

Universidad de las Ciencias Informáticas

Facultad 6



***Sistema de Información de Gobierno. Mercado de
datos para las series históricas del Sector externo***

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

Autora:

Yanisleidis Jorge Oduardo

Tutora:

Ing. Dunia Ortiz Nodal

Ciudad de la Habana, Junio 2011

“Año 53 de la Revolución”



*"En la tierra hacen falta personas que trabajen más y critiquen menos,
que construyan más y destruyan menos, que prometan menos y resuelvan más,
que esperen recibir menos y dar más, que digan mejor ahora que mañana"*

Che

Declaración de Autoría

Declaración de Autoría

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

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

Autora: _____

Yanisleidis Jorge Oduardo

Tutora: _____

Ing. Dunia Ortiz Nodal

Datos de Contacto

Datos de Contacto

Nombre y Apellidos: Ing. Dunia Ortiz Nodal.

Especialidad: Ingeniero en Ciencias Informáticas.

Graduada en el 2009 de Ingeniero en Ciencias Informáticas de la Universidad de Ciencias Informáticas (UCI). Actualmente trabaja adjunta a la UCI, pues esta brindando sus servicios como especialista de la Oficina Nacional de Estadísticas en los proyectos que se realizan en la universidad para esa entidad.

Agradecimientos

Agradecimientos

Ante todo quiero que sepan que mis palabras son incapaces de decir todo lo que siento en estos momentos, la emoción y la alegría que me da de verlos aquí en este uno de mis mejores momentos de toda la vida, eso puedo asegurarlo.

La primera persona que quiero mencionar en mis agradecimientos es a mi mamita, que es la principal razón por la cual ingresé en la UCI, por su apoyo y confianza durante mis años de estudio, sin eso no estuviese aquí. Gracias por darme la oportunidad de demostrarte que no fue en vano todo tu sacrificio. Gracias por darme el privilegio de ser tu hija, te quiero mucho.

Agradezco a mi papito que también quiero mucho, y a mi madrastra por todo su apoyo y buenos consejos.

A mi hermanita que sin darse cuenta por su edad, me impulsa a ser cada día mejor persona porque soy su ejemplo.

A toda mi familia sin excepción del más mínimo pariente, por confiar en mí y siempre darme ánimo para seguir adelante.

A mi novio y compañero por estar siempre a mi lado, alegrándome y apoyándome en los momentos más difíciles.

A mis suegros por los buenos consejos, y animarme para seguir hasta el final.

A todo mis amigos, los del grupo, las niñas y niños del apartamento, a mis inolvidables compañeros del secretariado de la FEM de tercer año, de la facultad y de la universidad.

A mis compañeras del trío de tesis de Comercio exterior por toda su ayuda.

A mi tutora por todo el apoyo y la preocupación, al igual a todos los profesores del proyecto por sacarme de algunos apuros.

A los profesores que de una forma u otra contribuyeron en mi desarrollo profesional, les agradezco por haberme preparado para el futuro.

Y a todo los que me ayudaron y apoyaron para que se hiciera posible la realización de este trabajo.

Y por último y no menos importante a Fidel y a la Revolución por haberme dado la oportunidad de estudiar en la UCI y formarme profesionalmente.

Dedicatoria

Dedicatoria

Le dedico esta investigación a toda mi familia, en especial a mi mamita, mi papito y mi hermanita por ser las personas que me impulsan para lograr todo lo que me propongo y mucho más. También a mi novio y a los amigos que me acompañaron durante la carrera. A todos por apoyarme para cumplir este sueño que siempre he deseado.

Resumen

Resumen

La Oficina Nacional de Estadística (ONE) es la empresa encargada del control estadístico en Cuba. Dicha empresa se ha trazado metas para lograr informatizar todas sus áreas, es el caso del Sector externo, específicamente para las series históricas. En esta área los datos son almacenados en formato Excel, por lo que se hace muy complejo el proceso de acceder y divulgar la información, obstaculizando así el proceso de toma de decisiones. Los datos no están integrados y solo pueden ser consultados por especialistas de informática e información que posean conocimientos del negocio. Además no existe una aplicación informática que ofrezca reportes flexibles con información actualizada. El presente trabajo de diploma tiene como objetivo desarrollar el mercado de datos y la capa de visualización de datos para las series históricas del Sector externo del Sistema de Información de Gobierno (SIGOB) que contribuya a la toma de decisiones. Se realizó el análisis, diseño, implementación y prueba de la aplicación arrojando como resultado un mercado de datos y una capa de visualización de datos, que brinda reportes actualizados para apoyar el proceso de toma de decisiones en el área del Sector externo en la ONE, específicamente para las series históricas.

Palabras claves: Mercado de datos, Sector externo, series históricas.

Índice

Índice

Introducción	1
Capítulo 1: Fundamentación Teórica del Almacén de Datos	5
1.1 Introducción	5
1.2 Indicadores del Sector externo	5
1.2.1 Indicadores del Sector externo a nivel Mundial	5
1.2.2 Indicadores del Sector externo en Cuba	6
1.3 Sistemas que controlan estadísticas del Sector externo en el mundo	7
1.4 Tecnologías de almacenamiento de datos	8
1.4.1 Almacenes de Datos	9
1.4.2 Mercados de datos	10
1.4.3 Procesos básicos de un AD	10
1.5 Inteligencia de Negocio (BI)	11
1.6 Metodología de Desarrollo	11
1.7 Modelo de Base Datos	13
1.7.1 Modelo dimensional	13
1.8 Modo de Almacenamiento de Datos	16
1.8.1 Proceso OLAP	16
1.8.2 Tipos de procesos OLAP	17
1.9 Sistema Gestor de Base Datos (SGBD)	18
1.9.1 PostgreSQL 8.4	19
1.10 Herramienta CASE	19
1.10.1 Visual Paradigm 6.4	20
1.11 Herramienta utilizada para la integración de datos	20
1.11.1 Pentaho Data Integration 4.0.1	20
1.12 Herramientas utilizadas para el proceso de BI	21
1.12.1 Pentaho Schema Workbench 3.2.0 (PSW)	21
1.12.2 Mondrian OLAP Server 3.0.4	21
1.11.3 Apache Tomcat 6.0	21
1.11.4 Pentaho BI Server 3.6.0	22
1.12 Conclusiones	22

