

Universidad de las Ciencias Informáticas

Facultad 6



Título: Sistema de Información de Gobierno. Mercado de datos Comercio exterior.

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

Autoras: Leidys Susel Mascareño Hodge
Patricia Peña Consuegra

Tutores: Ing. Esley León Valdés
Ing. Roberto Tellez Ibarra

Ciudad Habana, Junio de 2011

La inteligencia consiste no sólo en el conocimiento, sino también en la destreza de aplicar los conocimientos en la práctica.

Aristóteles

DECLARACIÓN DE AUTORÍA

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

Para que así conste firmo la presente a los ____ días del mes de _____ del año _____.

Leidys Susel Mascareño Hodge

Patricia Peña Consuegra

Firma del Autor

Firma del Autor

Ing.Roberto Tellez Ibarra

Ing. Esley León Valdés

Firma del Tutor

Firma del Tutor

DATOS DE CONTACTO

Tutores:

Tutor: Ing. Roberto Tellez Ibarra
Especialidad de graduación: Ingeniería en Ciencias Informáticas
Categoría docente: Instructor Recién Graduado
Años de experiencia en el tema: 1
Años de graduado: 2
Correo Electrónico: rtibarra@uci.cu

Tutor: Ing. Esley León Valdés
Especialidad de graduación: Ingeniería en Ciencias Informáticas
Categoría docente: Instructor Recién Graduado
Años de experiencia en el tema: 1
Años de graduado: 2
Correo Electrónico: elvaldes@uci.cu

Patricia:

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

En primer lugar quiero agradecer a mis padres: a mi mami querida que ha sido y será mi sostén, mi vida, por su amor, mil gracias. A mi papá por su nobleza y entrega en cada una de las cosas que hace, por ser mi maestro y mi ejemplo.

Agradezco a mi segunda madre, mi tía Haydee, es la persona que me ha hecho sentir más querida en toda mi vida. Me enseñó que debo aprender a valorarme como persona, de la misma manera en que valoro a otras. Que hay que darles a las personas el lugar que se merecen y no el lugar que una cree que tienen. Por demostrarme, al igual que mis padres, que no existe dificultad alguna cuando a mi lado la familia está.

Quiero agradecer a mis abuelos que me han apoyado en todo momento, y han estado a mi lado en todas y cada una de mis decisiones.

A mi hermano Marquito es lo que más quiero en mi vida, porque aun siendo el menor es mas celoso conmigo que nadie. A mis hermanas Daína, Laritza y Haydee, por su amor constante.

Gracias a mi tía Delkis por sus atenciones y preocupaciones durante los últimos cinco años. A mi tío más querido, a mis tías, mis primas y primos, los quiero mucho.

A los tutores por su apoyo en todo momento. A Tellez, no solo fuiste tutor, fuiste un amigo, gracias por las muchas horas del día que nos dedicaste.

Quiero dar las gracias a una persona muy querida, por estar a mi lado en los momentos más tristes y felices, por ser ejemplo en el estudio, por apoyarme en cada paso que doy y darme los ánimos necesarios para dar los siguientes, mi gran amiga: Laritza.

Muchas gracias a Susy, mi compañera de tesis, todo el tiempo fuiste la guía y el motor impulsor de este trabajo, también una buena amiga. A Alejandro, gracias por decirme siempre SI cuando necesité tu ayuda. A los profesores y amigos del departamento Almacenes de Datos, de quienes aprendí mucho.

Doy mis más sinceros agradecimientos a mis compañeros de año, a todos muchas gracias por permitirme compartir tantos momentos junto a ustedes. A las personas que siempre confiaron en mí y siempre tuvieron un gesto de preocupación por mis cosas. A todos los que hicieron posible este sueño. A las personas que no se encuentran físicamente en mi vida pero si en mi corazón: abuelo Oscar, tía Rayma y tía Sory.

Leidys Susel:

Quiero agradecer a mi mamá por todo su amor y cariño, por guiarme siempre en la vida ante las buenas y malas situaciones. Ella es mi razón de ser y mi motivación para seguir adelante, mamá te amo.

A mis dos papás por siempre confiar en mí, por darme apoyo en todos los momentos de mi vida, no quiero resaltar a uno por encima del otro para que no se pongan celosos, en realidad los amo a los dos por igual y significan mucho para mí.

A mi abuelita Ñingo por haberme dado la madre maravillosa que tengo y por estar siempre al tanto de mi vida.

A mi tía Martica por ser la mejor tía del mundo, ella siempre se ha preocupado por mí como si fuera su hija y se ha encargado de que no me falte nada.

A mi prima Mailén por ser tan especial y tan natural, por ser mi guía durante estos 5 años, por darme siempre la esperanza de seguir adelante y estar siempre presente para mí en todo momento.

A todos mis primos, mis tíos, mis abuelos, en fin, a toda mi familia en general por el apoyo que siempre me han dado y por preocuparse por mí.

Quiero agradecer a Yula, una de mis mejores amigas, por ser mi ejemplo a seguir.

A Laidy que es mi hermanito más chiquito y mi mejor amiga, por estar presente en los últimos años de la universidad, por todo lo que hemos compartido en los buenos momentos y todo el apoyo que me ha brindado en los días de tristeza, especialmente en estos últimos meses. Ella siempre ha entendido todas mis molestias, mis inquietudes y ha soportado mis cambios de carácter como una verdadera hermana.

Agradecerle de todo corazón a Tellez porque a pesar de que es nuestro tutor, para mí ha sido más que eso. Me ha acompañado siempre como un verdadero amigo y aunque en ocasiones me ha mortificado un poquito mandándome a hacer cosas con el cable de red, siempre he podido contar con su apoyo incondicional en todo momento.

A las mejores compañeras de cuarto que he tenido, Maryin, Lisandra y Lipsy, para mi han sido más que amigas, han sido hermanas.

A Patri por ser la mejor CT del mundo, creo que sin su paciencia y su apoyo nada de esto hubiera sido posible. También ha sido una excelente amiga.

Quiero agradecer también a José por todo el apoyo que me brindó durante los primeros años de la carrera, a Lupe por brindarme siempre su apoyo en todo momento y preocuparse por mí, espero ser para ella como una hija porque la quiero y la respeto como si fuera mi madre.

A toda la gente del proyecto especialmente a Leonel, a David, a Yoendy y a Yulio porque pude contar con ellos para todo, nunca obtuve un no de su parte, gracias mis amores.

A Yanisleidy porque siempre estuvo ahí para nosotras, dándonos sugerencias y acompañándonos en todo momento, formando un trío de tesis explosivo.

Agradezco a todas mis amistades en sentido general, no quiero decir nombres por temor a que se me quede alguien sin mencionar.

Patricia

A mi familia.

Leidys Susel

A mi madre adorada, que es mi razón de ser y mi fuerza para seguir adelante.

A mis dos padres, que los amo con locura, gracias por siempre confiar en mi y brindarme todo el apoyo del mundo.

A los tres pedacitos de mi vida, mis niñas Edisleidys, Arianna y Ailenis, me encantaría ser para ellas un ejemplo a seguir.

A mis hermanitas del alma, Normi, Dayi, Dunia y Dailín.

En fin, a toda mi familia, por todo su amor y preocupación.

La Oficina Nacional de Estadísticas (ONE) es la entidad encargada de procesar y controlar la información estadística en Cuba. Está compuesta por varias áreas, siendo una de ellas Comercio exterior, encargada de controlar la información referente a las importaciones y exportaciones efectuadas. En esta área los datos son almacenados en diferentes formatos; el volumen de información crece constantemente y existen múltiples versiones de los datos influyendo negativamente en la posibilidad que estos estén integrados, estandarizados y que posean buena calidad. En la presente investigación se propone una solución basada en el desarrollo de un mercado de datos que contribuya en el proceso de toma de decisiones de los directivos de Comercio exterior en la ONE. Para ello se realizó el análisis, diseño, implementación y prueba de la aplicación que constituye el resultado principal de este trabajo de diploma. Conjuntamente se hizo uso de varias herramientas y de una metodología para guiar el desarrollo de la solución. Se obtuvo un mercado de datos poblado y funcional, con una capa de inteligencia de negocio que brinda vistas de análisis actualizadas, permitiéndole a los especialistas del Comercio exterior un mejor estudio de la información.

Palabras clave: Mercado de datos, inteligencia de negocio.

TABLA DE CONTENIDO

INTRODUCCIÓN.....	1
CAPÍTULO 1: LOS ALMACENES DE DATOS.....	4
1.1 Almacenes de datos.....	4
1.1.1 Características, ventajas y desventajas de los almacenes de datos.....	5
1.2 Modelo multidimensional.....	6
1.2.1 Características del modelo multidimensional.....	6
1.2.2 Esquemas multidimensionales.....	7
1.3 Procesamiento Analítico en Línea.....	9
1.4 Mercado de datos o Datamart.....	10
1.4.1 Características y ventajas de los mercados de datos.....	11
1.5 Procesos básicos de un almacén de datos.....	12
1.5.1 Integración de datos.....	12
1.5.2 Inteligencia de negocio.....	14
1.6 Metodología de desarrollo.....	16
1.7 Herramientas seleccionadas para la solución.....	17
1.7.1 Herramienta de modelado.....	17
1.7.2 Sistema Gestor de Bases de Datos.....	18
1.7.3 Herramientas para la Integración de datos.....	19
1.7.4 Herramientas para Inteligencia de negocio.....	22
1.8 Conclusiones del capítulo.....	23
CAPÍTULO 2: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS COMERCIO EXTERIOR.....	24
2.1 Análisis del mercado de datos Comercio exterior.....	24
2.1.1 Estudio preliminar del negocio.....	24
2.1.2 Temas de análisis.....	24
2.1.3 Necesidades de los usuarios.....	25
2.1.4 Casos de Uso del Sistema.....	27
2.1.5 Reglas de Negocio.....	29
2.2 Diseño del mercado de datos Comercio exterior.....	30
2.2.1 Dimensiones, Hechos y Medidas.....	30
2.2.2 Matriz dimensional.....	36
2.2.3 Modelo de datos.....	36
2.2.4 Política de respaldo y recuperación.....	38
2.2.5 Esquema de seguridad.....	38
CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBA DEL MERCADO DE DATOS COMERCIO EXTERIOR.....	41

3.1 Implementación de la base de datos	41
3.1.1 Modelo de datos físico.....	41
3.1.2 Roles y privilegios establecidos en la base de datos	44
3.2 Implementación del subsistema de integración	44
3.2.1 Arquitectura de Integración	44
3.2.2 Proceso de Integración de datos	45
3.2.3 Trazas de las cargas de los datos	50
3.3 Implementación del subsistema de visualización	50
3.3.1 Cubos multidimensionales.....	50
3.3.2 Arquitectura de Información	53
3.4 Resultados obtenidos.....	55
3.5 Validación de la solución.....	56
3.5.1 Pruebas aplicadas al mercado de datos Comercio exterior	56
3.6 Conclusiones del capítulo	58
CONCLUSIONES GENERALES.....	60
RECOMENDACIONES	61
REFERENCIAS BIBLIOGRÁFICAS.....	62
BIBLIOGRAFÍA	64
GLOSARIO DE TÉRMINOS	66

ÍNDICE DE TABLAS

Tabla 1 Descripción de las dimensiones	32
Tabla 2 Descripción de los hechos	34
Tabla 3 Descripción de las medidas	35
Tabla 4 Matriz dimensional	36
Tabla 5 Usuarios en la base de datos	38
Tabla 6 Usuarios en la aplicación	39
Tabla 7 Convenciones de nombrado	41
Tabla 8 Correspondencia entre esquemas y tablas del mercado de datos.	43

ÍNDICE DE FIGURAS

Figura 1 Ejemplo de esquema estrella	7
Figura 2 Ejemplo de esquema snowflake o copo de nieve	8
Figura 3 Ejemplo de esquema constelación de hechos	9
Figura 4 Arquitectura de Pentaho Data Integration	21
Figura 5 Diagrama de Casos de Uso del Sistema.....	29
Figura 6 Modelo de datos	37
Figura 7 Arquitectura del proceso de integración.	45
Figura 8 Reporte estándar de medidas.	46
Figura 9 Reporte análisis de cadena.....	46
Figura 10 Reporte análisis numérico.....	47
Figura 11 Transformación para la carga de la dimensión aduana.	47
Figura 12 Transformación para la carga del hecho exportación	48
Figura 13 Trabajo del mercado de datos.....	49
Figura 14 Diseño de los cubos multidimensionales.....	51
Figura 15 Elementos del cubo importación	52
Figura 16 Elementos del cubo exportación	52
Figura 17 Elementos del cubo intercambio	53
Figura 18 Arquitectura de información	53
Figura 19 Mercancías agrupadas en GCE por la CUCI.....	55

INTRODUCCIÓN

El desarrollo de las Tecnologías de la Informática y las Comunicaciones (TIC) y sus grandes avances tecnológicos han causado un aumento considerable de la información. Todo ese volumen de datos es almacenado principalmente por las empresas para utilizarlo en dependencia de sus necesidades y aprovechar la información útil para la toma de decisiones. Pero por lo general este proceso se dificulta por el manejo de grandes volúmenes de información de la forma más óptima posible.

Los datos constituyen un eslabón de gran importancia para el creciente desarrollo que se evidencia, pero los mismos por sí solos no aportan ningún beneficio a la entidad u organización que los posea. Para que sean útiles se deben convertir en información y en conocimiento que contribuyan al buen desarrollo del negocio; para esto, las empresas deben realizar análisis sobre los mismos aplicando diversas tecnologías con el fin de reducir la incertidumbre sobre algún aspecto de la realidad y dar soporte a los procesos básicos de la organización.

En Cuba se encuentra la Oficina Nacional de Estadísticas (ONE), entidad encargada de recoger, controlar y emitir información estadística del país. La fuente fundamental de los datos que se exponen procede del Sistema Estadístico Nacional (SEN) donde colaboran órganos, organismos e instituciones de la nación cubana.

La Universidad de las Ciencias Informática (UCI) se encuentra desarrollando un producto para dicha entidad, este se nombra Sistema de Información de Gobierno, el cual está compuesto por módulos. Uno de estos es un almacén de datos que contiene varios mercados de datos correspondientes con las diferentes áreas socioeconómicas de la isla. Una de estas áreas es Comercio exterior, que se ocupa de controlar la información referente a las importaciones y exportaciones efectuadas en Cuba. Anteriormente en la UCI, se realizó un análisis y diseño para la elaboración de un mercado de datos que almacenara los indicadores estadísticos referentes al Comercio exterior, pero debido a la presencia de nuevas necesidades y exigencias de los usuarios de dicha área, se hizo necesario realizar el refinamiento del análisis y diseño existentes. En dicha área se detectaron los siguientes problemas: los datos que se registran están contenidos en archivos de diferentes formatos, que solo pueden ser consultados por especialistas de informática e información que posean conocimientos del negocio; el volumen de los datos crece constantemente, lo que dificulta el análisis de los resultados a las consultas que se realicen; la existencia de múltiples versiones de los datos interviene negativamente en la posibilidad de que estos se encuentren integrados y que posean buena calidad.

Además no existe una aplicación informática que brinde reportes flexibles con información actualizada que permita apoyar el proceso de toma de decisiones de los directivos del Comercio exterior en Cuba.

Teniendo en cuenta lo planteado anteriormente, surge como **problema de investigación**: ¿Cómo contribuir a la toma de decisiones del área Comercio exterior del Sistema de Información de Gobierno? El problema planteado delimita como **objeto de estudio**: los almacenes de datos, enmarcado en el **campo de acción**: mercado de datos para el área Comercio exterior del Sistema de Información de Gobierno.

Para dar respuesta al problema de investigación se plantea como **objetivo general**: Desarrollar el mercado de datos Comercio exterior del Sistema de Información de Gobierno que contribuya a la toma de decisiones. Del mismo se derivan los siguientes **objetivos específicos**:

1. Refinar el análisis y diseño del mercado de datos del área Comercio exterior.
2. Implementar el mercado de datos del área Comercio exterior.
3. Validar el mercado de datos del área Comercio exterior.

Para darle cumplimiento a los objetivos propuestos se plantearon las siguientes **tareas de investigación**:

1. Caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.
2. Refinamiento de los requisitos.
3. Refinamiento de los hechos, las medidas y las dimensiones del mercado de datos.
4. Refinamiento del modelo de datos.
5. Refinamiento de la arquitectura del mercado de datos.
6. Diseño del subsistema de integración.
7. Diseño del subsistema de visualización.
8. Diseño de los casos de pruebas.
9. Implementación de la base de datos.
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.

Como resultado se espera obtener un mercado de datos que contenga la información consolidada del área Comercio exterior, con el propósito de facilitar el control estadístico de la información y mejorar el proceso de toma de decisiones para los directivos del Comercio exterior en Cuba. El presente trabajo consta de 3 capítulos, estructurados de la siguiente manera:

Capítulo 1: Los almacenes de datos

Durante el desarrollo de este capítulo se abordan diversos elementos teóricos relacionados con los almacenes de datos y los mercados de datos, así como de los procesos por los que están compuestos los mismos. Se describen todas las herramientas a utilizar para dar solución al problema planteado. Se realiza una revisión bibliográfica de las metodologías presentes a nivel mundial y en Cuba para el desarrollo de los almacenes de datos.

