa el total agregado de los datos del nivel inferior.

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: defecto, error o sugerencia que se le hace al equipo de desarrollo una vez encontrada alguna dificultad en lo que se está evaluando.

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

Código abierto: término con el que se conoce al software distribuido y desarrollado libremente.

Matriz bus: es donde se muestra en forma de matriz la relación que existe entre los hechos y las dimensiones.

Caso de Uso (CU): es una secuencia de interacciones que se desarrollarán entre un sistema y sus actores en respuesta a un evento que inicia un actor principal sobre el propio sistema.

Lenguaje Unificado de Modelado (UML): lenguaje visual para especificar, construir y documentar un sistema de software.

Ministerio del Comercio Interior (MINCIN): es el ministerio que maneja todo lo relacionado con el Comercio Interior en el país.