Índice

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo	23
2.1 Introducción	23
2.2 Análisis	23
2.2.1 Estudio preliminar del negocio	23
2.2.2 Tema de análisis.....	24
2.2.3 Necesidades de información	24
2.2.4 Casos de Uso del Sistema	29
2.2.5 Reglas de Negocio.....	31
2.3 Diseño.....	32
2.3.1 Identificación de dimensiones, hechos y medidas.....	32
2.3.2 Matriz Bus o dimensional.....	35
2.3.3 Modelo de datos	35
2.3.4 Diseño del subsistema de integración de datos	37
2.3.5 Diseño del subsistema de visualización de datos.....	38
2.3.6 Esquema de seguridad	39
2.4 Política de respaldo y recuperación	40
2.5 Conclusiones	41
Capítulo 3: Implementación del Mercado de datos para las series históricas del Sector externo	42
3.1 Introducción	42
3.2 Implementación de la Base de Datos.....	42
3.3 Implementación de los procesos de ETL.....	43
3.3.1 Arquitectura del subsistema de integración.....	44
3.3.2 Implementación de los flujos de transformación.....	44
3.4 Implementación del subsistema de visualización de datos (BI)	47
3.4.1 Implementación de los cubos OLAP	47
3.4.2 Implementación de los reportes candidatos	48
3.4.3 Arquitectura de información	48
3.4.4 Configurar la seguridad de los usuarios	49
3.5 Conclusiones	50
Capítulo 4: Validación del Mercado de datos para las series históricas del Sector externo	51
4.1 Introducción	51

Índice

4.2 Validación del sistema	51
4.2.1 Pruebas aplicadas al Mercado de datos para las series históricas del Sector externo	51
4.3 Conclusiones	56
Conclusiones	57
Recomendaciones	58
Referencias Bibliográficas.....	59
Bibliografía.....	61
Anexos.....	64
Glosario	68

Índice de Figuras

Índice de Figuras

Figura 1 Relación de los grupos con los flujos de trabajo.....	13
Figura 2 Esquema en estrella	15
Figura 3 Esquema copo de nieve.....	15
Figura 4 Esquema constelación de hechos.....	16
Figura 5 Diagrama de Casos de Uso del Sistema.....	31
Figura 6 Modelo de datos	36
Figura 7 Diseño de transformación	37
Figura 8 Diseño del subsistema de visualización de datos.....	39
Figura 9 Estructura física de la BD.....	43
Figura 10 Arquitectura del subsistema de integración	44
Figura 11 Carga del hecho exportaciones.....	46
Figura 12 Trabajo.....	47
Figura 13 Implementación de los cubos OLAP	48
Figura 14 Consulta MDX Obtener exportaciones según país destino de las mercancías y áreas geográficas en un tiempo dado.....	48
Figura 15 Reporte Obtener exportaciones según país destino de las mercancías y áreas geográficas en un tiempo dado	49
Figura 16 Adicionar rol administrador.....	50
Figura 17 Adicionar rol analista.....	50
Figura 18 Comportamiento de los indicadores por secciones	53
Figura 19 Resultados de los casos de prueba	55

Índice de Tablas

Índice de Tablas

Tabla 1 Dimensiones identificadas.....	32
Tabla 2 Hechos identificados	34
Tabla 3 Medidas identificadas.....	34
Tabla 4 Relación entre los hechos y las medidas identificadas	34
Tabla 5 Matriz Bus o dimensional	35
Tabla 6 Usuarios en la base de datos	39
Tabla 7 Roles y permisos en la aplicación	40
Tabla 8 Esquemas y tablas del mercado de datos	42
Tabla 9 Sección Reportes candidatos.....	54
Tabla 10 Descripción de las variables.....	55
Tabla 11 Lista de Chequeo	64
Tabla 12 Caso de prueba para el caso de uso Consultar información de exportaciones.....	65
Tabla 13 Caso de prueba para el caso de uso Consultar información de importaciones.....	66

Introducción

Introducción

La informática ha sido una de las áreas del conocimiento que mayor desarrollo ha adquirido en los últimos tiempos, casi la totalidad de las disciplinas científicas requieren de ella. Cuando se habla de informática se define como unas de las materias más importantes en la actualidad, ya que por medio de esta la humanidad vive en una sociedad comandada por las nuevas tecnologías.

Las tecnologías de la información actualmente son elementos fundamentales para la superación y desarrollo de un país. Han sido conceptualizadas como la integración y convergencia de la informática, las telecomunicaciones y la técnica para el procesamiento de datos. Estas son el núcleo central de una transformación multidimensional que experimenta la economía y la sociedad. El estudio y dominio de las tecnologías de la información son de gran importancia, pues tiende a modificar los hábitos, patrones de conducta y la forma de pensar del ser humano como ente social.

Actualmente Cuba apuesta al desarrollo de la informática con el énfasis en la comunidad y como contribución a la cultura tecnológica de niños y jóvenes. Con este fin se han creado un sin número de programas entre los que se pueden mencionar: los joven club de computación, el equipamiento informático montado en todas las escuela del país, así como lo creación de la Universidad de las Ciencias Informáticas (UCI) en el año 2002. A pesar de su poco tiempo de fundada la UCI ha hecho grandes aportes al desarrollo del software cubano.

La UCI es una institución que se caracteriza por ser docente y productiva a la vez. En ella se forman profesionales comprometidos con la revolución y altamente calificados en la rama de la informática. En la misma se han creado los centros de desarrollo, entre los que se pueden mencionar el Centro de Tecnologías de Gestión de Datos (DATEC), el cual brinda su apoyo para el desarrollo de software. Este centro organiza la producción en líneas de trabajo, entre ellas la línea “Desarrollo de Almacenes de Datos”.

Actualmente llevar la estadística de un país o institución es muy importante, ya que se puede tener constancia de los crecimientos o decrecimientos que se producen de un año para otro en cualquier aspecto de la economía u otra esfera. En el mundo existen empresas élites encargadas del control de estadísticas, es el caso de la Oficina Nacional de Estadística (ONE) en Cuba.

La ONE fue creada como resultado de la reorganización de los organismos de la administración central del estado, a partir de la desaparición del comité estatal de estadísticas. Dicha entidad es una unidad presupuestada, creada para proponer, organizar y ejecutar, según corresponda, la aplicación de la política estatal en materia de estadística del país. El proyecto Sistema de Información del

Introducción

Gobierno (SIGOB) nace de la necesidad de centralizar toda la información histórica 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.

En la línea de trabajo encargada del desarrollo de almacenes de datos se trabaja en conjunto con la ONE en la elaboración de un Almacén de Datos (AD). Se han propuesto mejorar el acceso y manipulación del gran volumen de información que es llevada por esta entidad. Dicho AD es uno de los módulos que integra el proyecto SIGOB.

Atendiendo a las necesidades de la informatización en todo el país, la ONE se ha trazado metas para lograr informatizar todas sus áreas, como es el caso del Sector externo¹, específicamente para las series históricas, donde los datos son almacenados en formato Excel lo que dificulta la forma en que se realizan los principales reportes y cruces de variables, indicadores, tasas, porcentajes y demás aspectos de interés, afectándose considerablemente la disponibilidad de la información para órganos del estado. La información va acumulándose año tras año y es cada vez más complicado analizarla. Los datos no están integrados y solo pueden ser consultados por especialistas de informática e información que posean conocimientos del negocio. Además no existe una aplicación informática que brinde reportes flexibles con información actualizada para apoyar el proceso de toma de decisiones de los directivos del Sector externo en la ONE.