Capítulo 2: Análisis y diseño del mercado de datos Comercio exterior

Durante el desarrollo de este capítulo se realiza una revisión de los requisitos, los casos de uso, los hechos, las medidas y las dimensiones del mercado de datos identificados en la tesis precedente, para realizar el refinamiento de los mismos. Una vez concluida esta etapa se construye el modelo de datos correspondiente y se diseña el subsistema de integración, el de visualización y los casos de prueba.

Capítulo 3: Implementación y prueba del mercado de datos Comercio exterior

El presente capítulo comprende la implementación de la base de datos, del subsistema de integración y del subsistema de visualización. Mediante los procesos Extracción, Transformación y Carga e Inteligencia de Negocio se le dará solución a las necesidades del negocio. Además se plasma la validación de la solución propuesta mediante las listas de chequeo y la aplicación de los casos de pruebas.

CAPÍTULO 1: LOS ALMACENES DE DATOS

Durante el desarrollo del presente capítulo se abordan diversos elementos teóricos relacionados con los almacenes de datos y los mercados de datos, así como de los procesos por los que están compuestos los mismos. Se describen todas las herramientas a utilizar para dar solución al problema planteado. Se realiza una revisión bibliográfica de las metodologías presentes a nivel mundial y en Cuba para el desarrollo de los almacenes de datos.

1.1 Almacenes de datos

La evolución tecnológica existente actualmente ha supuesto un gran desarrollo en las tecnologías de la información. En las empresas y organismos de todo el mundo existe un gran cúmulo de información, por lo que nace la necesidad de digitalizar y almacenar la misma con el fin de que su manipulación y análisis sean más eficientes. Para lograr satisfacer dicha necesidad surgen los almacenes de datos.

Existen múltiples definiciones de los almacenes de datos, a continuación se citan algunas de ellas:

Un Almacén de Datos o Datawarehouse (DW), es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta. (Sinnexus 2008)

Es además un almacén de información temática orientado a cubrir las necesidades de aplicaciones de los Sistemas de Soporte de Decisiones (DSS) y de la Información de Ejecutivos (EIS), que permite acceder a la información corporativa para la gestión, control y apoyo a la toma de decisiones. (Torres 2006)

También es: una colección de datos orientada a temas o materias, integrada, variable en el tiempo y no volátil que será utilizada fundamentalmente en el proceso de toma de decisiones. (Inmon 2005)

Según Ralph Kimball en la segunda edición de su libro *The Datawarehouse Toolkit*, un almacén de datos es una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis. (Kimball and Ross 2002)

Luego de haber analizado diferentes bibliografías relacionadas con la disciplina, se ha observado que diversos autores definen el término almacén de datos de forma muy particular y expresando lo que

significa para ellos, pero el hecho es que a consideración de los autores, un almacén de datos es una colección de datos orientada a un tema específico, en la cual la información va a prevalecer de manera efectiva a lo largo del tiempo cuyo principal objetivo es dar soporte en el proceso de toma de decisiones a las empresas y organizaciones.

1.1.1 Características, ventajas y desventajas de los almacenes de datos

Bill H. Inmon, creador del término almacenes de datos y considerado como el padre de dicha disciplina, define algunas de sus principales características:

1. Temático: Los datos contenidos en el almacén están orientados a un tema o materia específicos. Esto posibilita a los usuarios finales un mejor entendimiento y acceso a la información. Además, las respuestas a los pedidos de información de un tema en particular se realizan de una forma más ágil ya que toda esta contenida en un mismo lugar.
2. Integrado: Los datos que se encuentran en el almacén están integrados en una estructura sólida. No contiene datos que no son de utilidad para los clientes, tampoco tiene datos redundantes.
3. No volátil: En el almacén de datos solo se realizan tres operaciones fundamentales: carga periódica, carga inicial o histórica y consulta. No se actualizan ni se eliminan los datos.
4. Histórico: Los datos contenidos en un almacén reflejan el estado de la actividad del negocio en el momento presente, a medida que el almacén tiene más datos será más útil pues brindará la posibilidad de establecer comparaciones en diferentes períodos. (Sinnexus 2008)
5. Contiene metadatos, es decir, datos sobre los datos. Los metadatos permiten saber la procedencia de la información, su periodicidad de actualización y su fiabilidad.

Ventajas y desventajas

Los almacenes de datos han abierto un mundo de posibilidades a las empresas, permitiéndoles mantener control sobre sus actividades. Algunas de las ventajas que propicia el uso de los almacenes de datos son:

1. Mediante la implementación de un almacén de datos se proporciona una herramienta que posibilita dar soporte al proceso de toma de decisiones a diversas empresas basándose en información general e integrada del negocio.

2. Es una solución a la cual se le puede sacar un gran provecho porque como contiene datos históricos, se puede aprender de estos y con este conocimiento se predicen situaciones que se presentarían en un futuro.
3. Permite a los usuarios medir los resultados de una mejor forma y a partir de esto, establecer prioridades en las acciones a realizar y decisiones a tomar con respecto a un determinado cliente.
4. En un contexto de objetivos definidos en los negocios permite a empresas, explorar automáticamente, visualizar y comprender los datos e identificar patrones, relaciones y dependencias que impactan en los resultados finales de la cuenta de resultados (tales como el aumento de los ingresos, incremento de los beneficios, contención de costes y gestión de riesgos).(Torres 2006)
5. El uso de los almacenes de datos en los negocios posibilita que los usuarios comprendan mejor el mismo, así como también se garantiza la ejecución de las consultas de manera rápida y eficiente.

Una de las principales desventajas de un proyecto de almacenes de datos es el costo. Los almacenes de datos por lo general no son estáticos, lo que mantenerlos resulta muy costoso. Además, es difícil la estandarización de los proyectos relacionados con los almacenes de datos debido a que cada entidad posee características diferentes lo que hace complejo el desarrollo de esta solución.

1.2 Modelo multidimensional

La tecnología de los almacenes de datos debido a su orientación analítica, impone un procesamiento y pensamiento distintos, la cual se sustenta por un modelamiento de bases de datos propio, conocido como Modelamiento multidimensional, el cual busca ofrecer al usuario su visión respecto de la operación del negocio. El modelamiento dimensional es una técnica para modelar bases de datos simples y entendibles al usuario final. La idea fundamental es que el usuario visualice fácilmente la relación que existe entre los distintos componentes del modelo.(Wolff 2002)

1.2.1 Características del modelo multidimensional

La estructura básica de un almacén de datos está definida por dos elementos principales, esquemas y tablas(Wolff 2002). Como en cualquier base de datos relacional, un DW se compone de tablas. Hay dos tipos básicos de tablas en el modelo multidimensional, las tablas de hecho y las dimensiones. Las

tablas de hechos contienen los valores de las medidas de negocios, por ejemplo: ventas promedio en dólares y número de unidades vendidas. Las tablas de dimensiones contienen el detalle de los valores que se encuentran asociados a la tabla hecho. Por otra parte, la colección de tablas en el almacén se conoce como esquema. Los esquemas se pueden agrupar dentro de tres categorías básicas: esquemas estrellas, esquemas snowflake o copo de nieve y esquema constelación de hechos.

1.2.2 Esquemas multidimensionales

Esquema estrella

El modelo multidimensional es conocido también como esquema estrella, ya que su estructura es similar: una tabla central y un conjunto de tablas que la atienden radialmente. El nombre proviene debido a que el diagrama forma una estrella, con puntos radiales desde el centro (Wolff 2002). El centro de la estrella consiste en una tabla de hecho y las puntas son las tablas de dimensiones. Las tablas de dimensiones sólo tienen conexión con la tabla hecho. En la figura 1, extraída de (Dario 2009), se muestra una imagen que representa un esquema estrella.

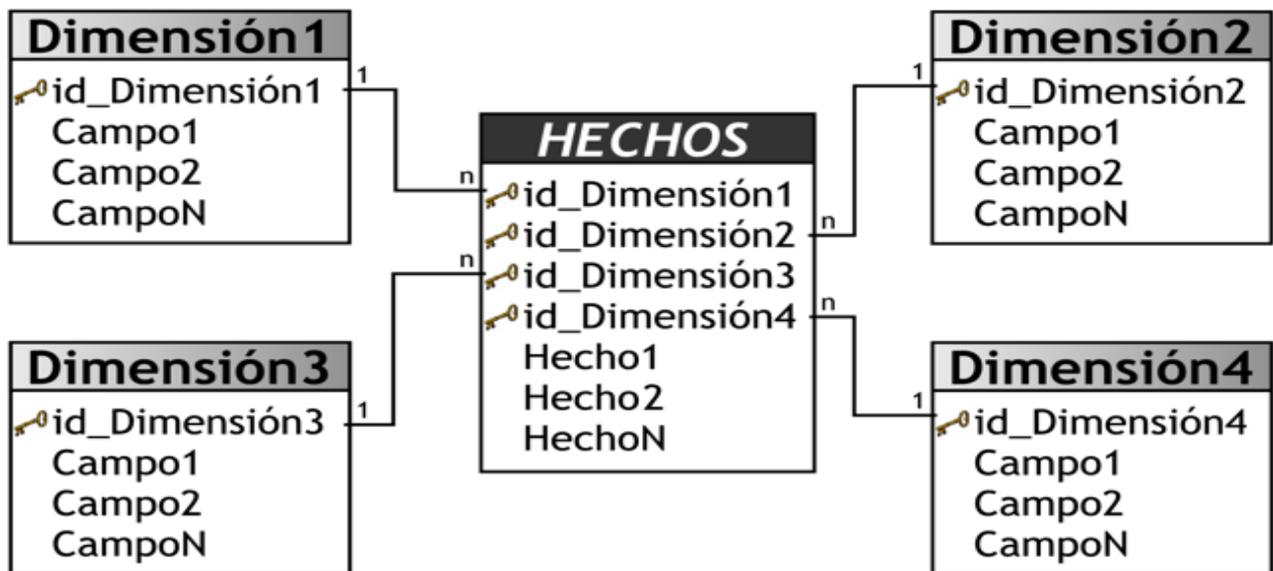


Figura 1 Ejemplo de esquema estrella

Esquema snowflake o copo de nieve

La diferencia que existe entre el esquema estrella y el copo de nieve es que la estructura de las tablas de dimensiones no es la misma. En el esquema copo de nieve las dimensiones están normalizadas, cada una posee sólo el nivel que es clave primaria en la tabla y la llave foránea de su parentesco del nivel más cercano. En la figura 2, extraída de (Dario 2009), se muestra una imagen que representa un esquema copo de nieve.

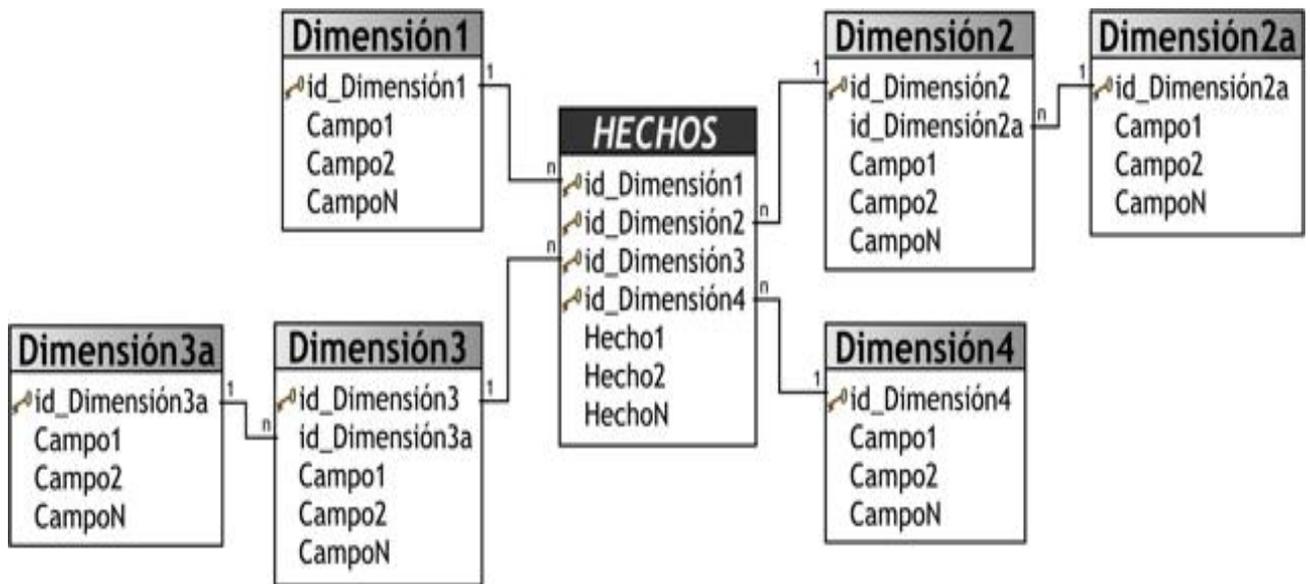


Figura 2 Ejemplo de esquema snowflake o copo de nieve

Esquema constelación de hechos

Su diseño y cualidades son muy similares a las del esquema en estrella, la diferencia radica en que la constelación de hechos contiene más de una tabla de hechos, por lo cual se podrán analizar más aspectos claves del negocio con un mínimo esfuerzo adicional de diseño y se reutilizan las tablas de dimensiones. En la figurara 3, extraída de (Dario 2009), se muestra una imagen que representa un esquema constelación de hechos.

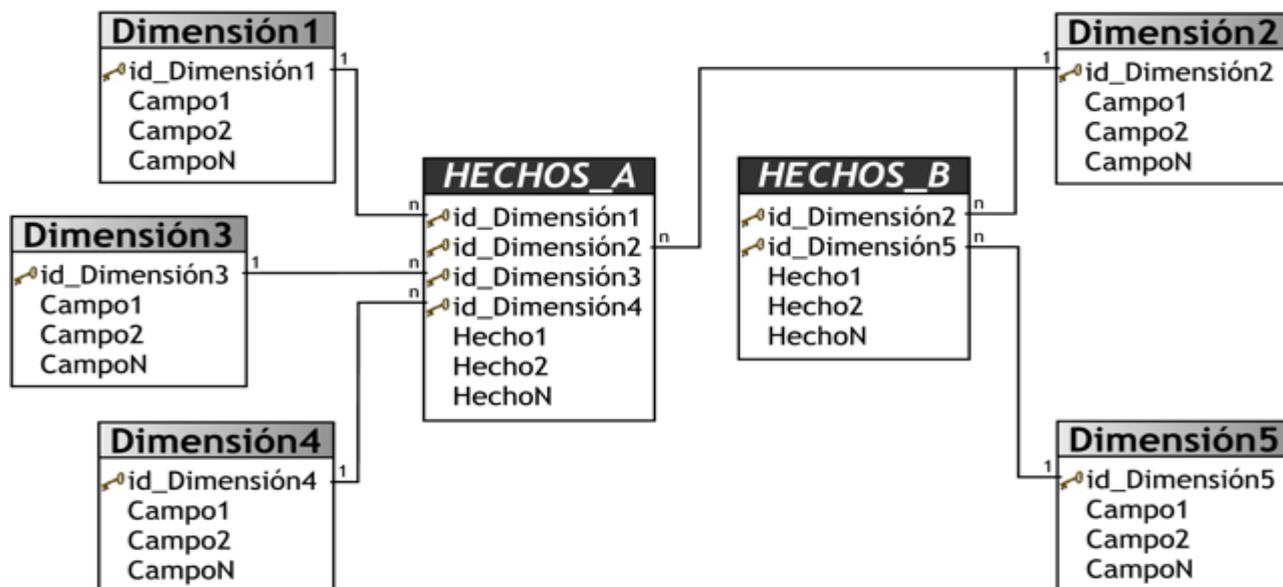


Figura 3 Ejemplo de esquema constelación de hechos

1.3 Procesamiento Analítico en Línea

OLAP es el acrónimo en inglés de procesamiento analítico en línea (On-Line Analytical Processing). Es una solución utilizada en el campo de la inteligencia de negocios que define a una tecnología que se basa en el análisis multidimensional de los datos y que le permite al usuario tener una visión más rápida e interactiva de los mismos. Es muy utilizado en el área de marketing, ventas, informes, etc., especialmente porque las respuestas a consultas complejas se obtienen muy rápidamente.

Existen diversos tipos de implementaciones de la tecnología OLAP, las que varían según el tipo de motor en el que se almacenan los datos. De esta manera, están clasificadas en:

ROLAP: Almacena los datos en un motor relacional por lo que se puede tener un acceso veloz a ellos. La principal ventaja de esa arquitectura es que permite el análisis de una enorme cantidad de datos.

MOLAP: Almacena los datos en una base de datos multidimensional. Para optimizar los tiempos de respuesta, el resumen de la información es usualmente calculado por adelantado. Estos valores precalculados o agregaciones son la base de las ganancias de desempeño de este sistema. Algunos sistemas utilizan técnicas de compresión de datos para disminuir el espacio de almacenamiento en disco debido a los valores pre-calculados.

HOLAP (Hybrid OLAP): Los datos son almacenados de dos formas, algunos en un motor relacional y otros en una base de datos multidimensional.

Operaciones sobre OLAP

Hay tres operaciones básicas que se pueden realizar en un cubo OLAP:

Rotar: Cambia las dimensiones del cubo que se está viendo y se obtiene una nueva vista de información. Por ejemplo, una vista de 'Importaciones por productos' puede cambiarse fácilmente a 'Importaciones por empresas'.

Rebanar: Cambiar el valor de una dimensión por otro valor, por ejemplo, de las ventas de Enero a las ventas de Febrero.

Taladrar o Drilling: Los datos de las dimensiones se pueden abrir para obtener más detalle. Si se ve información geográfica, puede pasar de un continente a un país y luego a una ciudad en particular.

Con estas operaciones, se puede indagar en la información generada por un negocio o información corporativa para todo el personal tomador de decisiones de varias formas. (OlapX 2002)

Utilidades de OLAP

1. Tienen acceso a grandes cantidades de datos.
2. Analizan las relaciones entre muchos tipos de elementos empresariales.
3. Presentan los datos en diferentes perspectivas.
4. Involucran cálculos complejos entre elementos de datos. (Ibarra 2005)
5. Puede responder con rapidez a consultas de usuarios.
6. Los datos están organizados en diferentes dimensiones, lo que permite un mejor análisis.

1.4 Mercado de datos o Datamart

Uno de los términos más abordados cuando se habla de almacenes de datos es el mercado de datos. Cuando se trata de almacenar información, esta es una de las posibles soluciones que está muy vinculada a los almacenes de datos. Existen varias definiciones acerca de lo que es un mercado de datos, a continuación se citan algunas de ellas:

Un mercado de datos es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para

analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento.(Sinnexus 2008)

Se entiende un mercado de datos como un subconjunto de los datos del almacén con el objetivo de responder a un determinado análisis, función o necesidad y con una población de usuarios específica.(Curto 2007)

Según Bill Inmon, un mercado de datos es una estructura departamental que alimenta un almacén en el cual los datos no están normalizados y se basa en las necesidades de información del departamento.(Inmon 2005)

“...Un *Datamart* es un almacén de datos históricos relativos a un departamento de una organización, así que puede ser simplemente una copia de parte de un Datawarehouse para uso departamental.”(Gascón and Castellar 2003)

Según Ralph Kimball, un mercado de datos es un subconjunto lógico de un almacén de datos completo. Frecuentemente es visto como una restricción de un almacén a un grupo específico del negocio. Está construido, orientado y organizado alrededor de una parte específica del negocio.(Kimball 2005)

Luego de haber analizado los diferentes conceptos citados, se concluye que un mercado de datos es un subconjunto de un almacén de datos, que contiene datos operacionales de un área específica del negocio sobre el cual se puede realizar análisis para dar soporte al proceso de toma de decisiones y al que tiene acceso un grupo limitado de personas.

1.4.1 Características y ventajas de los mercados de datos

1. Poseen las mismas características de integración, no volatilidad y orientación temática que un almacén de datos.
2. Representan una estrategia de divide y vencerás para ámbitos muy genéricos de un almacén de datos.
3. Los mercados de datos se enfocan a los requisitos de los usuarios que están asociados a un departamento específico de la empresa.
4. Su utilización y comprensión es sencilla debido que contienen menor número de información que los almacenes de datos.

5. Poseen menor alcance histórico que los almacenes de datos.

Ventajas del uso de los mercados de datos

1. Poseen menos volumen de información, lo que posibilita mayor rapidez en las respuestas a las consultas y que su elaboración sea más fácil.
2. Se realizan sobre ellos consultas MDX¹ o SQL sencillas que facilitan el acceso a los datos que son utilizados con frecuencia.
3. Simplifican el desarrollo de todo el mecanismo de su base de datos y con ello baja substancialmente todo el coste del proyecto, así como su duración. (Greenhouse 2010)
4. Son simples de implementar.
5. Conllevan poco tiempo de construcción y puesta en marcha.
6. Permiten manejar información confidencial.
7. Reflejan rápidamente sus beneficios y cualidades.

1.5 Procesos básicos de un almacén de datos

1.5.1 Integración de datos

La integración de datos es el proceso de unificación de los datos provenientes de múltiples fuentes, según Ralph Kimball. Cuando la información proveniente de diferentes sistemas es inconsistente y con baja calidad, se plantea que esto, es el resultado de que los datos sean heterogéneos y se encuentren distribuidos y dispersos y en la mayoría de los casos no estandarizados. La tarea de aunar los datos con el fin de centralizarlos de manera única, es costosa y compleja.

Una de las formas en que puede realizarse este proceso es mediante la Extracción, Transformación y Carga de Datos que además es la que se emplea para la solución.

Extracción, Transformación y Carga

Es la tecnología enfocada a la integración de datos, tanto por lote como a tiempo real hacia almacenes de datos. (Kimball and Caserta 2004) Estos procesos se combinan para extraer datos de bases de datos fuentes o archivos, y colocarlas en bases de datos destino. Los procesos ETL se utilizan para

¹Multi-Dimensional eXpressions, más conocida por sus siglas MDX.

migrar datos de una o más bases de datos a terceros y también para convertir bases de datos de un tipo o formato a otro. Se utilizan además para sincronizar datos desde diversas aplicaciones.

Para lograr los objetivos trazados, es necesario utilizar esta tecnología ya que las fuentes de datos que se quieren integrar trabajan con diferentes nomenclaturas, por lo que el proceso de integración implica aplicar reglas de transformación para depurar los datos y almacenarlos en un destino de forma tal que la información que se genere tenga preparación, organización y esté disponible para ser consultada en el momento requerido.

Los procesos ETL son los componentes más importantes de una infraestructura de inteligencia de negocio. Pueden ser invisibles por los usuarios, recuperan los datos de todos los sistemas operativos y los pre-elaboran para las herramientas de análisis y de reporte.

Algunas de las características que posee son:

1. Es un mecanismo de carga muy eficiente y efectivo orientado a los almacenes de datos.
2. Enfocado a migrar y mezclar datos.
3. Reduce la exposición a desarrollos manuales producto de la existencia en el mercado de herramientas potenciales para la implementación visual, con manejo de excepciones, gestión y planificación de tareas.(Tandrón 2008)
4. Necesita pocos servicios de administración y mantenimiento.
5. Gran capacidad para llevar a cabo transformaciones.
6. Tecnología enfocada a la integración de datos en bases de datos versátiles hacia los almacenes de datos.

Subprocesos ETL

El proceso ETL se divide en tres subprocesos fundamentales, los cuales se exponen a continuación.

Extracción: consiste en extraer los datos de la fuente de origen. Se convierten los datos a un formato preparado para iniciar la transformación. Aquí se verifican la calidad de los datos extraídos y si cumplen los estándares conforme a los requerimientos. En caso de extraer muchos datos el sistema origen podría ralentizar.

Limpieza y Transformación: En esta fase se aplican una serie de Reglas del negocio² sobre los datos extraídos, con el objetivo de convertirlos en datos aptos para ser cargados. Aquí es necesario lograr una buena calidad de los datos y para ello es necesario el control de los valores válidos, garantizar la coherencia entre los valores, la eliminación de duplicaciones y comprobar que las reglas del negocio no han sido forzadas.

Luego que los datos han sido limpiados se procede a realizar las transformaciones mediante las reglas de transformación que pueden ser: combinar los datos de distintas fuentes, realizar búsqueda de valores en distintas tablas, darle tratamiento a valores nulos, entre otras.

Carga: Momento en el que los datos, provenientes de la fase anterior, son incluidos en el sistema de destino, dependiendo de los requerimientos de la organización. El principal objetivo de esta fase es lograr que los datos estén listos para ser consultados. Este subproceso puede abarcar una amplia variedad de acciones diferentes. En algunas bases de datos se sobrescribe la información antigua con nuevos datos. En los almacenes de datos al mantener un historial de los registros se puede hacer una auditoría de los mismos y disponer de un rastro de toda la historia de un valor a lo largo del tiempo.

1.5.2 Inteligencia de negocio

Una de las claves para que las empresas lleguen a tener éxito es la capacidad para tomar decisiones precisas y rápidas. Para ello se apoyan en una colección de tecnologías que posibilitan sin lugar a duda una mejor gestión de la información. A este conjunto de tecnologías se le denomina Inteligencia de Negocios, Inteligencia Empresarial o BI (del inglés Business Intelligence). Dicho término no está exento de poseer varias interpretaciones.

Se puede definir el término inteligencia de negocio como: "...conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización o empresa para facilitar la toma de decisiones."(Neves 2007)

"...conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio."(Sinnexus 2008)

² Describe las políticas, normas, operaciones, definiciones y restricciones presentes en una organización y que son de vital importancia para alcanzar los objetivos misionales.

La Inteligencia de Negocios es un concepto que hace referencia a las técnicas de análisis de datos destinados a encontrar información útil para la toma de decisiones, incluido el conjunto del software que aporta las interfaces y funciones necesarias que apoyan dicho proceso. (Romero 2004)

Otro concepto al que se hace alusión es: La inteligencia de negocio es una arquitectura compuesta por procesos, tecnologías y herramientas que transforman los datos de los factores claves de éxito en información, información en conocimiento y conocimiento en planes que impulsarán las acciones que harán rentable al negocio. (Onuva 2008)

Luego de realizar una, revisión bibliográfica de diferentes autores se puede concluir que la inteligencia de negocio es un conjunto de herramientas y métodos que permiten transformar los datos en información una vez que el usuario haya encontrado en los mismos algún significado especial, además posibilitan también que dicha información sea muy útil a los directivos para contribuir en los procesos de toma de decisiones de la empresa.

Este conjunto de herramientas y estrategias debe poseer las siguientes características: (Huamantumba 2007)

1. Deben garantizar el acceso de los usuarios a los datos sin importar cuál sea la procedencia de los mismos.
2. No solo deben posibilitar la presentación de la información, sino también permitir que el usuario realice análisis sobre los datos más relevantes para el negocio.
3. Independientemente de los conocimientos técnicos que posean los usuarios, estos deben ser capaces de manipular estas herramientas.

Ventajas que ofrece una solución de inteligencia de negocio

1. Una solución de inteligencia de negocio completa permite observar y comprender que está ocurriendo en la entidad y por qué ocurre, ayuda a predecir qué ocurrirá y que caminos debería tomar el equipo para el mejoramiento del negocio.
2. Al tener una sola solución que controle todos los programas que se encuentren en los diferentes departamentos de la empresa se tiene un mejor control de los costes.

3. El proceso de toma de decisiones mejora, porque se tiene acceso a toda la información de todos los niveles de la organización, lo que ayuda a los ejecutivos a planear y pronosticar el trabajo, proporcionándole a los mismos un amplio conocimiento del negocio.

1.6 Metodología de desarrollo

Cuando se hace uso del vocablo metodología, se refiere a un método o conjunto de operaciones utilizados en una investigación científica que conllevan a alcanzar un objetivo determinado.

A medida que las soluciones de almacenes de datos han ido madurando y alcanzando un alto nivel entre los sistemas destinados al análisis de información histórica y soporte a los procesos de toma de decisiones, también han evolucionado las metodologías para el desarrollo e implantación de este tipo de soluciones. Existen dos de ellas bien definidas y que han servido de guía a las entidades en cuanto a este tema, las mismas son: la metodología Kimball en honor a su creador Ralph Kimball y la metodología de Inmon debido a su creador Bill H. Inmon, dos de las personalidades que más influyeron en el área de los almacenes de datos. Ralph Kimball es reconocido a nivel mundial en el diseño de los almacenes de datos y creador del enfoque multidimensional y Bill H. Inmon es el creador del término almacén de datos y considerado como el padre de dicha disciplina.

Existe una diferencia entre ambas metodologías, la misma se evidencia en la forma de enfrentar el problema. Inmon se basa en un enfoque descendente, plantea que primero se debe construir el almacén de datos y a partir de este los mercados de datos. Mientras que la metodología de Ralph Kimball se enfoca en el diseño de la base de datos que almacenará la información para la toma de decisiones, el mismo está compuesto por las tablas de hecho y dimensiones, las cuales contienen toda la información cuantitativa y cualitativa respectivamente de los indicadores. Una de sus características principales es que posee una arquitectura ascendente ya que plantea que primeramente se deben crear mercados de datos independientes orientados a temas específicos y la unión de todos ellos conformaría el almacén.

Existen diversas metodologías que pretenden dar un acercamiento a una propuesta ideal para el desarrollo de almacenes. Cada autor la orienta a la optimización del rendimiento y a su visión de los principales procesos que se deben tener en cuenta para construir un almacén de datos flexible y dinámico. En dependencia de la problemática presentada es la política de selección por una y otra metodología. (DATEC 2009)

Para este tipo de soluciones DATEC³ emplea un modelo de desarrollo de soluciones de almacenes de datos que cubre todas las fases por las que pasa la construcción de un almacén de datos, desde el levantamiento de información inicial hasta la capa de visualización. Este modelo toma como guía la metodología de Kimball por los siguientes elementos:

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

1.7 Herramientas seleccionadas para la solución

El avance y la diversificación de las herramientas de desarrollo actualmente son crecientes, tanto comerciales como de código abierto⁴. Apreciadas como un conjunto de programas y ayudas que dan asistencia a los analistas y desarrolladores, durante todos los pasos del ciclo de vida de desarrollo de un software.

1.7.1 Herramienta de modelado

Visual Paradigm for UML 6.4

Es una herramienta que utiliza UML como lenguaje de modelado. Soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. Su uso posibilita una construcción de aplicaciones rápida, con buena calidad, y a un menor coste. Proporciona abundantes tutoriales. Además, posibilita la integración con sistemas de control de versiones que almacenan centralmente los artefactos y realizan un seguimiento de los cambios realizados sobre un proyecto.

³DATEC: Centro de Tecnologías de Gestión de Datos

⁴ En inglés open-source es el término con el que se conoce al software distribuido y desarrollado libremente.

Acelera el desarrollo de aplicaciones, sirviendo de puente visual entre arquitectos, analistas y diseñadores de sistemas de información, haciendo el trabajo más fácil y dinámico. Además, automatiza tareas tediosas que pueden distraer a diseñadores del desarrollo. (Cuervo and Moreno 2005)

Algunas de las características que posee esta herramienta son las siguientes:

1. Navegación intuitiva entre el código y el modelo visual. (Cuervo and Moreno 2005)
2. Poderoso generador de informes PDF/HTML.
3. Tiempo real en la demanda.
4. Un ambiente modelador visual superior.
5. Sofisticado diseño de diagramas.

Entre las ventajas más significativas de Visual Paradigm se pueden citar:

1. Es una herramienta multiplataforma.
2. Soporta ingeniería inversa, generación de código e importación desde Rational Rose.
3. Permite la generación de bases de datos (transformación de diagramas de Entidad-Relación en tablas de base de datos).

Para la solución se podrá usar esta herramienta en su versión para UML 6.4 Enterprise Edition debido a que la UCI posee licencia para la utilización de la misma.

1.7.2 Sistema Gestor de Bases de Datos

Un Sistema Gestor de Bases de Datos es un conjunto de programas que tiene como objetivo fundamental servir de interfaz entre la base de datos, el usuario y las aplicaciones.

PostgreSQL 8.4

Es una de las herramientas de gestión más utilizada, considerada como la base de datos Open Source más avanzada del mundo. Opera bajo licencia BSD⁵. Actualmente muchas organizaciones, incluyendo grandes corporaciones, instituciones gubernamentales y pequeños negocios en línea usan PostgreSQL para manejar sus datos.

⁵BSD (Berkeley Software Distribution), sólo requiere que el código fuente licenciado mantenga la información de derechos de autor y licenciamiento.

Algunas de las ventajas que presenta son:

1. Puede ser instalado un número ilimitado de veces.(Díaz 2003) Esto trae consigo otras ventajas como son:
 - Modelos de negocios más rentables con instalaciones a gran escala.
 - No existe la posibilidad de ser auditado para verificar cumplimiento de licencia en ningún momento.
 - Flexibilidad para hacer investigación y desarrollo sin necesidad de incurrir en costos adicionales de licenciamiento.
2. Permite la gestión de diferentes usuarios, como también los permisos asignados a cada uno de ellos, lo que evidencia su confiabilidad.
3. Es multiplataforma.
4. Incorpora funciones de diversa índole: manejo de fechas, geométricas, orientadas a operaciones con redes.
5. Permite la declaración de funciones propias, así como la definición de disparadores.
6. Es un gestor de bases de datos relacionales que permite manejar grandes volúmenes de información.

Debido a las ventajas antes expuestas, se decidió utilizar como gestor de bases de datos PostgreSQL en su versión 8.4 para la solución. Además, la ONE actualmente se encuentra inmersa en un proceso de migración hacia la independencia tecnológica, sus medios de almacenamiento están siendo orientados hacia la plataforma PostgreSQL, debido a que es una propuesta Código Abierto (Open Source) que sobrepasa a muchas propietarias y que posee gran potencialidad y adaptabilidad.

1.7.3 Herramientas para la Integración de datos

Pentaho Data Integration 4.0.1

Esta herramienta reúne un conjunto de componentes que permiten modelar y ejecutar transformaciones sobre flujos de datos. Es una de las herramientas ETL de código abierto más antigua, cuenta con una gran comunidad de usuarios y su interfaz gráfica permite un aumento de la productividad. Puede funcionar sobre varias plataformas a través de un sistema que soporte un

entorno de ejecución de Java 1.4 o una versión superior. Se integra con ficheros de Microsoft Office. Incluye procesamiento optimizado de los ficheros planos.