Por todo lo antes expuesto se propone como **problema a resolver**: ¿Cómo contribuir a la toma de decisiones en el área Sector externo del Sistema de Información de Gobierno específicamente para las series históricas?

Se propone como **objeto de estudio** los Almacenes de Datos e Inteligencia de Negocio y como **campo de acción** el Mercado de datos y capa de visualización para las series históricas en el área Sector externo del Sistema de Información de Gobierno.

Se plantea como **objetivo general** de esta investigación desarrollar el Mercado de datos para las series históricas del Sector externo del Sistema de Información de Gobierno que contribuya a la toma de decisiones.

Se definen como **objetivos específicos**:

¹ Sector externo: área en la que se gestionan todos los datos estadísticos referidos a las actividades de exportación e importación efectuadas en nuestro país.

Introducción

- Realizar el análisis y el diseño del mercado de datos para las series históricas del Sector externo.
- Implementar el mercado de datos para las series históricas del Sector externo.
- Implementar la capa de visualización de datos.
- Validar el mercado de datos para las series históricas del Sector externo.

Para el cumplimiento de los objetivos específicos de esta investigación se trazaron las siguientes **tareas**:

1. Fundamentación teórica de Almacenes de Datos.
2. Caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de Almacenes de Datos.
3. Levantamiento de requisitos.
4. Descripción de los casos de uso del mercado de datos.
5. Definición de los hechos, las medidas y las dimensiones del mercado de datos.
6. Diseño del modelo de datos.
7. Definición de la arquitectura del mercado de datos.
8. Diseño del subsistema de integración.
9. Diseño del subsistema de visualización.
10. Diseño de los casos de pruebas.
11. Implementación del subsistema de integración.
12. Implementación del subsistema de visualización.
13. Aplicación de las listas de chequeo.
14. Aplicación de los casos de pruebas.

El documento está compuesto por 4 capítulos, estos están estructurados de la siguiente manera:

Capítulo 1: Fundamentación Teórica del Almacén de Datos

Este capítulo está referido al marco teórico de la investigación. Se realiza el estudio de los sistemas de AD y mercados de datos y sus principales características, así como sus funcionalidades. Además se

abordan elementos de los procesos básicos de los AD y se realiza un análisis de la Inteligencia de Negocio (BI). También se define la metodología a utilizar para el desarrollo de la solución y las principales herramientas.

Capítulo 2: Análisis y diseño del Mercado de datos para la series históricas del Sector externo

En este capítulo se definen los temas de análisis, el negocio, los casos de uso del sistema y aspectos del diseño. Entre los contenidos a los que se hace referencia se encuentran los permisos, los requisitos de información, multidimensional, funcional y no funcional. Se definen las dimensiones del mercado de datos, los hechos y las medidas asociadas a los hechos definidos. Se estructura la matriz bus, el modelo dimensional y se realiza la elaboración tanto del diseño lógico como físico para dicho mercado de datos. Además se diseña el subsistema de integración, el de visualización y los casos de prueba.

Capítulo 3: Implementación del Mercado de datos para la series históricas del Sector externo

Los temas que aborda este capítulo están relacionados con todos los elementos de extracción, transformación y carga de los datos. Se implementa el modelo de datos, los flujos de transformación, los cubos multidimensionales y los reportes candidatos. Además incluye la configuración de la seguridad de los usuarios.

Capítulo 4: Validación del Mercado de datos para la series históricas del Sector externo

En este capítulo se hace referencia a la validación de la solución. Se describen pruebas existentes para la validación de una aplicación. Además se detalla el proceso de diseño y aplicación de de los casos de pruebas, y queda diseñada la lista de chequeo. También se hace referencia a la validación por parte del cliente mediante las pruebas de aceptación.

Capítulo 1: Fundamentación Teórica del Almacén de Datos

1.1 Introducción

Este capítulo está referido al marco teórico de la investigación. Se realiza el estudio de los sistemas de AD y mercados de datos y sus principales características, así como sus funcionalidades. Además se abordan elementos de los procesos básicos de los AD y se realiza un análisis de la Inteligencia de Negocio (BI). También se define la metodología a utilizar para el desarrollo de la solución y las principales herramientas.

1.2 Indicadores del Sector externo

1.2.1 Indicadores del Sector externo a nivel Mundial

El Sector externo en economía es el conjunto de actividades y estrategias que realizan los estados en el ámbito internacional para poder mantener un adecuado intercambio productivo, comercial y financiero, que asegure el desarrollo de sus pueblos. Este intercambio integra los diferentes sistemas de producción y consumo que existen dentro del marco de la economía mundial. Su existencia revela que la economía de un país está condicionada por la influencia de la economía de otros países, ya sea por los intercambios de productos, como de tecnología, servicios o capitales.

Durante cuarenta años las economías occidentales han confiado en el crecimiento de los Estados Unidos como motor del engranaje económico mundial. Ninguna de las economías capitalistas más prósperas han crecido sin una expansión de sus exportaciones, siendo el mercado norteamericano el principal factor estimulante.

La economía norteamericana ha jugado el papel de "locomotora" de la economía mundial. Esta situación ha ocasionado un inmenso costo en términos de desequilibrios macroeconómicos: un descomunal déficit fiscal y comercial que apenas parece ser controlable, y una creciente dependencia de los capitales extranjeros, tanto de préstamo como de riesgo, que hacen de los Estados Unidos el mayor deudor del mundo.

En las últimas décadas han comenzado a destacarse en el ámbito económico los países emergentes² que vienen mostrando una considerable pujanza y estabilidad económica que realza su importancia en el contexto de la economía internacional. Hay que destacar como actores principales de este grupo heterogéneo los denominados países BRIC (Brasil, Rusia, India, China). Estos países tienen un Sector externo que genera amplios flujos de exportación hacia los países desarrollados debido a los bajos

² Países emergentes: todos aquellos países que estando todavía en vías de desarrollo, vienen mostrando en los últimos años una considerable pujanza y estabilidad económica que realza su importancia en el contexto de la economía internacional.

costes de la mano de obra y a la creciente deslocalización de los centros productivos de los países desarrollados. Los principales destinos de sus relaciones comerciales son Estados Unidos y la Unión Europea. Se prevé un rápido crecimiento de estas economías y, posiblemente, superarán en conjunto al G-7³ ya en el 2032. (1)

En el 2010 la economía mundial comenzó a recuperarse de la crisis financiera. El progreso ya es evidente, aunque los ritmos de la recuperación económica son muy diferentes en sitios diferentes. Hoy en día, los mejores indicadores se registran en los países asiáticos, sobre todo en China, que supera a EEUU. La crisis económica afectó tanto a las economías desarrolladas como a las emergentes. Los países del BRIC actualmente tienen muchas posibilidades para superar la depresión en breve, utilizando cada uno su propio camino para hacerlo.