Ofrece soporte para metadatos e incluye operaciones de transformación, así como funciones que posibilitan operar con los campos en el flujo de datos, renombrando, calculando campos en función de otros, correlacionando valores y realizando búsquedas auxiliares en bases de datos.

Esta herramienta es fácil de usar, brinda la posibilidad de copiar y leer del mismo fichero en paralelo, permitiendo maximizar la capacidad de entrada/salida en el entorno ETL. Añade un debugger⁶ integrado diseñado para mejorar la productividad del desarrollador, pudiendo agregar puntos de ruptura condicionales en la ejecución de las transformaciones, dando la posibilidad de pausar y resumir la ejecución de la transformación.

Se debe destacar que su rendimiento se puede ver afectado cuando se realizan operaciones de join⁷ con numerosos volúmenes de datos, pues maneja pequeñas cantidades de información en el flujo. Permite ejecutar código JavaScript dentro de las transformaciones e incorpora un evaluador de expresiones regulares.

Presenta algunas desventajas porque no cuenta con un componente de calidad de datos, no automatiza el proceso de separación y redistribución de datos para el procesamiento paralelo, además que para realizar búsquedas de mayores volúmenes necesita utilizar una base de datos de búsqueda donde se ejecutan un gran número de sentencias SQL que frenan el rendimiento de ETL. La arquitectura de Pentaho Data Integration se puede observar en la figura 4 (Ferreira and Schmidt 2009).

⁶Herramienta informática con la que se puede leer el código fuente en el que están escritos los programas.

⁷ Sentencia en SQL, que permite combinar registros de dos o más tablas en una base de datos relacional.

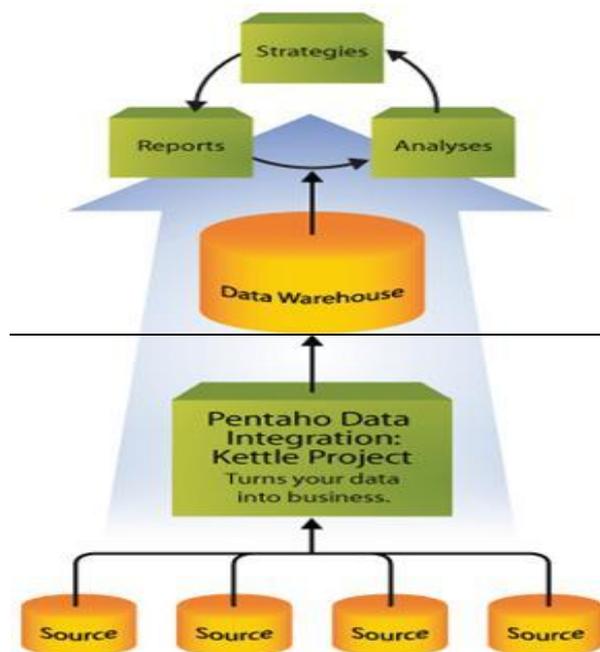


Figura 4 Arquitectura de Pentaho Data Integration

DataCleaner 1.5.3

Esta herramienta requiere un entorno de ejecución de Java 5.0 o una versión superior y Drivers de JDBC⁸. La misma permite la evaluación del nivel de calidad de los datos contenidos en el sistema de información preparando los mismos para la transformación en el proceso ETL. Es una aplicación muy fácil de usar, genera sofisticados informes que permiten a los usuarios determinar de un vistazo el nivel de calidad de los datos, identificar y analizar la estructura del origen de datos y combinar resultados creando vistas fáciles de interpretar para evaluar la calidad de los datos.

Entre sus características se pueden mencionar:

1. Los perfiles de datos se utilizan para calcular y analizar diversas medidas importantes basadas en los valores de los datos.
2. Posee un validador de datos el cual dará un resultado que puede ser interpretado como bueno o malo.
3. Soporta acceso de lectura a muchos tipos de almacenes de datos.

⁸ JDBC: Es una biblioteca de clases que permite la conexión con bases de datos utilizando Java.

1.7.4 Herramientas para Inteligencia de negocio

Pentaho Schema Workbench 3.0

Es una herramienta para el desarrollo y prueba de cubos OLAP de forma visual. La definición del XML no es extremadamente compleja, pero en la práctica resulta engorroso recordar cada uno de los elementos junto a sus atributos y sub-elementos tal y como se encuentran en el almacén. Con esta aplicación, se puede configurar una conexión con el modelo físico, para luego elaborar el esquema lógico de manera simple y efectiva. Para ello la herramienta ofrece un editor de esquemas con la fuente de datos subyacente para su validación. (Pentaho 2005)

Mondrian OLAP Server 3.0.4

La aplicación web OLAP Mondrian combinado con Jpivot⁹, permite realizar consultas a almacenes de datos, los resultados son presentados mediante un navegador de modo que el usuario pueda realizar las actividades típicas de navegación. Mondrian utiliza MDX como lenguaje de consulta, que fue un lenguaje propuesto por Microsoft. Funciona sobre las bases de datos estándar del mercado: Oracle, DB2, SQL-Server, MySQL, PostgreSQL, entre otras, lo cual habilita y facilita el desarrollo del negocio basado en la plataforma Pentaho. (Pentaho 2005)

Pentaho BI Server 3.6.0

Con esta herramienta se suministra soporte e infraestructura para crear soluciones de inteligencia de negocio. Proporciona servicios básicos además de incluir autenticación, registro, auditoría y servicios web. Incorpora 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). Además, está diseñada para integrarse fácilmente en cualquier proceso de negocio. Permite que puedan ejecutarse los informes y aplicaciones, se puede usar como base para construir un sistema propio de inteligencia de negocios. (Summan 2006)

⁹Jpivot es una biblioteca de etiquetas JSP personalizada, representa una tabla OLAP y permite a los usuarios realizar operaciones como: rotar, rebanar, taladrar.

Tres de sus principales ventajas son:

1. Administra y programa reportes.
2. Administra seguridad de usuarios.
3. Brinda la posibilidad de guardar la consulta que se ejecute.

Apache Tomcat 6.0

Tomcat es el servidor Web más utilizado a la hora de trabajar con Java en entornos web. Es desarrollado en un contexto abierto y participativo, publicado bajo la licencia Apache en su versión 2. Posee conectores http lo que posibilita la gestión de solicitudes y respuestas. Funciona en cualquier sistema operativo que disponga de la máquina virtual Java porque fue escrito en ese lenguaje. (Tomcat 2010)

1.8 Conclusiones del capítulo

Luego de finalizar el desarrollo de este capítulo, se puede concluir lo siguiente:

- Se abarcó una panorámica general del proceso de desarrollo de los almacenes de datos y mercados de datos, así como algunas de las definiciones que exponen estudiosos del tema. De aquí se ultima que el desarrollo de un mercado de datos soluciona el problema contribuyendo a la toma de decisiones.
- Se decidió adoptar la Metodología para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio propuesta por DATEC, la cual está basada en la metodología de Ralph Kimball y se considera que es la más adecuada para el desarrollo de la solución.
- Se seleccionaron varias herramientas como Visual Paradigm, PostgreSQL 8.4, otras pertenecientes a Pentaho BI Suite 3.5 Community Edition y algunas para la solución de inteligencia de negocio, de las cuales se expusieron un grupo de características que las identifican como las más adecuadas para la solución.

CAPÍTULO 2: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS COMERCIO EXTERIOR

Durante el desarrollo de este capítulo se realiza una revisión de los requisitos, los casos de uso, los hechos, las medidas y las dimensiones del mercado de datos identificados en la tesis precedente, para realizar el refinamiento de los mismos. Una vez concluida esta etapa se construye el modelo de datos correspondiente y se diseña el subsistema de integración y el de visualización.

2.1 Análisis del mercado de datos Comercio exterior

2.1.1 Estudio preliminar del negocio

La Oficina Nacional de Estadísticas (ONE) es una entidad creada para proponer, organizar y ejecutar, según corresponda, la aplicación de la política estatal en materia de estadística del país. Mediante el Sistema Estadístico Nacional (SEN), ejerce una adecuada dirección, ejecución y control de la captación de las cifras económicas y sociales, así como su difusión de acuerdo con los requerimientos de la economía y las demás necesidades del país en información estadística. Tiene como visión construir un sistema estadístico profesional, capaz de responder con calidad a las necesidades de información del país, para enfrentar los objetivos de desarrollo económico y social así como su adecuado reflejo internacional.

El proyecto Sistema de Información del Gobierno nace de la necesidad de centralizar toda la información existente en la ONE para lograr un mejor monitoreo y control de los datos. Se enfoca en la creación de una herramienta que permita acceder a toda la información, con el objetivo de apoyar la toma de decisiones en las diferentes áreas socioeconómicas. Una de estas áreas es Comercio exterior, en la que se gestionan todas las estadísticas relacionadas con las actividades de exportación e importación efectuadas en Cuba.

Comercio exterior recibe directamente de la aduana nacional toda la información, la cual se nutre de las aduanas que radican en las diferentes provincias del país. En el proceso de envío de información no existen entidades ni organismos intermediarios debido a que toda la información que se gestiona es de carácter secreto.

2.1.2 Temas de análisis.

Para el desarrollo de los mercados de datos es necesaria la identificación de los temas de análisis. Con estos se obtienen varias perspectivas que orientan el avance del cumplimiento de las tareas planteadas y garantizan la utilidad y el éxito del diseño de las estructuras que se desarrollan. En el

trabajo de diploma precedente se hace referencia a dos temas de análisis: Análisis de las exportaciones y Análisis de las importaciones. Básicamente la ONE maneja toda la información relacionada con las exportaciones y las importaciones del área Comercio exterior, pero debido a los nuevos requisitos de información solicitados por el cliente, se identificó un nuevo tema de análisis relacionado con el intercambio comercial. A continuación se muestra como quedaron definidos en la presente investigación:

- Análisis de las exportaciones.
- Análisis de las importaciones.
- Análisis del intercambio comercial.

2.1.3 Necesidades de los usuarios

Para el buen desarrollo del análisis en el proceso del negocio es preciso conocer que es lo que necesitan los usuarios. La implicación de los mismos durante el ciclo de vida del producto es de gran importancia, de ahí proviene la posibilidad de que los resultados sean satisfactorios o insatisfactorios en correspondencia a sus necesidades. En la ONE para el análisis del Comercio exterior, los especialistas se centran en las exportaciones e importaciones efectuadas en Cuba. De estas dos operaciones se almacenan varios parámetros en los cuales ellos enfocan sus estudios. Todo este análisis está destinado a satisfacer necesidades gubernamentales.

Requisitos de información.

Los requisitos de información son las principales funcionalidades que el sistema debe tener disponible a la hora de realizar análisis sobre los datos. Constituyen una entrada fundamental para el proceso de inteligencia del negocio y para futuros reportes. Se definen a partir de la comparación entre las necesidades de información y las reglas del negocio con los elementos disponibles en las fuentes. En la investigación precedente fueron identificados 10 requisitos de información relacionados con las importaciones y las exportaciones efectuadas en Cuba.

Debido a las nuevas necesidades de información del cliente surgieron nuevos requisitos en la presente investigación, a continuación se muestran los mismos:

- RI 1. Obtener importaciones de productos agrupados por GCE, en un tiempo dado.
- RI 2. Obtener importaciones de productos por GCE y país de origen en un tiempo dado.
- RI 3. Obtener importaciones de productos seleccionados en un tiempo dado.

- RI 4. Obtener exportaciones de productos por GCE y países en un tiempo dado.
- RI 5. Obtener exportaciones de productos de la industria azucarera sin incluir donativos en un tiempo dado.
- RI 6. Obtener exportaciones de productos de la industria azucarera incluyendo donativos en un tiempo dado.
- RI 7. Obtener exportaciones de productos seleccionados en un tiempo dado.
- RI 8. Obtener exportaciones de productos por GCE incluyendo donativos en un tiempo dado.
- RI 9. Obtener exportaciones de productos por GCE sin incluir donativos en un tiempo dado.
- RI 10. Obtener exportaciones de productos sin donativos agrupados en GCE según la CUCI en un tiempo dado.
- RI 11. Obtener el intercambio de bienes con el exterior por empresas en un tiempo dado.
- RI 12. Obtener el intercambio de bienes con el exterior por países en un tiempo dado.

Requisitos funcionales.

Los requisitos funcionales son aquellos que definen las funciones que el sistema va a llevar a cabo. Deben estar orientados a las necesidades de los usuarios finales. En la tesis precedente fueron identificados 4 requisitos funcionales, a continuación se muestran los mismos:

- RF 1. Obtener el comportamiento de las importaciones y las exportaciones.
- RF 2. Elaborar una copia de seguridad mensual en el repositorio.
- RF 3. Permitir cargar, extraer y transformar los datos.
- RF 4. Realizar los reportes de los modelos.

Para ese entonces estos requisitos respondían a las funcionalidades solicitadas por el cliente, pero debido a nuevas peticiones de los usuarios y al nuevo análisis realizado han sido identificados nuevos requisitos funcionales que la solución debe poseer, estas son:

- RF 1. Autenticar usuario.
- RF 2. Adicionar usuario.
- RF 3. Eliminar usuario.
- RF 4. Adicionar rol.
- RF 5. Eliminar rol.

- RF 6. Adicionar reporte.
- RF 7. Eliminar reporte.
- RF 8: Modificar reporte.
- RF 9. Realizar la extracción de los datos.
- RF 10. Realizar la transformación y carga de los datos.
- RF 11. Configurar vista de análisis OLAP.
- RF 12. Editar consulta MDX.
- RF 13. Mostrar propiedades.
- RF 14. Suprimir filas y columnas vacías.

Requisitos no funcionales.

Los requisitos no funcionales son propiedades o cualidades que el producto debe tener. En la tesis precedente fueron identificados 9 requisitos no funcionales relacionados con la fiabilidad, eficiencia y capacidad de la solución. En la presente investigación, fueron redefinidos para el desarrollo del mercado de datos 12 requisitos no funcionales. Para conocimiento de los mismos ver artefacto Especificación de Requisitos que se adjunta en el expediente de proyecto.

2.1.4 Casos de Uso del Sistema

Los casos de uso del sistema representan información de manera visual. Describen lo que debe hacer el sistema relacionado con el usuario, representando generalmente requisitos funcionales. Para una mejor comprensión de los mismos se hace una representación gráfica que contiene los casos de uso y la relación que mantiene con los usuarios que interactúan con él.

En la investigación anterior se definieron los casos de uso informativos Solicitar información de importaciones y Solicitar información de exportaciones, además 32 casos de uso funcionales. Para la realización del mercado de datos Comercio exterior en la presente investigación fue necesario redefinir casos de uso de información y funcionales, donde se realizó la especificación de cada uno de estos, la misma se adjunta en el expediente de proyecto, en el artefacto Modelo de Casos de Uso.

Actualmente los casos de uso informativos se agrupan por el tipo de información que se maneja en la ONE específicamente en el área Comercio exterior y en dependencia de las necesidades de los usuarios. Estos se nombran:

- C.U.I Analizar información de importaciones.
- C.U.I Analizar información de exportaciones.
- C.U.I Analizar información de intercambio comercial.

Los casos de uso funcionales identificados son:

- C.U.F Administrar usuario.
- C.U.F Administrar rol.
- C.U.F Administrar reporte.
- C.U.F Realizar la extracción de los datos.
- C.U.F Realizar la transformación y carga de los datos.
- C.U.F Analizar reporte.
- C.U.F Autenticar usuario.

A continuación se muestra el Diagrama de Casos de Uso del Sistema, en el cual se representa gráficamente a los procesos del negocio y su interacción con los clientes:

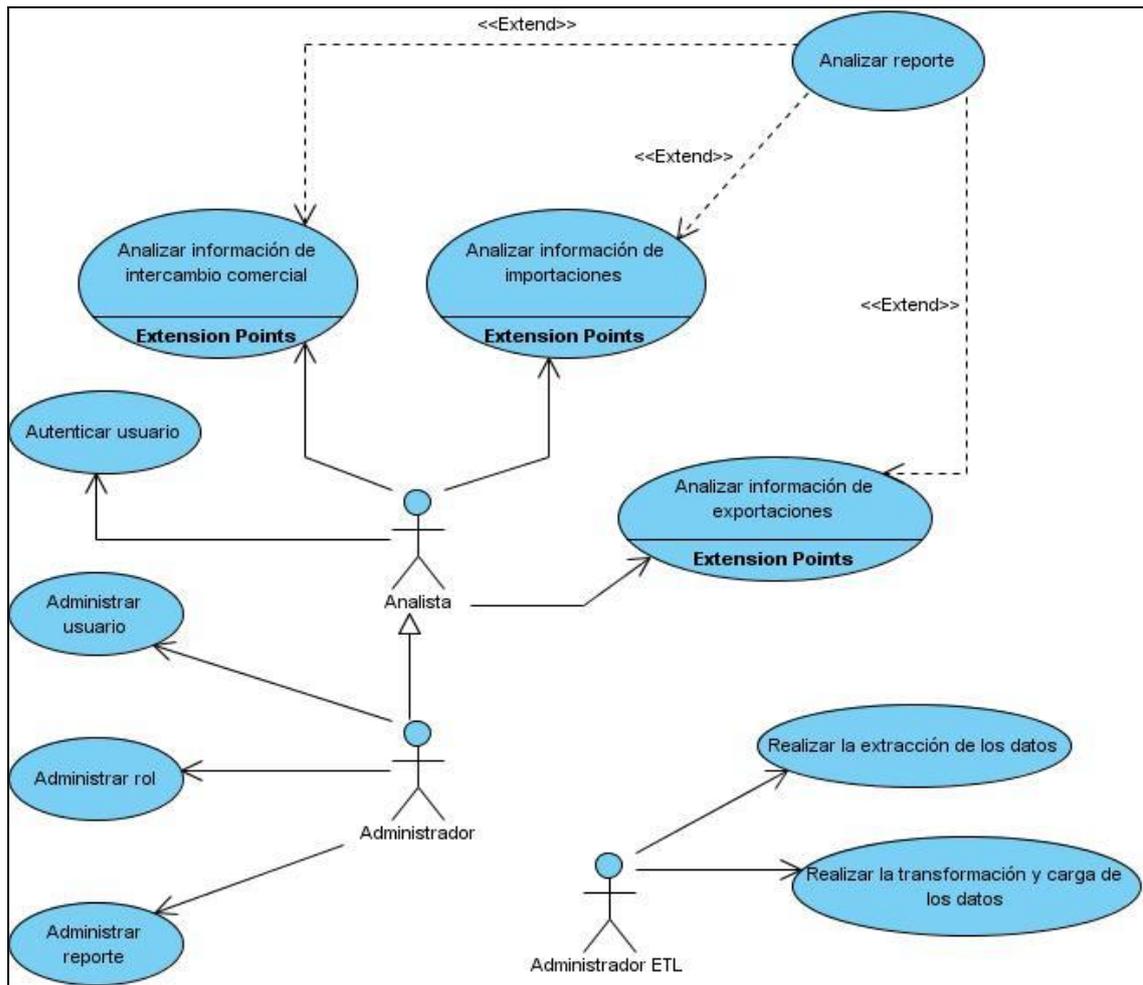


Figura 5 Diagrama de Casos de Uso del Sistema.

2.1.5 Reglas de Negocio

Las reglas del negocio describen las políticas, normas, operaciones, definiciones y restricciones presentes en una organización. Son de gran importancia para el logro de los objetivos en la misma. En el proceso de almacenamiento de los indicadores del Comercio exterior se guarda la información de las exportaciones, de las importaciones y del intercambio comercial en forma de clasificadores para un mejor control de estos parámetros. Varios de estos clasificadores tienen significados propios, que conllevan a las reglas del negocio. A continuación se relacionan algunas de ellas:

Clasificador Subpartida

RN 1: Describe los productos que están siendo comercializados ya sea en una exportación o en una importación. Tiene un código de ocho dígitos, de estos, el primero pertenece a la sección, los dos primeros al capítulo, los cuatro primeros pertenecen a la partida, estos cuatro unidos a los dos siguientes corresponden a la subpartida y los ocho en total representan el código SACLAP¹⁰.

Clasificador Grandes Categorías Económicas (GCE)

RN 2: Divide los productos en grandes categorías económicas. Posee un código de tres dígitos. El mismo está compuesto de la siguiente forma: primer número de este código representa a la categoría, este unido al segundo corresponden a la sub-categoría y los tres unidos significan la categoría básica.

Clasificador Uniforme para el Comercio Internacional (CUCI)

RN 3: Este es el clasificador por el cual se rigen todas las entidades a nivel mundial que llevan algún tipo de control sobre el comercio internacional. Posee un código de cuatro cifras de las cuales la primera representa la sección, las dos primeras corresponden al capítulo, las tres primeras al grupo y las cuatro cifras unidas significan el subgrupo.

Clasificador Industrial Internacional Uniforme (CIIU)

RN 4: Posee un código de cuatro dígitos. El primero representa la categoría, los dos primeros la división, los tres primeros el grupo y los cuatro juntos corresponden a la clase.

En la investigación precedente solo fueron identificadas tres reglas del negocio, las mismas coinciden con las tres primeras mencionadas. Se identificaron otras reglas del negocio en la presente investigación, las que se documentan en el artefacto Reglas del negocio que se adjunta en el expediente de proyecto.

2.2 Diseño del mercado de datos Comercio exterior

2.2.1 Dimensiones, Hechos y Medidas

Los términos hecho, dimensión y medida son muy usados en la disciplina de los almacenes de datos. Un hecho representa la ocurrencia de un determinado proceso dentro de la organización; una dimensión es una perspectiva por la cual se puede realizar análisis sobre el hecho y una medida es un valor o indicador de análisis del hecho.

En la investigación anterior fueron identificadas 8 tablas de dimensiones, estas fueron:

¹⁰SACLAP: Sistema Armonizado de Clasificación de Productos, constituye una nomenclatura internacional orientada a los aranceles de aduana y a las estadísticas del Comercio exterior. Es equivalente a la subpartida.

- dim_temporal. Jerarquía: anno → mes
- dim_países. Jerarquía: nombre_area_geo → nombre_país
- dim_moneda. Jerarquía: moneda
- dim_empresa. Jerarquía: código
- dim_subpartidas. Jerarquía: capítulo → partida → subpartidagce → sgce → gcec → cuci → scuc → dcuc → gcuc → ccuc
- dim_con_pag. Jerarquía: código
- dim_regimen_exp. Jerarquía: código
- dim_régimen_imp. Jerarquía: código

Debido a los nuevos requerimientos de información y a nuevas peticiones del cliente se identificaron 18 tablas de dimensiones, algunas coinciden con 7 de las antes mencionadas. En la siguiente tabla se describen las 18 dimensiones que fueron identificadas para el mercado de datos Comercio exterior en la presente investigación:

Tabla 1 Descripción de las dimensiones

Dimensión	Descripción
dim_régimen	Esta dimensión refleja los tipos de comercio (Importación, Exportación, Reimportación, Reexportación, Reimportación de bienes devueltos, Reexportación de bienes devueltos) que se realizan con la mercancía.
dim_cond_pago	Contiene el nombre y el código que corresponda según la contrapartida financiera derivada de la transacción comercial realizada.
dim_subpartida	Subgrupos en los que se dividen las mercancías de una partida, se identifica por 8 dígitos. Jerarquía: código SACLAP → subpartida → partida → capítulo → sección
dim_prod_sele	Se indica código y descripción de los productos seleccionados.
dim_prod_azuc	Se indica el código y la descripción de los productos azucareros que son comercializados.
dim_empresa	Indica el código y nombre de la empresa que realiza la actividad de importación o de exportación.
dim_país	Esta dimensión indica el nombre del país y del área geográfica al cual pertenecen los mismos, así como los diferentes códigos ISO de cada uno de los países de los cuales provienen las importaciones y a los cuales se realizan las exportaciones.
dim_temporal_mes	Se registran un grupo de atributos relacionados con la fecha en la cual se realiza la importación o la exportación. Jerarquía: año → semestre → trimestre → mes
dim_moneda	Se indica la descripción y código de la moneda en que se

	declara la mercancía.
dim_aduana	Se refleja el nombre y código de la aduana donde se presentan los documentos para su autorización, registro y concesión del levante.
dim_modos_transp	Se refleja el código y la descripción el modo de transporte en que las mercancías arriban o salen del país.
dim_nac_transp	Se refleja la descripción y el código de la nacionalidad del medio de transporte que traslada la mercancía importada o exportada.
dim_gran_cat_econ	Se refleja el código y la descripción de las grandes categorías económicas por las cuales se agrupan los productos. El código está compuesto por 3 dígitos: Jerarquía: categorías básicas → subcategoría → categoría
dim_cuci	Se especifica el código y la descripción de los productos según la Clasificación Uniforme para el Comercio Internacional (CUCI). Jerarquía: subgrupo → grupo → capítulo → sección
dim_clase_bulto	Se reflejará la descripción y código del embalaje ¹¹ en que viene contenida la mercancía declarada.
dim_grupo_prod	Se indica el código y la descripción de los grupos de productos por los que se agrupa la mercancía.
dim_ciiu	Refleja el código y el nombre de la Clasificación Industrial Internacional Uniforme (CIIU) de todas las actividades económicas.

¹¹Embalaje: Cualquier medio material, que sirve para acondicionar, presentar, embalar, manipular, almacenar, conservar y transportar una mercancía.

	Jerarquía: clase → grupo → división → categoría
dim_um	Se indica el código y la descripción de las unidades de medidas en las que están dadas las mercancías.

En la tabla 2 se describen los hechos importación y exportación, los mismos fueron identificados en la investigación precedente y actualmente se mantienen en el mercado de datos Comercio exterior:

Tabla 2 Descripción de los hechos

Hecho	Descripción
hech_importacion	Es objeto de análisis del negocio. Comprende todo lo relacionado con las mercancías que van a entrar al país.
hech_exportacion	Es objeto de análisis del negocio. Comprende todo lo relacionado con las mercancías que salen del país.

En la investigación precedente se identificaron 12 medidas, actualmente el mercado de datos cuenta con 14 medidas, las que se definieron anteriormente y 2 más que fueron añadidas a petición del cliente. En la tabla 3 se describen las 14 medidas contenidas en cada uno de los hechos anteriormente descritos. El tipo de datos de todas estas es double.

Tabla 3 Descripción de las medidas

Medidas	Descripción
precio_unit	Es el precio en condiciones FOB ¹² de cada unidad del producto que se declara según factura comercial.
serv_usd	Se refleja el importe total en moneda libremente convertible que por derechos de aduana se hayan dejado de pagar por todas las subpartidas arancelarias.
serv_mn	Se refleja el importe total a pagar en moneda nacional del impuesto o tasa que ampara esa declaración.
aran_usd	Identifica a los totales por tipo de impuesto o tasa aplicable (Arancel y Servicio de Aduanas).
aran_mn	Se refleja la tasa de cambio vigente que fije el Banco Central de Cuba entre la moneda nacional y la moneda en que se declara.
base_imp	Se refleja la base para el cálculo de los derechos de aduana, o sea, el valor en aduana de las mercancías calculado según las normas de valoración vigentes.
val_otrogas	Se reflejan los gastos que se enumeran en las normas de valoración en aduanas.
val_flete	Indica el importe del flete ¹³ en dólar estadounidense que corresponde al producto.
cant_comp	Refleja la cantidad de artículos correspondiente a la mercancía declarada.
peso_bruto	Es la unidad de medida de la mercancía expresada en kilogramos, incluyendo el embalaje.
peso_net	Es la unidad de medida de la mercancía expresada en kilogramos, sin incluir el embalaje.
val_est	Se reflejará el valor (CIF ¹⁴ para las importaciones y FOB para las exportaciones) de las mercancías incluyendo todos los elementos de

¹²FOB: Término de comercialización internacional que indica el precio de venta de un determinado artículo incluye el valor de la mercancía y los gastos de transporte y maniobra.

¹³ Flete: Tarifa básica pactada entre el transportador y el usuario del servicio, en el cual el primero se compromete a trasladar la mercancía desde un punto de origen hasta el de destino acordado previamente.

	costos.
val_fact	Se indicará el valor factura en dólar estadounidense según condición de entrega FOB, que corresponda al producto.
val_seg	Se indicará el importe del seguro en dólar estadounidense que corresponde al producto.

2.2.2 Matriz dimensional

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

Tabla 4 Matriz dimensional

hech/dim	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
hech 1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
hech 2	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

2.2.3 Modelo de datos

A continuación se muestra una porción del modelo de datos actual, en el se relacionan las principales dimensiones y los dos hechos. Cada hecho esta relacionado con cada una de las dimensiones redefinidas, a diferencia del modelo de datos de la tesis precedente en el que las dimensiones dim_régimen_exp y dim_régimen_imp solo se relacionaban con sus hechos correspondientes.

¹⁴ CIF: Término de comercialización internacional que indica el precio de la mercancía incluyendo el costo, seguro y fletes.

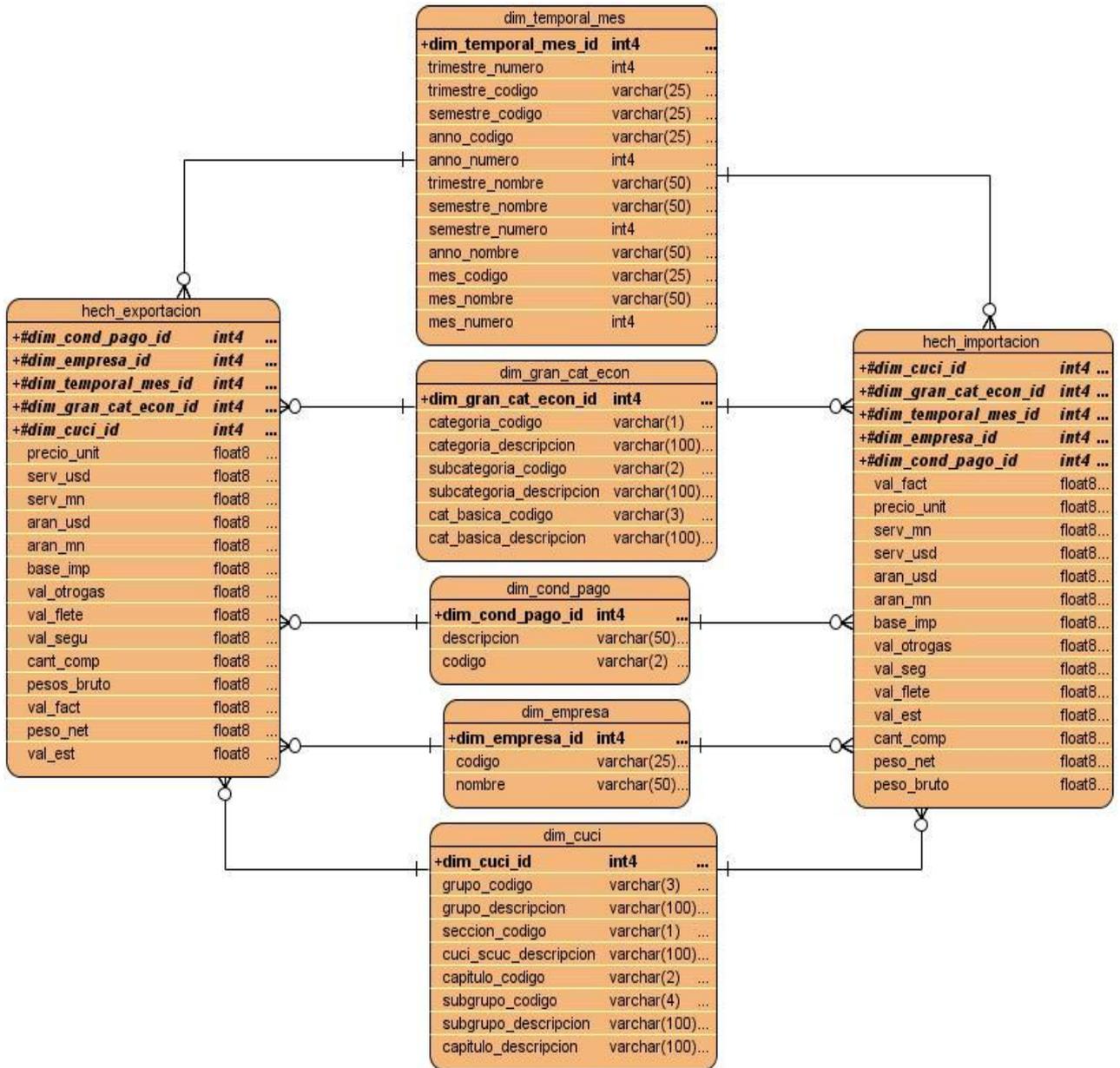


Figura 6 Modelo de datos

2.2.4 Política de respaldo y recuperación.

La política de respaldo y recuperación que se empleará en el mercado de datos está condicionada por 3 puntos fundamentales:

- Periodicidad de salvados del sistema: los salvados se harán en diferentes momentos del año, en un período aproximado de 30 días (copias parciales) y en los meses de julio y diciembre (copia total).
- Tablas involucradas: las tablas que normalmente se involucran en la realización son `hech_importacion` y `hech_exportacion`. Las tablas de dimensiones también estarán implicadas en caso de cambiar.
- Salvados existentes: actualmente en el área Comercio exterior se realizan salvados con la periodicidad mencionada anteriormente. La información salvada se almacena en dos lugares diferentes: en la propia entidad y fuera de ella.

2.2.5 Esquema de seguridad

Uno de los aspectos fundamentales para las aplicaciones es aplicar una buena política de seguridad. Si esto no se realiza de forma correcta, puede llegar a ser una carga que afectará el desarrollo del sistema. La seguridad para el mercado de datos Comercio exterior está dada en gran medida por los niveles de acceso al sistema, apoyándose específicamente en los roles definidos para la interacción de los usuarios con la base de datos y con la aplicación.

Seguridad en la base de datos

Se definió un grupo de usuarios que se muestran en la tabla 5. Estos manejan el acceso a las funcionalidades que le corresponde a cada rol para la interacción con la base de datos.