1.2.2 Indicadores del Sector externo en Cuba

Al triunfar la Revolución Cubana el Gobierno revolucionario promulgó mediante la Ley 793 del 25 de abril de 1960 la creación del Banco para el Comercio exterior de Cuba (BANCEC). BANCEC nace con el objeto de contribuir y colaborar con la política de comercio internacional del Estado Revolucionario y la aplicación de las medidas sobre el Comercio exterior adoptadas por el banco nacional.

El establecimiento del monopolio estatal del Comercio exterior de Cuba se logra con la creación del Ministerio del Comercio Exterior (MINCEX), el 23 de febrero de 1961 mediante la Ley 934. Para la economía de Cuba las relaciones externas y en específico el Comercio exterior ejercen un determinante impacto, debido a que la economía es abierta y mantiene una alta dependencia del mismo.

El bloqueo de los Estados Unidos ha privado a Cuba de importantes ingresos por exportaciones de bienes y servicios. Obstaculiza el acceso del país a las fuentes externas de financiamiento y provoca un incremento en los precios por concepto de reubicación geográfica del comercio. En el caso de las empresas de este sector, una de las principales afectaciones se deriva del encarecimiento del financiamiento externo, por las implicaciones que lleva consigo el elevado riesgo con el que son calificadas las operaciones cubanas. Por consiguiente, los financiamientos sólo se logran con tasas de interés por encima de las prevalecientes en el mercado internacional. Las afectaciones registradas desde abril 2009 a marzo de 2010 por las empresas del Comercio exterior se cifran en 155 millones 500 mil dólares. (2)

³ G-7: grupo de países industrializados del mundo cuyo peso político, económico y militar es muy relevante a escala global. Está conformado por Alemania, Canadá, Estados Unidos, Francia, Italia, Japón y Reino Unido.

Capítulo 1: Fundamentación Teórica del Almacén de Datos

En la actualidad se ha logrado una reorientación notable del intercambio comercial de Cuba, en busca de una diversificación de los socios comerciales. El proceso de descentralización del Comercio exterior se ha llevado a cabo de una forma organizada y paulatina.

1.3 Sistemas que controlan estadísticas del Sector externo en el mundo

El comercio “sin papel” gracias al uso de la tecnología digital y los documentos electrónicos, incrementan la seguridad y la transparencia de las operaciones comerciales internacionales. Permite a los gobiernos, mejorar su capacidad recaudatoria, mediante la disminución de los fraudes y evasiones fiscales; mientras que a las empresas les permitirá maximizar sus ingresos. Además los costos asociados a la documentación comercial, sean directos o indirectos, tenderán a reducirse drásticamente, mejorándose así los plazos de entrega a lo largo de la cadena de suministro y garantizando menos espacio para la corrupción. Con el objetivo de generar una información actualizada y de fácil acceso para su consulta, se han desarrollado diferentes sistemas que controlan estadísticas en el Sector externo, entre los que se pueden mencionar:

Sistema Informático MARIA (SIM)

El sistema informático basado en el Arancel Integrado Aduanero de Argentina, más conocido por MARIA, es utilizado para formalizar el registro y darle seguimiento a las declaraciones y operaciones vinculadas a la exportación e importación de mercaderías y al tránsito en frontera. Se ha extendido su alcance a Internet y se ha implementado una versión Web llamada Sistema Informático MARIA Online, la cual constituye el mecanismo de consulta a destinaciones de importaciones y exportaciones del MARIA y está abierto al público en general. Es un sistema desarrollado para funcionar sobre el sistema operativo UNIX y emplea, como gestor de bases de datos relaciones, el sistema ORACLE.

SIDUNEA

El Sistema Aduanero Automatizado (SIDUNEA) es la herramienta informática para el control y administración de la gestión aduanera, desarrollada por UNCTAD (La Conferencia de las Naciones Unidas sobre el Comercio y el Desarrollo), y que actualmente es usada con éxito en más de 80 países. Permite realizar un seguimiento automatizado de las operaciones aduaneras y controlar efectivamente la recaudación de los impuestos aduaneros. Al ser un sistema multidisciplinario, está especializado en cada área del trabajo aduanero para ser la herramienta de trabajo de todos los clientes de la aduana, sean usuarios internos o externos, privados o públicos.

SIDUNEA se puede configurar de acuerdo a las características nacionales de cada régimen aduanero, al arancel nacional y a la legislación de cada país, además de implementar los estándares

Capítulo 1: Fundamentación Teórica del Almacén de Datos

internacionales para procesar los datos de Comercio exterior ya acordados por la Organización Mundial de Aduanas (WCO) y por la Organización Internacional para la Estandarización (ISO).

Al ser cliente-servidor, SIDUNEA es un programa que se instala en su computador para su trabajo aduanal y a la vez es una red de información nacional y mundial. Genera datos estadísticos sobre Comercio exterior y permite el intercambio electrónico de datos entre comerciantes y aduana. (3)

ISIDORA

ISIDORA, cuyas siglas significan “Integración de Sistemas por Internet para el Desarrollo de las Operaciones y Regulaciones Aduaneras”, persigue, como principal objetivo, la simplificación de los procesos aduaneros de Chile. A través de él se pueden realizar todos los trámites vinculados a las operaciones de Comercio exterior, tanto de importación como de exportación, al tiempo de permitir la coordinación de todos los agentes y operadores económicos entre sí y con la Aduana, en forma remota, a través de Internet.

Es un sistema informático basado en la arquitectura cliente/servidor, que integra todos los trámites, procedimientos y consultas que se realizan en el Servicio Nacional de Aduanas de Chile (SNA). (4)

Siex

Es el Sistema Estadístico de Comercio exterior de la Dirección de Impuestos y Aduanas Nacionales de Colombia (Siex) que brinda el servicio de consulta de información estadística de Comercio exterior. Siex ofrece las estadísticas de importaciones, exportaciones y balanza comercial, para facilitar la consulta de la información y brindar a los usuarios la posibilidad de utilizarla oportunamente en sus análisis. La fuente oficial de las cifras de importaciones es la Dirección de Impuestos y Aduanas Nacionales (DIAN) colombiana y de las cifras de exportaciones, el Departamento Administrativo Nacional de Estadística (DANE). (5)

Después de realizar el análisis del funcionamiento de estos sistemas estadísticos no se encontró ninguno que gestione los procesos de toma de decisiones a partir de un AD, que aportara funcionalidades para su utilización en la gestión de series históricas del área del Sector externo en la ONE. Por lo que se propone la creación de un mercado de datos para mejorar el proceso de toma de decisiones en el área del Sector externo.

1.4 Tecnologías de almacenamiento de datos

Las tecnologías de almacenamiento de datos permiten consolidar grandes volúmenes de información y así proporcionar un mejor análisis de la misma. Integran eficientemente tecnologías de bases de datos

con análisis de datos.