Tabla 5 Usuarios en la base de datos

Actor	Descripción
Administrador ETL	Tiene permiso solo para realizar las funciones que se incluyen en los procesos ETL.
Administrador	Tiene acceso total a la base de datos. Realiza las operaciones de administración en la misma.
Analista	Tiene acceso de solo lectura a la base de datos. Accede a la misma solo mediante la aplicación.

Seguridad en la aplicación.

Debido al creciente incremento que se evidencia actualmente de las aplicaciones desplegadas en el Servidor de Inteligencia de Negocios de Pentaho y de los usuarios que demandan acceso a estas, es de gran importancia que se defina un esquema de seguridad que haga sostenible el manejo de la misma en este servidor. A continuación se muestran los actores definidos para que el acceso a la aplicación sea de forma organizada.

Tabla 6 Usuarios en la aplicación

Actores	Descripción
Administrador	Tiene acceso total el Sistema de Información del Gobierno. Realiza las operaciones de administración sobre el mismo.
Analista	Tiene acceso de solo lectura al Área de Análisis Comercio exterior. Analiza la información contenida en los reportes.

Para lograr una mejor organización y mayor rendimiento en la seguridad del sistema, la herramienta Pentaho BI Server establece tres aspectos fundamentales, estas son: (Pentaho 2005)

Seguridad de acceso a datos de objetos: Incluye usuarios, contraseñas, autorizaciones permitidas, recursos web y protección a datos.

Autenticación: Tiene que ver con el procesamiento de información interactiva de inicio de sesión (por ejemplo nombre de usuario y contraseña) comparándola con la información recuperada del almacén de datos de seguridad.

Autorización de recursos web (URL): Brinda protección a las URL para responder a cada usuario si pueden o no acceder a una determinada página. Esto es decidido por el administrador de recursos web, el cual le brinda a cada usuario autenticado un permiso de seguridad, delimitando las páginas a las que tiene acceso y a las que no.

2.3 Conclusiones del capítulo

Una vez finalizado el capítulo y con este el análisis y diseño del mercado de datos, se puede concluir que:

- Se realizó el análisis del mercado de datos Comercio exterior, mediante el cual se refinaron los requisitos obtenidos en la tesis precedente y se identificaron además otros requisitos para lograr satisfacer las nuevas necesidades de los clientes.
- Fueron refinadas las reglas de negocio existentes. Además, se identificaron otras que serán empleadas en la construcción del mercado de datos, permitiendo definir y controlar la estructura y el funcionamiento de la organización.
- Se realizó la especificación de los casos de uso del sistema, identificándose nuevos casos de uso de información y funcionales mediante los cuales se logró una mejor estructuración de los procesos del negocio.
- Se refinó el modelo de datos existente, identificando nuevas dimensiones y refinando algunas existentes, garantizando el correcto funcionamiento del sistema.
- Se realizó el diseño de los subsistemas de integración y de visualización que posteriormente serán implementados como finalidad del sistema, así como también el diseño de los casos de prueba mediante los cuales se realizará la validación del sistema.

CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBA DEL MERCADO DE DATOS COMERCIO EXTERIOR

El presente capítulo comprende la implementación de la base de datos, del subsistema de integración y del subsistema de visualización. Mediante los procesos Extracción, Transformación y Carga e Inteligencia de Negocio se le dará solución a las necesidades del negocio. Además se plasma la validación de la solución propuesta mediante las listas de chequeo y la aplicación de los casos de pruebas.

3.1 Implementación de la base de datos

3.1.1 Modelo de datos físico.

El modelo de datos es una colección integrada de conceptos que permite describir los datos y las relaciones que existen entre ellos. Permite describir las estructuras de datos, las restricciones de integridad y las operaciones de manipulación de los datos. El modelo físico se obtiene a partir del modelo lógico dimensional. Recoge las relaciones existentes entre las tablas de hechos y las dimensiones que conforman el mercado de datos. Se utiliza para describir los datos a un nivel interno, proporciona detalles de cómo son almacenados los datos en el ordenador.

Convenciones de nombrado

Los elementos de los mercados de datos se nombran de una manera estándar teniendo en cuenta el tipo de estructura que se maneje. Existen convenciones de nombrado que son empleadas con el objetivo de establecer un vocabulario común y homogéneo que permita un entendimiento claro y conciso de las estructuras de los almacenes de datos. En la siguiente tabla se muestran las convenciones de nombrado empleadas en la implementación del mercado de datos Comercio exterior; los elementos a los cuales se hace referencia son las tablas de hechos, tablas de dimensiones, llaves primarias y atributos compuestos.

Tabla 7 Convenciones de nombrado

Estructura	Descripción	Ejemplo
Tablas de hechos.	Todas las tablas de hechos tendrán una cadena que demuestra que son hechos y el concepto que describen. Separados por el carácter _.	hech_<concepto>
Tablas de dimensiones.	Todas las tablas de dimensiones	dim_<concepto>

	tendrán una cadena que demuestra que son dimensiones y el concepto que describen, separadas por el carácter _.	
Llaves primarias.	Todas las llaves primarias están conformadas por el nombre de la tabla a la que pertenecen y una cadena que demuestra que son llaves primarias, separados por el carácter _.	<nombre_tabla>_id
Atributos compuestos.	En los atributos donde el nombre es compuesto se debe especificar el primer componente del atributo separado del segundo por el carácter _.	<Primer nombre>_<Segundo nombre>

Índices

Un índice es una estructura de disco asociada con una tabla o una vista que se utiliza, principalmente, para mejorar el rendimiento de una base de datos. Contiene claves generadas a partir de una o varias columnas de la tabla o la vista. Se utilizan principalmente en los campos sobre los cuales se hacen búsquedas con más frecuencia. Para la implementación del mercado de datos Comercio exterior se propone como indexado la utilización de las llaves primarias. Todas las llaves primarias poseen índices de tipo “b-tree” (Árboles balanceados), significa que cualquier búsqueda que se realice utilizando las llaves se optimizará mediante este método.

Esquemas y tablas

Los esquemas son una representación gráfica de la estructura de bases de datos. Posibilitan que la información almacenada en la base de datos este organizada, estos son contenedores de funciones, tipos de datos y operadores.

Para el desarrollo del mercado de datos Comercio exterior se definieron 3 esquemas:

- Esquema dimensiones: Contiene todas las tablas de dimensiones de la base de datos de la ONE que son comunes para los mercados de datos.
- Esquema mart_comercio_ext: Contiene todas las tablas de hechos propuestas del mercado de datos Comercio exterior, además de algunas tablas de dimensiones que son propias del mismo.

- Esquema metadatos: Contiene las tablas que guardan la información de las trazas de las cargas de los datos.

En la solución se definieron 2 tablas de hechos, 18 tablas de dimensiones y 2 tablas adicionales que contienen la información de las trazas de las cargas de los datos, estas quedaron distribuidas en los esquemas mencionados de la siguiente forma:

Tabla 8 Correspondencia entre esquemas y tablas del mercado de datos.

Esquemas	Tablas
dimensiones	dim_aduana dim_cuci dim_gran_cat_econ dim_pais dim_regimen dim_subpartida dim_temporal_mes dim_um
mart_comercio_ext	dim_ciiu dim_clase_bulto dim_cond_pago dim_empresa dim_grupo_prod dim_modo_transp dim_moneda dim_nac_transp dim_prod_azuc dim_prod_selec hech_exportacion hech_importacion
metadatos	meta_job_info meta_transf_info

3.1.2 Roles y privilegios establecidos en la base de datos

La seguridad de la base de datos mart_comercio_ext queda establecida de acuerdo a los privilegios asignados a los usuarios pertinentes según el rol a desempeñar.

A continuación se especifican los roles y sus privilegios:

- Administrador de ETL: realiza los procesos de ETL en la base de datos del sistema.
- Administrador: tiene acceso a la base de datos en su totalidad, ya sea para la administración o configuración de la misma.
- Analista: solo tiene permiso para consultar la información contenida en la base de datos. Realiza la operación de tipo selección sobre los datos.

3.2 Implementación del subsistema de integración

3.2.1 Arquitectura de Integración

El proceso de desarrollo de un software requiere de la definición previa de una arquitectura. Esta representa la guía de construcción de un sistema. No es más que un conjunto de estructuras que posibilita a los desarrolladores una línea común para alcanzar los objetivos cubriendo todas las necesidades.

La arquitectura de este proceso está conformada por algunos elementos necesarios en la implementación de dicha solución.

Fuente de datos: son los ficheros provenientes del sistema fuente, los cuales se encuentran en formato DBF (Archivo de Base de Datos). Estos sufrirán un proceso de extracción para posteriormente ser modificados según las necesidades de información existentes.

Área temporal: es el área en la cual se prepararán los datos. Esta preparación incluye la limpieza y la transformación de los mismos. También es llamada Staging area, y es lo que separa el sistema fuente del almacén de datos.

Mercado de datos: es el destino hacia donde se cargaran los datos integrados.

La arquitectura del proceso queda representada de la siguiente manera:

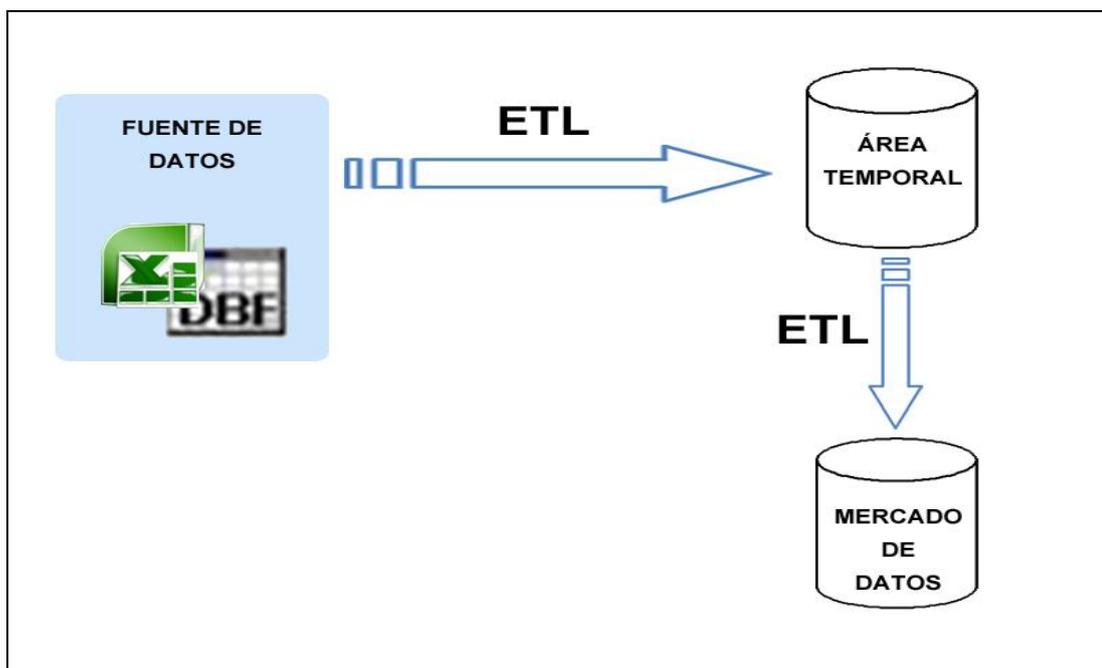


Figura 7 Arquitectura del proceso de integración.

3.2.2 Proceso de Integración de datos

Perfilado de datos

El perfilado de los datos se realiza para tener un conocimiento del estado de los datos fuentes. A partir de esto se establecen reglas de acuerdo a las necesidades del cliente. Este proceso se llevó a cabo mediante la herramienta Data Cleaner en el cual se generaron los reportes Estándares de medidas, Análisis de cadenas y Análisis numérico permitiendo hacer el análisis pertinente. A continuación se exponen las imágenes que muestran los resultados obtenidos:

En el reporte estándar de medidas se evidencia la cantidad de filas que posee la fuente de datos, en este caso consta con 100 filas. Además se dan a conocer los campos con valores nulos, vacíos, y el mayor y menor valor que trae cada uno de estos en la fuente de datos.

Standard measures									
	est_fiscal	operac	nodm	val_fact	cond_pag	val_flet	peso_net	cond_acu	cod_unim
Row count	100	100	100	100	100	100	100	100	100
Null values	0	0	0	0	0	0	0	100	0
Empty valu...	0	0	0	0	0	0	0	0	0
Highest val...	0000	E 00018	6584.18	01	0.00	1022.40000	<null>		03
Lowest val...	0000	E 00002	4.32	01	0.00	0.02400	<null>		01

Figura 8 Reporte estándar de medidas.

En el reporte análisis de cadena se muestra por campo la cantidad de caracteres que pueden tener, el máximo y mínimo de palabras, el porcentaje de mayúsculas entre otras informaciones.

String analysis					
	subpartida	cond_pag	provdest	noadic	paispr1d
Char count	800	200	600	200	200
Max chars	8	2	6	2	2
Min chars	8	2	6	2	2
Avg chars	8	2	6	2	2
Max white ...	0	0	0	0	0
Min white s...	0	0	0	0	0
Avg white s...	0	0	0	0	0
Uppercase ...	0%	0%	33%	0%	100%
Lowercase ...	0%	0%	0%	0%	0%
Non-letter ...	100%	100%	66%	100%	0%
Word count	100	100	100	100	100
Max words	1	1	1	1	1
Min words	1	1	1	1	1

Figura 9 Reporte análisis de cadena

El reporte análisis numérico tiene en cuenta el valor más alto y más bajo que toma cada campo, entre otras informaciones.

Number analysis					
	serv_usd	serv_mn	peso_bru	base_imp	val_est
Highest val...	<null>	548,75	3850	312963,15	312963,15
Lowest value	<null>	0,01	0,04	4,2	4,2
Sum	<null>	918,07	11786,86	1037315,91	1038580,46
Mean	<null>	12,08	117,87	10477,94	10385,8
Geometric ...	<null>	0,9	7,86	472,45	477,13
Standard d...	<null>	63,77	454,34	47388,77	47157,83
Variance	<null>	4067,08	206423,32	2245695983,08	2223861050,55

Figura 10 Reporte análisis numérico

Extracción, transformación y carga de los datos

Durante la realización del proceso ETL del mercado de datos se realizaron 20 transformaciones, 18 para la carga de las dimensiones y dos para la carga de los hechos. Posteriormente se cargaron los datos en el mercado. A continuación se muestra en la figura el ejemplo de una transformación para la carga de la dimensión aduana.

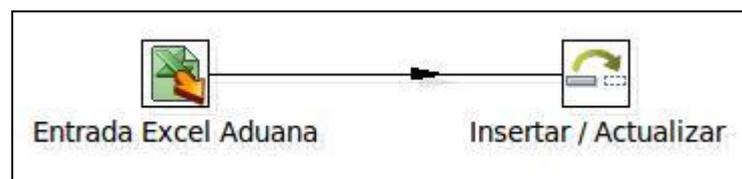


Figura 11 Transformación para la carga de la dimensión aduana.

Los datos correspondientes a la carga de la dimensión aduana no requieren de ser transformados, deben ser cargados en el mercado de datos exactamente cómo mismo son extraídos. El componente Entrada de Excel permite extraer los datos del fichero que se desea cargar. El Insertar / Actualizar es el componente que carga la tabla destino, su función es insertar o actualizar las filas de la tabla cada vez que se ejecute la transformación.

La figura 12 muestra la transformación realizada para la carga del hecho exportación. Inicialmente se extraen los datos de la fuente y se valida que no provengan nulos. Posteriormente se realiza una unión por clave, para el mismo hecho existen fuentes diferentes y es necesario unir los datos que son comunes para situarlos en un mismo componente. Se realizan además las búsquedas en la base de

datos para buscar las coincidencias entre la fuente y las dimensiones, se valida que los identificadores resultantes de esta búsqueda no provengan nulos y finalmente se inserta la información en la tabla destino.