1.4.1 Almacenes de Datos

Presentar una vista integrada y consolidada de los datos de una organización, proporcionando conocimiento al proceso de toma de decisiones, fueron las principales motivaciones por las que surgieron los AD.

W. H. Inmon definió un AD como *“una colección de datos clasificada por temas, integrada, variable en el tiempo y no volátil que se utiliza como ayuda al proceso de toma de decisiones por parte de quienes dirigen a una organización”*. (6)

Ralph Kimball define un AD como: *“una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis”*. (7)

Un AD es un conjunto de datos históricos, internos o externos, y descriptivos de un contexto o área de estudio. Están integrados y organizados de tal forma que permiten aplicar eficientemente herramientas para resumir, describir y analizar los datos con el fin de ayudar en la toma de decisiones estratégicas. Proporcionan acceso a datos para análisis complejos, revelación de conocimientos; dan respuesta a las demandas de alto rendimiento de datos e información de una organización.

Un AD de forma más general es: *“un conjunto de tecnologías de soporte a la toma de decisiones, cuyo objeto es que quien trabaja con los conocimientos (ejecutivo, directivo, analista) pueda tomar decisiones de manera más rápida y eficaz”*. (8)

Etapas del diseño de un Almacén de Datos

- Origen: define los orígenes de datos del AD, como los sistemas de Procesamiento de Transacciones en Línea (OLTP) y las fuentes de datos externas (datos sindicados, datos censales).
- Integración: define el mapeo entre los orígenes de datos y el propio AD.
- AD: define la estructura del AD.
- Adaptación: define el mapeo entre el AD y las estructuras empleadas por el cliente.
- Cliente: define las estructuras concretas que son empleadas por los clientes para acceder al AD, como los mercados de datos o aplicaciones OLAP. (9)

Los Almacenes de Datos contienen a menudo grandes cantidades de información que en ocasiones se subdividen en unidades lógicas más pequeñas dependiendo del subsistema de la entidad del que procedan o para el que sea necesario. Estas unidades son los llamados mercados de datos.

1.4.2 Mercados de datos

Un AD al integrar toda la información de la organización requiere tratar con más fuentes de datos, plataformas, requisitos de usuarios, presupuesto, tiempo de desarrollo y personal. Una solución es enfocarse en una unidad de negocios y crear un mercado de datos que satisfaga sus requisitos de información.

Los mercados de datos surgen por la complejidad y elevados recursos asociados a la implementación de los proyectos de AD. Estos son un subconjunto de un AD que soporta los requisitos de un departamento o área de negocio concretos. (10)

Algunas ventajas de la construcción de los mercados de datos son:

- Más simples de implementar que un AD.
- Pequeños conjuntos de datos y, en consecuencia, menor necesidad de recursos.
- Se encuentran rápidamente las necesidades de las unidades de negocio.
- Consultas más rápidas por menor volumen de datos.

Como desventaja se puede decir que, en algunos casos, añaden tiempo al proceso de actualización. En síntesis, son pequeños AD centrados en un tema o un área de negocio específico. En muchos casos, los AD comienzan siendo mercados de datos con el objetivo de minimizar los riesgos para luego ir ampliando su espectro gradualmente.

1.4.3 Procesos básicos de un AD

Los principales procesos que conforman un AD son: extracción, transformación y carga, son generalmente conocidos por sus siglas en inglés ETL (*Extraction, Transformation and Loading*).

Integración de datos

ETL es el proceso que organiza el flujo de los datos entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un AD, reformatearlos, limpiarlos y cargarlos en otra base de datos, mercado de datos o AD.

Los procesos ETL son los componentes más importantes de una infraestructura de Inteligencia de Negocios (BI). Recuperan los datos de todos los sistemas operativos y los pre-elaboran para las

herramientas de análisis y de reportes. La exactitud de la plataforma BI entera depende de los procesos ETL.

Los procesos de ETL constan de múltiples pasos, cuyo objetivo es transferir datos desde las aplicaciones de producción a los sistemas de BI:

- Extracción de los datos desde las aplicaciones y bases de datos de producción.
- Transformación de estos datos, realizar cálculos o análisis sintáctico de cadenas, enriquecerlos con información de búsqueda externa y, además, adaptarlos al formato preciso por el sistema objetivo.
- Carga de los datos resultantes en las diversas aplicaciones de BI: AD históricos generales o AD empresariales, AD históricos individuales, aplicaciones OLAP (Procesamiento analítico en línea) o cubos. (11)

1.5 Inteligencia de Negocio (BI)

La BI es el conjunto de productos y servicios que permiten a los usuarios finales acceder y analizar de manera rápida y sencilla la información. Agrupa 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.

La BI actúa como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva, que no es otra que proporcionar información privilegiada para responder a los problemas de negocio. (12)

1.6 Metodología de Desarrollo

“Las metodologías de desarrollo de software son un conjunto de procedimientos, técnicas y ayudas a la documentación para el desarrollo de productos software”. Para el desarrollo de un AD se requiere de una metodología que guíe el proceso de construcción. (13)

Existen diversas metodologías para la construcción de AD, donde cada una está orientada a la optimización del rendimiento y a la visión de los principales procesos que se deben tener en cuenta para construir un AD flexible y dinámico. Entre las diferentes metodologías se encuentra: SQLBI, DM2, CRISP-DM, Hefesto, Modelo para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio en DATEC, entre otras.

Capítulo 1: Fundamentación Teórica del Almacén de Datos

Para el desarrollo de la presente investigación se utilizó el Modelo para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio en DATEC, la cual surge tomando como base la metodología propuesta por Ralph Kimball en 1992, por los siguientes elementos:

- Crea los conceptos de hechos y dimensiones, lo que indudablemente es muy eficaz en el proceso de la toma de decisiones y proporciona mayor agilidad en el proceso de desarrollo.
- Propone ir construyendo el AD a través de la construcción de los mercados de datos departamentales, lo que constituye una estrategia buena y coincide con la división lógica de las empresas, entidades y organismos.
- Existe abundante documentación sobre la metodología propuesta por Ralph Kimball, la respuesta a todas las dudas y preguntas que puedan surgir se pueden encontrar en la Web, a través de los servicios que brindan el grupo creador de la metodología.
- Es una metodología madura y reconocida por el resto de la comunidad dedicada al tema. Tiene bien definidas las etapas, actividades, artefactos y roles.

Este modelo cuenta con los siguientes cinco grupos de trabajo: grupo de Análisis, grupo de Almacenes, Grupo de ETL, grupo de BI y grupo de Dirección. Cada uno realiza actividades específicas y bien delimitadas según sus responsabilidades dentro de un proyecto.

El proceso de desarrollo de este modelo está compuesto por 7 flujos de trabajos los cuales son:

- Estudio preliminar o planeación.
- Requerimientos.
- Arquitectura y diseño.
- Implementación.
- Prueba.
- Despliegue.
- Soporte y mantenimiento.
- Gestión y administración del proyecto.

Grupos/ Flujos	Estudio Preliminar	Requerimientos	Arquitectura y Diseño	Implementación	Prueba	Despliegue	Soporte y Mantenimiento
Análisis	Responsable	Responsable	Participa	No Participa	Responsable	No Participa	No Participa
Almacén	Participa	No Participa	Responsable	Responsable	Participa	Responsable	Participa
ETL	Participa	No Participa	Responsable	Responsable	Participa	Responsable	Participa
BI	Participa	No Participa	Responsable	Responsable	Participa	Responsable	Participa
Dirección	Responsable	Responsable	Responsable	Responsable	Responsable	Responsable	Participa

Legenda:
 **Responsable**
 **Participa**
No Participa

Figura 1 Relación de los grupos con los flujos de trabajo.

Según la cantidad de actividades y los resultados principales de los flujos, la Figura 1 muestra una gráfica que demuestra la responsabilidad, importancia y participación de cada grupo en los distintos flujos de trabajo.

Como complemento y fortaleciendo la etapa del levantamiento de requisitos; se tomó lo planteado por Leopoldo Zenaido Zepeda Sánchez en su Tesis de doctorado, orientando el trabajo a los Casos de Uso (CU) y logrando estar más alineado con las tendencias y normas de la Universidad.

1.7 Modelo de Base Datos

Un modelo de bases de datos es un conjunto de ideas lógicas utilizadas para representar la estructura de datos y las relaciones entre ellos dentro de la base de datos. (14)

1.7.1 Modelo dimensional

El modelado dimensional es una técnica de diseño lógico utilizada 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. Incluye las dimensiones de análisis y sus atributos, su organización jerárquica, así como los diferentes hechos de negocio que se quieren analizar. Además de ser una alternativa al modelado relacional.

Sus principales ventajas son:

- Enfocado en el negocio y sus actividades.
- Permite búsquedas a gran velocidad.

El modelo dimensional divide los datos en dos grandes tipos: las medidas y las categorías del entorno de estas medidas. Las medidas, que generalmente son numéricas, se almacenan en las tablas de hechos y las categorías o descripciones que son textuales se almacenan en las tablas de dimensiones. (15)

Tabla de Hechos

Las tablas de hechos son las tablas primarias en el modelo dimensional y contiene los valores del negocio. Los hechos son valores numéricos. Contiene dos o más claves foráneas que corresponden a sus respectivas tablas de dimensiones. El nivel de detalle de las transacciones que se almacenen dependerá del nivel de granularidad que se requiera. La granularidad se refiere al nivel de detalle admitido en el mercado de datos.

Tabla de dimensión

En este tipo de modelo, los datos se representan mediante dimensiones. Cada dimensión se define por su llave primaria que sirve para mantener la integridad referencial en la tabla de hechos a la que se acopla. Los atributos de estas tablas sirven de base a las solicitudes que se hacen al AD. Las tablas de dimensiones contienen información jerárquica que permitirán la realización de las agregaciones o las profundizaciones.

Los modelos dimensionales pueden ser representados mediante varios esquemas, como son:

Esquema en estrella

El esquema estrella deriva su nombre del hecho que su diagrama forma una estrella, con puntos radiales desde el centro. El centro de la estrella consiste de una tabla de hechos, y las puntas de la estrella son las tablas de dimensiones relacionadas a esta, a través de sus respectivas claves. Las tablas dimensionales sólo tienen conexión con la tabla hecho y ninguna más. (16)

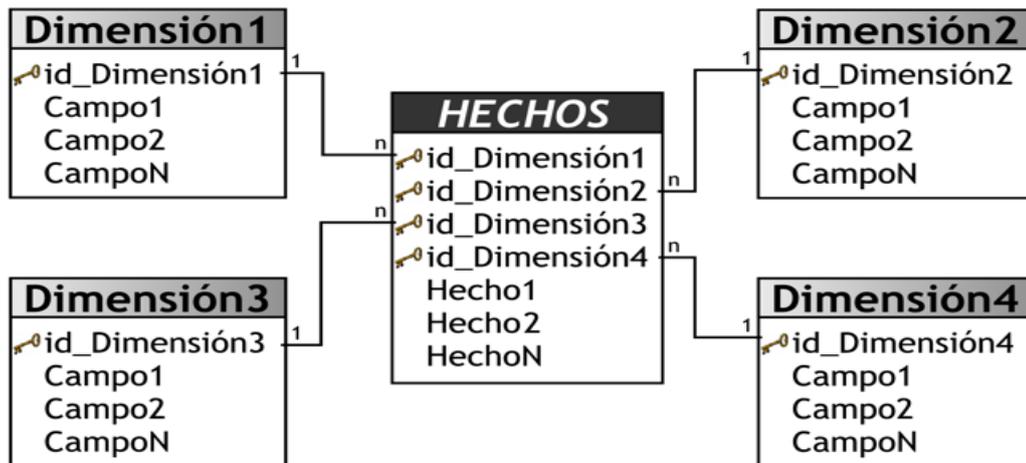


Figura 2 Esquema en estrella

Esquema copo de nieve

La diferencia del esquema copo de nieve o *snowflake* comparado con el esquema estrella, está dada en la estructura de las tablas dimensionales: las tablas dimensionales en el esquema copo de nieve están normalizadas. Cada tabla dimensional contiene sólo el nivel que es clave primaria en la tabla y la llave foránea de su parentesco del nivel más cercano del diagrama. (16)

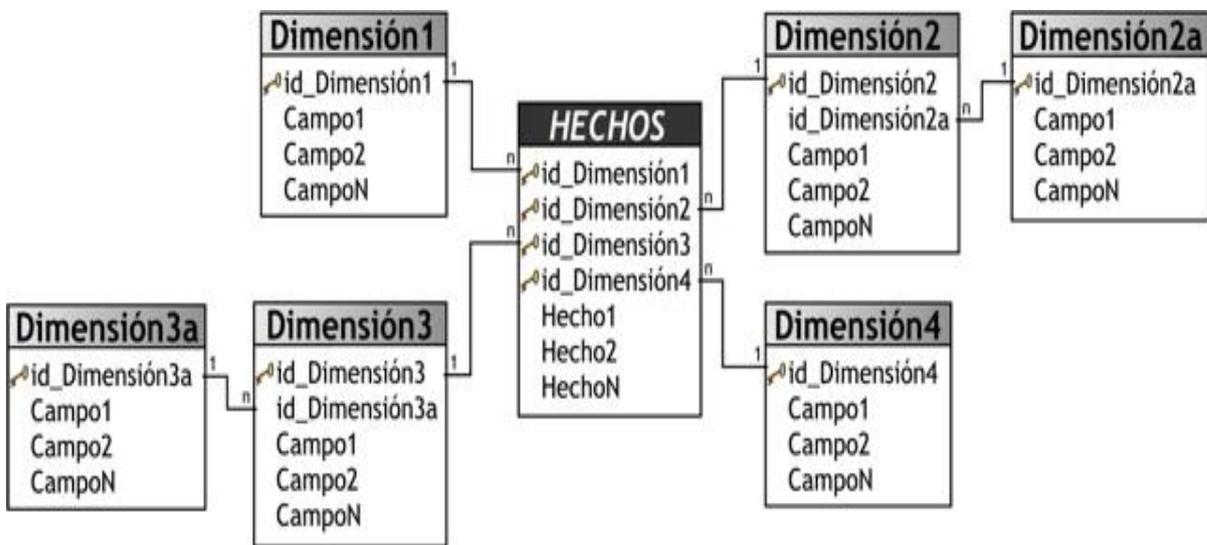


Figura 3 Esquema copo de nieve

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

dimensionales. (16)