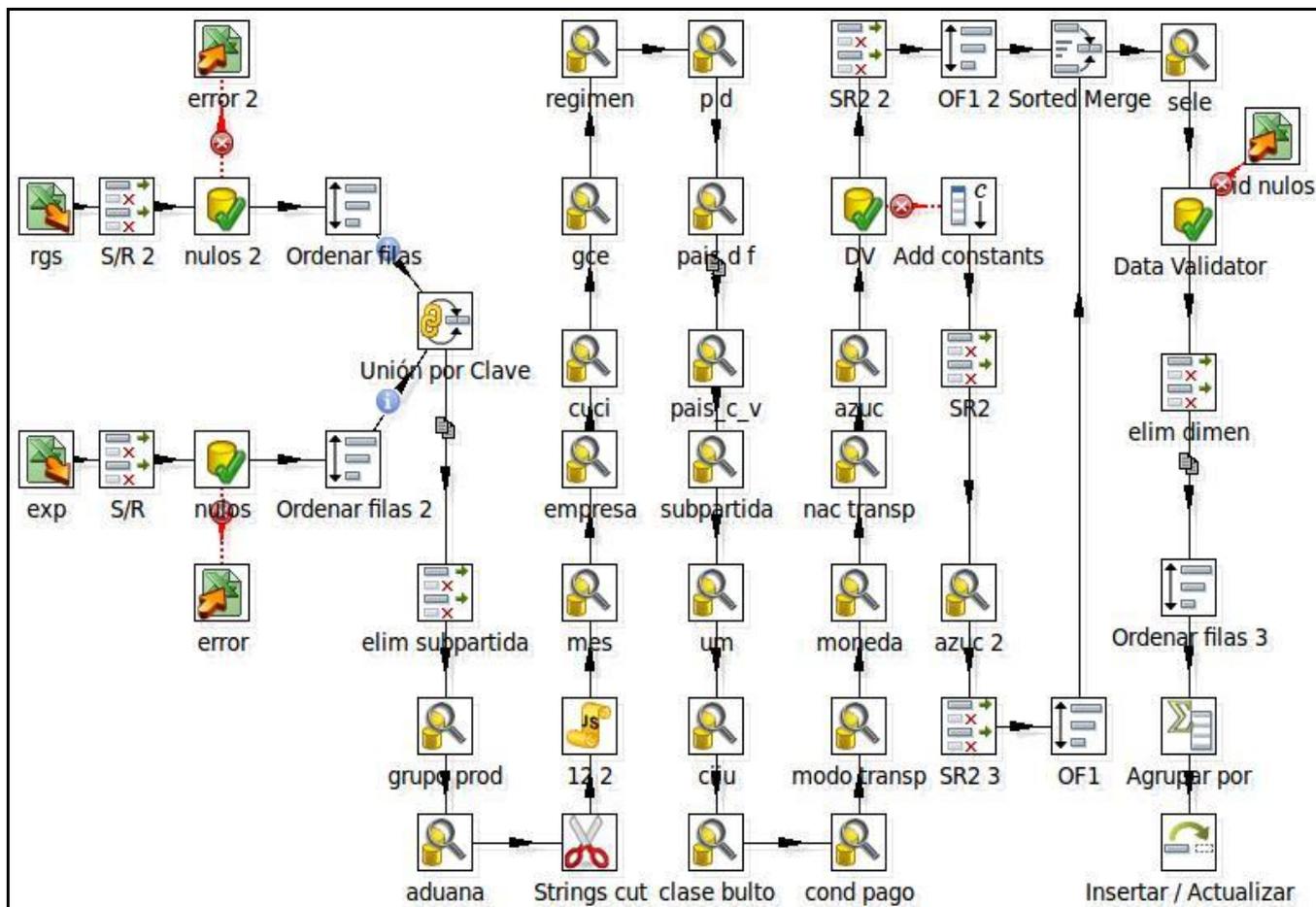


Figura 12 Transformación para la carga del hecho exportación

Trabajo para organizar el orden de la carga

El trabajo define el flujo de ejecución de las transformaciones. Una vez inicializado, se efectúa la conexión a la base de datos, posteriormente se ejecutan las transformaciones correspondientes a las dimensiones propias del mercado y por último se ejecutan las transformaciones para la carga de los hechos exportación e importación.

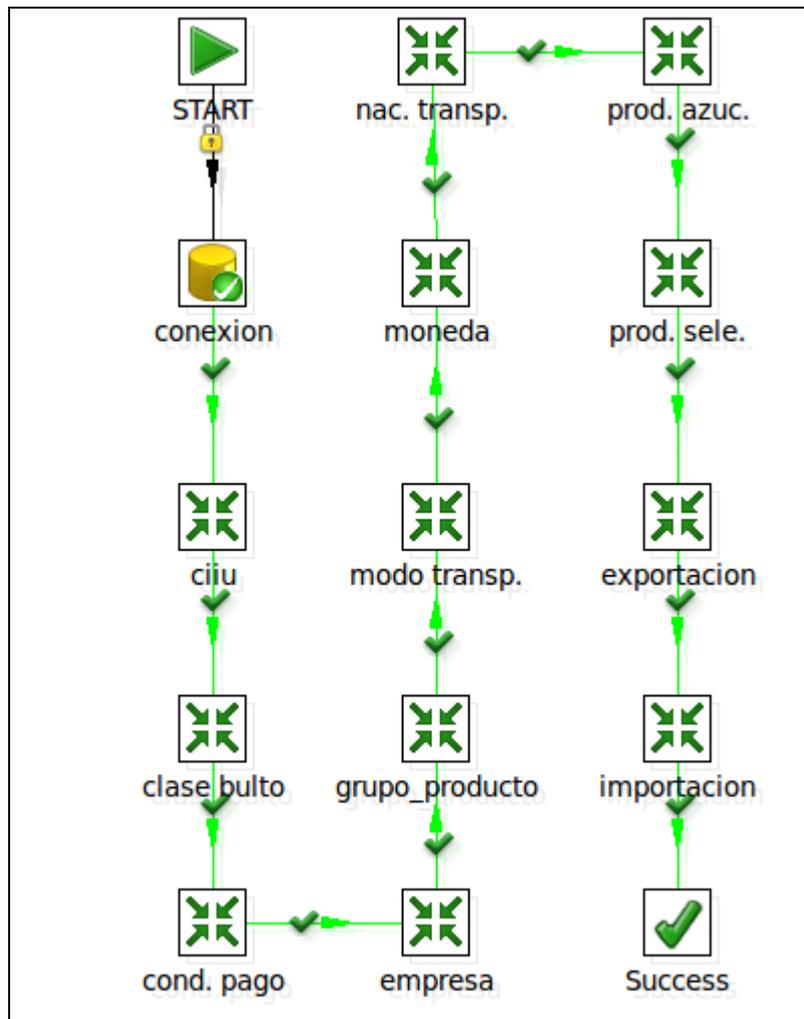


Figura 13 Trabajo del mercado de datos

Una vez finalizado el proceso de ETL, se registran en un directorio los datos con defectos para que posteriormente puedan ser analizados por los clientes y a partir de este análisis, darle algún tratamiento en dependencia del negocio.

3.2.3 Trazas de las cargas de los datos

Durante la implementación del subsistema de integración es importante almacenar las trazas de las cargas de los datos. Mediante estas se puede registrar un grupo de información que posteriormente le será muy útil al usuario (administrador_etl) para verificar la ejecución del proceso ETL; además, constituyen el punto de partida que posibilita la recuperación ante los fallos ocurridos durante las cargas de la información. Para el mercado de datos Comercio exterior, las trazas son almacenadas en dos tablas contenidas en el esquema metadatos:

- Tabla meta_transf_info: guarda toda la información relacionada con la ejecución de las transformaciones.
- Tabla meta_job_info: se almacena la información relacionada con la ejecución del trabajo.

3.3 Implementación del subsistema de visualización

3.3.1 Cubos multidimensionales

En el desarrollo de este proceso el esquema multidimensional es creado mediante la herramienta Pentaho Schema Workbench. Este permiten organizar los datos para que se ajusten al modo que tienen los usuarios de analizarlos: en categorías jerárquicas y en valores de resumen previamente calculados. En la solución se crearon 2 cubos multidimensionales, correspondientes a cada una de las tablas de hechos identificadas: importación y exportación y un cubo virtual nombrado intercambio, el mismo se crea porque para dos reportes se necesitan medidas que están en diferentes cubos, y el cubo virtual brinda la posibilidad de utilizar elementos de otros cubos.

En el esquema generado, el cubo importación contiene 10 dimensiones, 2 medidas y el hecho correspondiente. El cubo exportación posee 12 dimensiones, 2 medidas y la tabla de hecho que le corresponde. El cubo intercambio tiene 5 dimensiones, 2 medidas y 2 miembros calculados. La siguiente imagen muestra el esquema multidimensional que fue diseñado para el mercado de datos Comercio exterior, en el mismo están representados los tres cubos antes mencionados y las dimensiones que fueron definidas para ser usadas de forma general por estos.

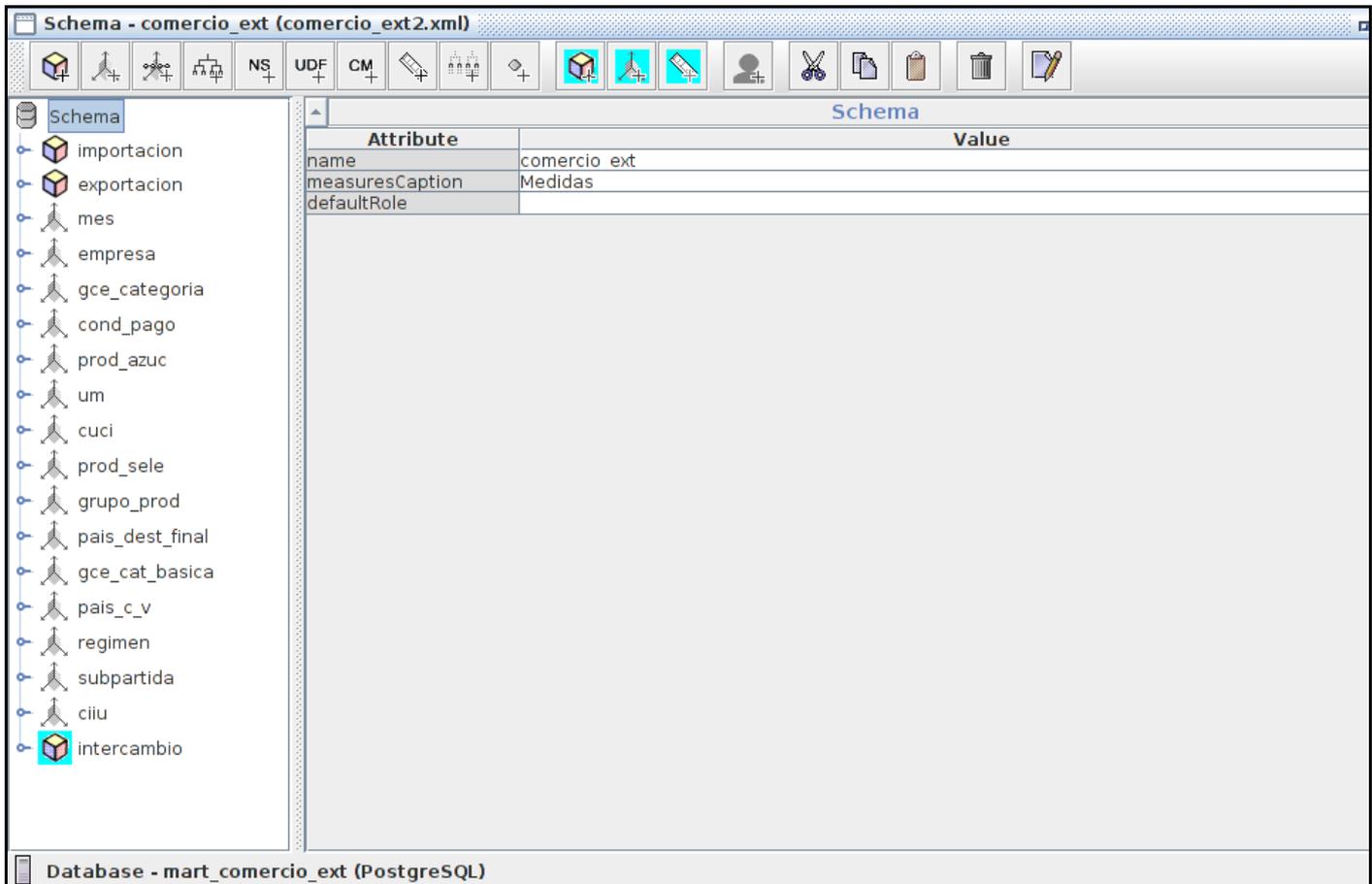


Figura 14 Diseño de los cubos multidimensionales

El cubo multidimensional importación contiene los valores de la tabla de hecho importación, además de las dimensiones y medidas de las vistas de análisis correspondientes.

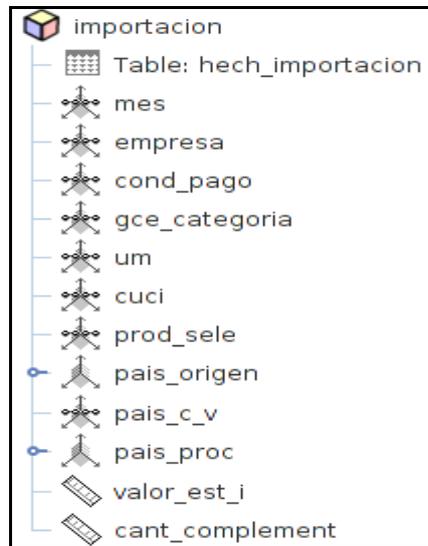


Figura 15 Elementos del cubo importación

El cubo multidimensional exportación contiene los valores de la tabla de hecho exportación; además de las dimensiones y medidas necesarias para las vistas de análisis correspondientes.

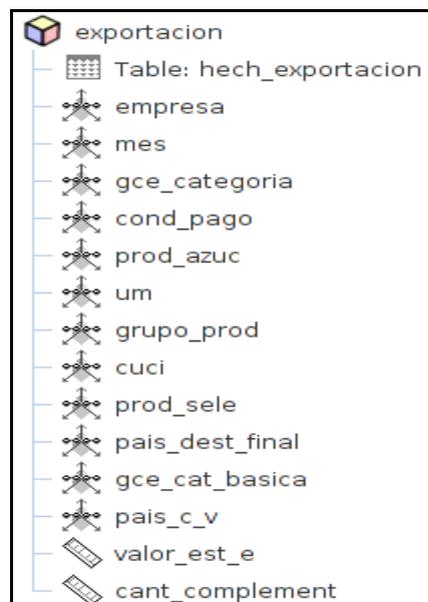


Figura 16 Elementos del cubo exportación

El cubo virtual intercambio contiene las dimensiones, medidas, y miembros calculados que son necesarios para las vista de análisis correspondientes al intercambio comercial.

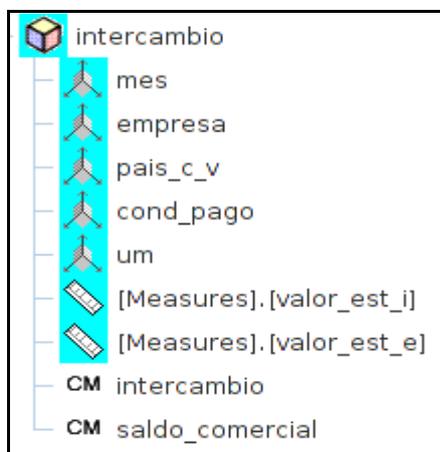


Figura 17 Elementos del cubo intercambio

3.3.2 Arquitectura de Información

La arquitectura de información permite estructurar y organizar los contenidos para ser utilizados por los usuarios de una manera más simple y directa. En el mercado de datos Comercio exterior la capa de visualización está representada por la estructura de navegación siguiente: 1 Área Análisis General (A.A.G), 1 Área de Análisis (A.A) y 3 Libros de Trabajo (L.T). Cada Libro de Trabajo tiene las vistas de análisis asociadas, suman entre todas 12 vistas.

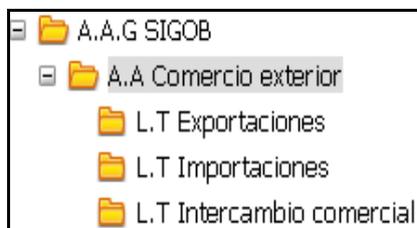


Figura 18 Arquitectura de información

A continuación se describe la estructura de navegación:

A.A.G SIGOB: Área Análisis General Sistema de Información del Gobierno, contiene la información de todos los mercados de datos que componen el almacén.

A.A Comercio exterior: Área de Análisis Comercio exterior, controla la información referente a las exportaciones e importaciones que están contenidas en los libros de trabajos.

L.T Exportaciones: Agrupa la información que se refiere a las exportaciones. Contiene los siguientes reportes:

- Mercancías agrupadas en GCE por la CUCI.
- Por GCE con donativos.
- Por GCE sin donativos.
- Por GCE y países.
- Por productos de la industria azucarera con donativos.
- Por productos de la industria azucarera sin donativos.
- Por productos seleccionados.

L.T Importaciones: Agrupa la información que se refiere a las importaciones. Contiene los siguientes reportes:

- Por GCE.
- Por GCE y país origen.
- Por productos seleccionados.

L.T Intercambio Comercial: Agrupa la información que se refiere al intercambio comercial. Contiene los siguientes reportes:

- Por empresas.
- Por países.

Vistas de análisis:

Ya finalizada la implementación del subsistema de visualización deben quedar conformadas las vistas de análisis que fueron solicitadas por el cliente. Estas poseen gran importancia, muestran la información que el usuario quiere que se visualice como finalidad del sistema para a partir de esta realizar el proceso de toma de decisiones. A continuación se muestra la imagen de la vista de análisis Mercancías agrupadas en Grandes Categorías Económicas por la Clasificación Uniforme para el Comercio Internacional perteneciente al Libro de Trabajo Exportaciones.

		Año	
		+ 2009	+ 2010
		Condición de pago	Condición de pago
		+ Todos	+ Todos
		Medidas	Medidas
CUCI	Categoría	● Valor estadístico EXP (MP)	● Valor estadístico EXP (MP)
- Todos	+ Todas	82870.41	61191.52
+ 0	+ Todas	54921.02	26333.12
+ 1	+ Todas	3038.81	11667.04
+ 4	+ Todas	89.89	36.53
+ 5	+ Todas	5545.69	2509.76
+ 6	+ Todas	3398.52	9114.48
+ 7	+ Todas	10905.05	2289.02
+ 8	+ Todas	4971.43	9241.56

Figura 19 Mercancías agrupadas en GCE por la CUCI.