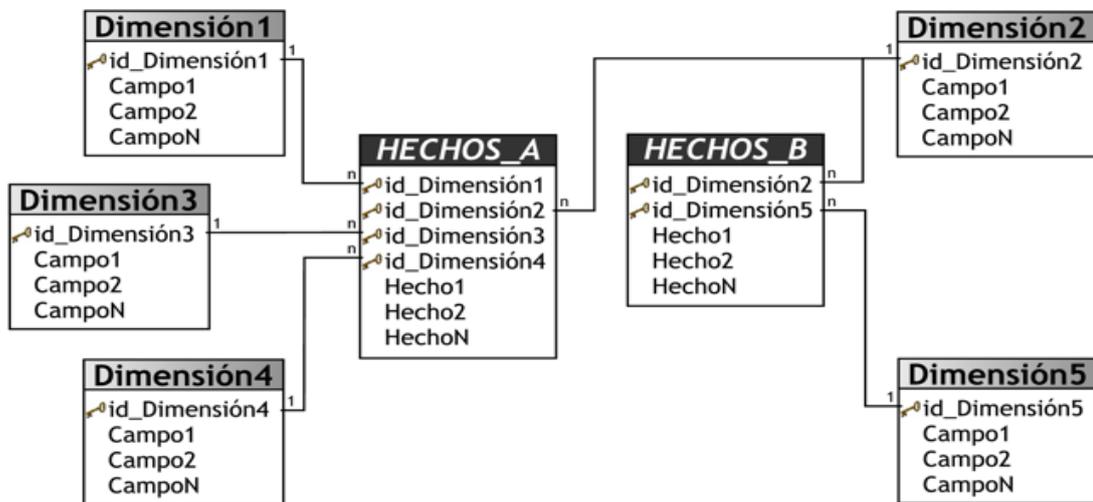


Figura 4 Esquema constelación de hechos

1.8 Modo de Almacenamiento de Datos

Almacenamiento de Datos o *Data Warehousing*, es el proceso que facilita la creación y explotación de un AD. Es una combinación de conceptos y tecnología que cambian significativamente la manera en que es entregada la información. Su objetivo principal es satisfacer los requerimientos de información internos de la empresa para una mejor gestión, con eficiencia y facilidad de acceso. (17)

1.8.1 Proceso OLAP

Los sistemas OLAP (*Online Analytical Processing*) son bases de datos orientadas al procesamiento analítico. OLAP es la síntesis, análisis y consolidación dinámica de grandes volúmenes de datos multidimensionales.

OLAP soporta consultas avanzadas, permitiendo el acceso rápido, coherente e interactivo a las vistas, posibilitando realizar análisis avanzado de datos y por ende facilita la comprensión de datos corporativos. Se caracteriza por permitir realizar un análisis multidimensional (el análisis multidimensional parte de una visión de la información como dimensiones de negocio) de datos corporativos, capaces de soportar el análisis de los usuarios y posibilidades de navegación, seleccionando la información a obtener. Para poder ofrecer estas características, estos sistemas deben:

- Soportar requerimientos complejos de análisis.
- Analizar datos desde diferentes perspectivas.

- Soportar análisis complejos contra un volumen extenso de datos.

En el mundo se han desarrollado principalmente dos tipos diferentes de sistemas OLAP: OLAP multidimensional (MOLAP) y OLAP relacionales (ROLAP). OLAP híbrido (HOLAP) se refiere a las tecnologías que combinan MOLAP y ROLAP. (18)

1.8.2 Tipos de procesos OLAP

Procesamiento Analítico Multidimensional (MOLAP)

La arquitectura MOLAP usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente. Un sistema MOLAP usa una base de datos propietaria multidimensional, en la que la información se almacena multidimensionalmente, para ser visualizada en varias dimensiones de análisis. Utiliza una arquitectura de dos niveles: la bases de datos multidimensionales y el motor analítico. La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato.

La arquitectura MOLAP requiere unos cálculos intensivos de compilación, lee de datos precompilados y tiene capacidades limitadas de crear agregaciones dinámicamente o de hallar ratios⁴ que no se hayan precalculados y almacenados previamente. (19)

Ventajas:

- Excelente rendimiento.
- Puede realizar cálculos complejos.

Desventajas:

- Limitado en la cantidad de datos que puede manejar.
- Requiere una inversión adicional.

Procesamiento Analítico En Línea Relacional (ROLAP)

La arquitectura ROLAP, accede a los datos almacenados en un AD para proporcionar los análisis OLAP. La premisa de los sistemas ROLAP es que las capacidades OLAP se soportan mejor contra las bases de datos relacionales.

⁴ Los Ratios financieros: también llamados razones financieras o indicadores financieros, son coeficientes o razones que proporcionan unidades contables y financieras de medida y comparación, a través de las cuales, la relación (por división) entre sí de dos datos financieros directos, permiten analizar el estado actual o pasado de una organización, en función a niveles óptimos definidos para ella.

El motor ROLAP se integra con niveles de presentación, a través de los cuáles los usuarios realizan los análisis OLAP. Esta arquitectura es capaz de usar datos precalculados si estos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del AD, y soporta técnicas de optimización de accesos para acelerar las consultas. Estas optimizaciones son, entre otras, particionado de los datos a nivel de aplicación, soporte a la desnormalización y *joins* múltiples. (19)

Ventajas:

- Puede manejar grandes cantidades de datos.
- Puede aprovechar las funciones inherentes a la base de datos relacional.

Desventajas:

- El rendimiento puede ser lento.
- Limitada por las funcionalidades de SQL.