En esta vista de análisis la información se analiza de acuerdo al valor estadístico las exportaciones que se efectúan según la Clasificación Uniforme para el Comercio Internacional y las Grandes Categorías Económicas; en los años 2009 y 2010, además de la condición de pago con que es efectuada la exportación.

La vista puede ser modificada según la estructura que el usuario desee ver. Mediante los componentes que se muestran en la barra de herramientas en la figura, se puede cambiar la estructura de la vista configurando el navegador OLAP, la consulta MDX o mostrando las propiedades.

3.4 Resultados obtenidos

Con el desarrollo del mercado de datos Comercio exterior se realizó el proceso de perfilado de los datos mediante el cual se obtiene una visión del estado en que se encuentra la fuente que posteriormente va a ser sometida al proceso ETL. La implementación del subsistema de integración permitió que la información procedente de varias fuentes estuviera integrada y con buena calidad, centralizada en un sistema de soporte de almacenamiento de datos el cual fue diseñado para ser flexible a cambios. Además, se implementaron varias vistas de análisis que poseen información actualizada la cual puede ser consultada por cualquier persona sin necesidad que esta sea un especialista en informática o información con altos conocimientos del negocio. Con estas vistas de

análisis se solucionaron las principales necesidades de información de los clientes, contribuyendo a la mejora del proceso de toma de decisiones de los directivos del Comercio exterior en la ONE.

3.5 Validación de la solución

Algunas de las etapas implicadas en el ciclo de vida de un software son: análisis de requerimientos, diseño, implementación y prueba. Una vez que se ha realizado la etapa del proceso de desarrollo, se debe dar paso a la validación de la solución. Las pruebas de software son los procesos que permiten verificar y relevar la calidad de un producto. Son utilizadas para identificar posibles fallos de implementación, calidad y usabilidad de la aplicación. Para determinar el nivel de calidad se deben efectuar pruebas que permitan comprobar el grado de cumplimiento de las especificaciones iniciales del sistema.

Existen diferentes tipos de pruebas que pueden ser aplicadas, algunas de ellas son:

Prueba unitaria: Es el proceso de probar los componentes individuales de la solución. El propósito es identificar diferencias entre la especificación de los artefactos y el comportamiento real de cada módulo.

Prueba de integración: es el proceso en el cual los componentes son agregados para crear componentes más grandes. Es la prueba realizada para mostrar que aunque los componentes hayan pasado satisfactoriamente las pruebas de unidad, la integración de los componentes es incorrecta.

Prueba de sistema: Se refiere al comportamiento del sistema integrado. Durante la etapa de las pruebas unitarias y de integración deben haberse identificado la mayoría de las no conformidades. La prueba de sistema se aplica generalmente para probar los requerimientos no funcionales de la solución.

Pruebas de aceptación: Se realizan para probar que el sistema cumpla con los requerimientos especificados por el cliente.

3.5.1 Pruebas aplicadas al mercado de datos Comercio exterior

Pruebas unitarias

A medida que se fueron implementando los subsistemas se les aplicaron por separado las pruebas unitarias. Durante la validación del subsistema de integración se detectaron llaves duplicadas. Para solucionar este problema se agregaron a la transformación los componentes Ordenar filas y Agrupar por los cuales permitieron ordenar las tuplas por el identificador y luego agrupar en una sola fila todos

los identificadores repetidos con las medidas sumadas. Una vez cargados los datos se verificó además que durante la ejecución de la transformación no se perdiera ninguno.

Se validó además el subsistema de visualización detectando inicialmente que los totales que se mostraban no coincidían con los que estaban en la fuente. El problema quedó solucionado al comprobar que los datos ausentes se encontraban en el directorio donde se almacenan los datos defectuosos. Finalmente se comprobó que el total visualizado unido al total de los datos con defectos coincidía exactamente con la fuente, por lo que no había pérdida de información.

Aplicación de los Casos de Prueba

Un caso de prueba se diseña según las funcionalidades descritas en los casos de usos de información. Este diseño se elabora previamente a realizar las pruebas funcionales a la aplicación. Para realizar las pruebas al mercado de datos del área Comercio exterior fueron diseñados tres casos de pruebas. Estos fueron aplicados por parte de los especialistas del departamento. En el proceso se identificaron 7 no conformidades, de ellas 6 funcionales y 1 referente a la estructura del documento. En esta primera validación no proceden 2 de las no conformidades detectadas, las 5 restantes fueron solucionadas.

El mercado de datos fue validado además por los especialistas de calidad del centro DATEC arrojando durante este proceso 8 no conformidades, clasificándose las mismas en funcionales, de estas se le dio solución a 1, las 7 restantes no proceden.

Finalmente se realizó la revisión por el grupo CALISOFT con el objetivo de liberar la solución. Durante este proceso existieron 2 iteraciones. En la primera fueron arrojadas 21 no conformidades, de estas 11 no procedieron, las otras 10 fueron solucionadas. En la segunda iteración se detectaron 11 no conformidades, clasificadas como funcionales, todas estas fueron resueltas.

Aplicación de la lista de chequeo

Las listas de chequeo constituyen un mecanismo para el control de los riesgos y su función básica es la de detectar condiciones peligrosas que puedan generar incidentes al producto de software. Es un documento que tiene un conjunto de parámetros a medir sobre un aspecto determinado, sea documentación o aplicación. Es un instrumento de medición y evaluación que consiste básicamente en un formulario de preguntas referentes al atributo de calidad que se está probando y a las

características del documento en el caso de la documentación. Cada pregunta tiene asociada una evaluación en una escala que da una medida del grado de cumplimiento y disponibilidad de la propiedad evaluada, de esta manera se determina la evaluación del elemento probado.

Para elaborar la lista de chequeo del mercado de datos Comercio exterior, se tuvieron en cuenta elementos de evaluación que son importantes una vez realizado el proceso de ETL y BI, permitiendo recoger los puntos eficientes e ineficientes que posean dichos procesos. La lista de chequeo contiene diferentes indicadores a evaluar los cuales se encuentran distribuidos en tres secciones fundamentales:

- Estructura del documento: abarca todos los aspectos definidos por el expediente de proyecto o el formato establecido por el proyecto.
- Indicadores definidos: abarca todos los indicadores a evaluar durante la etapa de desarrollo del mercado.
- Semántica del documento: contempla todos los indicadores a evaluar respecto a la ortografía, redacción y demás.

La lista de chequeo definida se aplicó luego de que el sistema pasara por calidad del centro DATEC y las no conformidades detectadas fueran resueltas. Luego de aplicar esta herramienta, contando con un especialista del departamento, no se detectaron no conformidades.

3.6 Conclusiones del capítulo

Una vez finalizado el capítulo y con este la implementación y validación del mercado de datos Comercio exterior se concluye lo siguiente:

- Se realizó el perfilado de datos mediante el cual se obtuvo, a través de varios reportes generados por el mismo, un conocimiento más profundo del estado de los datos fuentes.
- Se realizó la implementación de la base de datos, quedando establecidos los esquemas y tablas de la misma.
- Se desarrolló la implementación del subsistema de integración, obteniendo como resultado un mercado de datos poblado y funcional.
- Se diseñaron los cubos multidimensionales en correspondencia con cada una de las tablas de hecho del mercado de datos y con las necesidades de información.

- Se realizó la implementación del subsistema de visualización, quedando conformadas las vistas de análisis de acuerdo a las necesidades de información del cliente.
- Se diseñaron y aplicaron los casos de prueba para validar la solución, arrojando algunas no conformidades durante todas las etapas que fueron solucionadas.
- Se diseñó una lista de chequeo para posteriormente aplicarla al mercado de datos.
- Se realizó la prueba de aceptación con el cliente, en la que surgieron nuevas vistas de análisis a petición de este, las mismas fueron implementadas.

CONCLUSIONES GENERALES

Al finalizar el desarrollo de la solución, se concluye lo siguiente:

- ✓ Con la confección del mercado de datos se dio respuesta a los problemas planteados.
- ✓ El refinamiento del análisis y diseño del mercado de datos cumplió con todo lo requerido en el negocio, satisfaciendo las nuevas necesidades y exigencias del cliente.
- ✓ Las estructuras dimensionales que fueron implementadas cumplen las condiciones necesarias para el proceso de integración de los datos.
- ✓ La implementación de los subsistemas de integración y de visualización permitió obtener como resultado un mercado de datos poblado y funcional, con información disponible para ser consultada por parte de los usuarios, apoyando el proceso de toma de decisiones.
- ✓ Se realizaron distintos tipos de pruebas que permitieron validar la calidad del producto a diferentes niveles, arrojándose algunas no conformidades que fueron solucionadas.

RECOMENDACIONES

- Poblar el mercado de datos con información real y ponerlo en explotación para probar las funcionalidades del mismo.
- Poblar el mercado de datos con toda la información existente en la ONE del área Comercio exterior para posteriormente aplicar otras técnicas de Inteligencia de negocio.

REFERENCIAS BIBLIOGRÁFICAS

1. SINNEXUS. *Business Intelligence*, 2008. [2010]. Disponible en: http://www.sinnexus.com/business_intelligence/index.aspx
2. TORRES, L. P. *Curso Almacenes de Datos: importancia del estándar.*, 2006.
3. INMON, W. H. *Building the Data Warehouse*, Third Edition. John Wiley & Sons, 2005. 427 p. 0-471-08130-2
4. KIMBALL, R. and M. ROSS. *The Datawarehouse Toolkit*. ELLIOTT, R., John Wiley & Sons, 2002. Second edition.
5. WOLFF, C. G. *Modelamiento Multidimensional* 2002.
6. DARIO, B. R. *Datawarehouse manager*, 2009. [2010]. Disponible en: <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse/34-datawarehouse-manager>
7. LANZILLOTTA, A. *Definición de OLAP – Tecnología OLAP*, Master Magazine, 2004.
8. OLAPX. *Que es OLAP*, 2002. [2010]. Disponible en: <http://www.olapxsoftware.com/es/WhatIsOlap.asp>
9. IBARRA, M. D. L. A. *Procesamiento Analítico en Línea*, 2005.67.
10. CURTO, J. *Data Warehousing, Data Warehouse y Datamart*, 2007.
11. GASCÓN, M. D. L. H. and V. CASTELLAR. *Cómo diseñar grandes variables en bases de datos multidimensionales*, 2003.
12. GREENHOUSE, S. *Data Mart*, 2010. [2010]. Disponible en: <http://www.swgreenhouse.com/Productos/Hi-Spins/DataMart.html>
13. KIMBALL, R. and J. CASERTA. *The DW ETL Toolkit Practical for Extracting, Cleaning, Conforming, and Delivered Data*. Canadá, Wiley Publishing, 2004. 526 p. 0-764-57923-1
14. TANDRÓN, I. C. *Técnicas y herramientas de extracción, transformación y carga de datos aplicadas a la seguridad ciudadana*. Ciudad Habana, Universidad de las Ciencias Informáticas, 2008.81. p.
15. NEVES, F. D. *Los beneficios de la inteligencia de negocios*, 2007.
16. ROMERO, L. M. R. *Inteligencia de negocios para la toma de decisiones*, 2004.25.
17. ONUVA. *Inteligencia de Negocio: Conceptos y Aplicaciones con Pentaho*, 2008.25.
18. HUAMANTUMBA, R. *Datamart paso a paso*, 2007.
19. DATEC. *Metodología para el desarrollo de soluciones de almacenes de datos e inteligencia de negocios*, DATEC, 2009.13.

20. CUERVO, M. C. and O. Y. B. MORENO. *Herramientas libres para modelar software*, 2005.
21. DÍAZ, N. A. *PostgreSQL - La Base de Datos Open Source más avanzada del mundo*, 2003. [2010]. Disponible en: <http://soporte.tiendalinux.com/portal/Portfolio/postgresql.html>
22. FERREIRA, K. and J. SCHMIDT. *Sistemas de Información. Pentaho*, 2009.25.
23. PENTAHO. *Pentaho Open Source Business Intelligence*, 2005. [2010]. Disponible en: <http://www.pentaho.com>.
24. SUMMAN. *Pentaho BI Platform Server*, 2006. [2010]. Disponible en: <http://www.summan.com/index.php/productos/software/pentaho-.html>
25. TOMCAT, A. *Apache Tomcat*, 2010. [2010]. Disponible en: <http://tomcat.apache.org/>

BIBLIOGRAFÍA

- Cuervo, M. C. and O. Y. B. Moreno (2005) *Herramientas libres para modelar software*.
- Curto, J. (2007) *Data Warehousing, Data Warehouse y Datamart*.
- Dario, B. R. (2009) "Datawarehouse manager" Retrieved 24/11/2010, 2010, from <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse/34-datawarehouse-manager>
- DATEC (2009) *Metodología para el desarrollo de soluciones de almacenes de datos e inteligencia de negocios en DATEC*.
- Díaz, N. A. (2003) "PostgreSQL - La Base de Datos Open Source más avanzada del mundo." Retrieved 1/12/2010, 2010, from http://soporte.tiendalinux.com/portal/Portfolio/postgresql_html.
- Ferreira, K. and J. Schmidt (2009) *Sistemas de Información. Pentaho*.
- Gascón, M. d. I. H. and V. Castellar (2003) *Cómo diseñar grandes variables en bases de datos multidimensionales*
- Greenhouse, S. (2010) "Data Mart." Retrieved 1/12/2010, 2010, from <http://www.swgreenhouse.com/Productos/Hi-Spins/DataMart.html>.
- Huamantumba, R. (2007) *Datamart paso a paso*.
- Ibarra, M. d. I. A. (2005) *Procesamiento Analítico en Línea*.
- Inmon, W. H. (2005) *Building the Data Warehouse, Third Edition*, John Wiley & Sons, Inc.
- Kimball, R. (2005) *The Data Warehouse Lifecycle Toolkit*, John Wiley & Sons, Inc.
- Kimball, R. and J. Caserta (2004) *The DW ETL Toolkit Practical for Extracting, Cleaning, Conforming, and Delivered Data*. Canada, Wiley Publishing, Inc.
- Kimball, R. and M. Ross (2002) *The DataWarehouse Toolkit*. R. Elliott, John Wiley & Sons, Inc. Second edition.
- Neves, F. D. (2007) *Los beneficios de la inteligencia de negocios*
- OlapX. (2002) "Que es OLAP." Retrieved 2/12/2010, 2010, from <http://www.olapxsoftware.com/es/WhatIsOlap.asp>.

- Onuva (2008) *Inteligencia de Negocio: Conceptos y Aplicaciones con PENTAHO*.
- Pentaho. (2005) "Pentaho Open Source Business Intelligence." Retrieved 1/12/2010, 2010, from <http://www.pentaho.com>.
- Romero, L. M. R. (2004) *Inteligencia de negocios para la toma de decisiones*
- Sánchez, L. Z. C. (2008) *Metodología para el Diseño Conceptual de Almacenes de Datos*. Departamento de Sistemas Informáticos y Computación. Valencia, UNIVERSIDAD POLITÉCNICA DE VALENCIA: 211.
- Sinnexus. (2008). "Business Intelligence." Retrieved 11/11/2010, 2010, from http://www.sinnexus.com/business_intelligence/index.aspx.
- Summan. (2006). "Pentaho BI Platform Server" Retrieved 28/11/2010, 2010, from <http://www.summan.com/index.php/productos/software/pentaho-.html>.
- Tandrón, I. M. C. (2008) *Técnicas y herramientas de extracción, transformación y carga de datos aplicadas a la seguridad ciudadana*. Ciudad Habana, Universidad de las Ciencias Informáticas: 81.
- Tomcat, A. (2010). "Apache Tomcat." Retrieved 3/12/2010, 2010, from <http://tomcat.apache.org/>
- Torres, L. P. (2006) *Curso Almacenes de Datos: importancia del estándar*.
- Wolff, C. G. (2002) *Modelamiento Multidimensional*

GLOSARIO DE TÉRMINOS

BI: Business Intelligence.

CIIU: Clasificación Industrial Internacional Uniforme.

CUCI: Clasificación Uniforme para el Comercio Internacional.

Datamart: Mercado de datos.

Datawarehouse: Almacén de datos.

DBF: Archivo de Base de Datos.

DSS: Sistemas de Soporte a la Decisión.

DW: Datawarehouse.

EIS: Sistemas de Información de Ejecutivos.

ETL: Extracción Transformación y Carga.

Framework: Marco de trabajo.

FTP: Protocolo de Transferencia de Archivos.

GCE: Grandes Categorías Económicas.

HOLAP: Procesamiento Analítico en Línea Híbrido.

HTML: Lenguaje de Mercado de Hipertexto.

MOLAP: Procesamiento Analítico en Línea Multidimensional.

OLAP: Procesamiento Analítico en Línea.

ONE: Oficina Nacional de Estadísticas.

PDF: Formato de Documento Portátil.

RAM: Memoria de Acceso Aleatorio.

RN: Reglas de negocio.

ROLAP: Procesamiento Analítico en Línea Relacional.

SEN: Sistema Estadístico Nacional.

SQL: Lenguaje de Consulta Estructurado.

UCI: Universidad de las Ciencias Informáticas

UML: Lenguaje Unificado de Modelado.

URL: Localizador Uniforme de Recursos

XML: Lenguaje de Marcas Extensible.