Procesamiento Analítico Híbrido en Línea (HOLAP)

Un desarrollo un poco más reciente ha sido la solución OLAP híbrida (HOLAP), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado. (20)

1.9 Sistema Gestor de Base Datos (SGBD)

Un SGBD es un conjunto de programas que se encargan de manejar la creación y todos los accesos a las bases de datos. Se compone de un lenguaje de definición de datos, de un lenguaje de manipulación de datos y de un lenguaje de consulta. El objetivo primordial de un SGBD es proporcionar un entorno que sea a la vez conveniente y eficiente para ser utilizado al extraer y almacenar información de la base de datos. Es una aplicación que permite a los usuarios definir, crear y mantener la base de datos, y proporciona acceso controlado a la misma. (21)

1.9.1 PostgreSQL 8.4

PostgreSQL es un sistema de gestión de bases de datos objeto-relacional, distribuido bajo licencia BSD⁵ y con su código fuente disponible libremente. Es el sistema de gestión de bases de datos de código abierto más potente del mercado. PostgreSQL utiliza un modelo cliente/servidor y usa multiprocesos en vez de multihilos para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto y el sistema continuará funcionando. PostgreSQL funciona muy bien con grandes cantidades de datos y una alta concurrencia de usuarios accediendo a la vez al sistema. (22)

Ventajas:

- Es gratuito.
- Hace más sencillo el análisis de datos.
- Tiene mejor soporte que los proveedores comerciales.
- Diseñado para ambientes de alto volumen.
- Herramientas gráficas de diseño y administración de base de datos.
- Tiene una mayor seguridad.

Desventajas:

- La velocidad de repuesta es relativamente lenta.
- Tienen un límite de 8k por filas que se pueden aumentar hasta 32, lo que trae consigo que disminuya su rendimiento.
- Consume gran cantidad de recursos cuando se realizan muchas transacciones de datos.

1.10 Herramienta CASE

Las herramientas CASE⁶ se pueden definir como un conjunto de programas y ayudas que dan asistencia a los analistas, ingenieros de software y desarrolladores, durante todos los pasos del ciclo de vida de desarrollo de un software. Estas fueron desarrolladas para automatizar esos procesos y facilitar las tareas de coordinación de los eventos que necesitan ser mejorados en el ciclo de desarrollo de software. (23)

⁵ La licencia BSD (Berkeley Software Distribution): permite a los programadores utilizar, modificar y distribuir a terceros el código fuente y el código binario del programa de software original con o sin modificaciones. Los trabajos derivados pueden optar a licencias de código abierto o comercial.

⁶ CASE: Ingeniería de Software Asistida por Computación.

1.10.1 Visual Paradigm 6.4

Visual Paradigm para UML⁷ es una herramienta profesional que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. Soporta diferentes diagramas UML como son: los de componentes, despliegue, secuencia, casos de uso, de clase, actividad y estado. (24)

Sus principales características son:

- Entorno de creación de diagramas para UML 2.0.
- Diseño centrado en casos de uso y enfocado al negocio que genera un software de mayor calidad.
- Capacidades de ingeniería directa e inversa.
- Dispone de múltiples versiones dependiendo de la necesidad.
- Disponibilidad en múltiples plataformas.

1.11 Herramienta utilizada para la integración de datos

Las herramientas ETL juegan un papel fundamental en la creación de los AD que componen la base de todo BI. Estas se colocan en la intersección de diferentes fuentes y sistemas actuando como un embudo que amalgama información y datos heterogéneos convirtiéndola en información soportada en un formato consistente. (25)

1.11.1 Pentaho Data Integration 4.0.1

Pentaho Data Integration también conocido como *kettle* abre, limpia e integra informaciones disponibles en aplicaciones y base de datos separadas y la pone en manos del usuario. Provee una consistencia, una sola versión de todos los recursos de información, que es uno de los más grandes desafíos para las organizaciones. Permite un poderoso proceso de ETL. El uso de *kettle* evita grandes cargas de trabajo manual frecuentemente difícil de mantener y de desplegar. Es una herramienta *open source*, sin costes de licencia y multiplataforma. Incluye cuatro herramientas:

- *Spoon*: para diseñar transformaciones ETL usando el entorno gráfico.
- *PAN*: para ejecutar transformaciones diseñadas con *Spoon*.

⁷ UML (Lenguaje Unificado de Modelado): es un lenguaje gráfico para visualizar, especificar, construir y documentar un sistema.

- *CHEF*: para crear trabajos.
- *Kitchen*: para ejecutar trabajos. (26)

1.12 Herramientas utilizadas para el proceso de BI

Las herramientas de BI son un tipo de software de aplicaciones diseñado para colaborar con la BI en los procesos de las organizaciones. Específicamente se trata de herramientas que asisten el análisis y la presentación de los datos. (27)

1.12.1 Pentaho Schema Workbench 3.2.0 (PSW)

PSW es una herramienta de desarrollo que permite crear modificar y publicar un esquema de Mondrian. Es un programa Java multiplataforma orientado al desarrollador conocedor de la estructura de un esquema de Mondrian. Permite crear todos los objetos que soporta Mondrian: esquema, cubo, dimensiones, métricas. Tiene dos áreas: la zona en la que se muestra la estructura jerárquica del esquema OLAP y la zona de edición de las propiedades de cada elemento.

Además, entre sus características incluye:

- Realizar consultas *MDX*⁸ contra el esquema creado (requiere conocer la sintaxis del lenguaje).
- Consultar la base de datos que sirve de origen para el esquema de Mondrian.
- Publicar directamente el esquema en el servidor de Pentaho. (28)

1.12.2 Mondrian OLAP Server 3.0.4

Mondrian es una de las aplicaciones más importantes de la plataforma Pentaho BI. Es un servidor OLAP *open source* que gestiona comunicación entre una aplicación OLAP (escrita en Java) y la base de datos con los datos fuente. Permite analizar grandes cantidades de datos almacenados en bases de datos SQL de una forma interactiva sin necesidad de escribir las sentencias que serían necesarias para ello en SQL. Permite crear cubos de información para análisis multidimensional. Mondrian se ejecuta sobre un servidor Web. (29)

1.11.3 Apache Tomcat 6.0

Esta herramienta funciona como un contenedor de servlets⁹ desarrollado bajo el proyecto *Jakarta* en la *Apache Software Foundation*. Tomcat implementa las especificaciones de los *servlets* y de *Java*

⁸ MDX (expresiones multidimensionales): es un lenguaje de secuencias de comandos basado en instrucciones que se utiliza para definir, manipular y recuperar datos de objetos multidimensionales.

⁹ Servlets: son programas que se ejecutan en el servidor, realizando la función de una capa intermedia entre una petición proveniente de un navegador Web y las aplicaciones del servidor. Su función principal es proveer páginas web dinámicas y personalizadas.

Server Pages (JSP) de *Sun Microsystems*. Incluye el compilador *Jasper*, que compila JSPs convirtiéndolas en *servlets*. El motor de *servlets* de Tomcat a menudo se presenta en combinación con el servidor web Apache.

Tomcat es usado como servidor web autónomo en entornos con alto nivel de tráfico y alta disponibilidad. Dado que fue escrito en Java, funciona en cualquier sistema operativo que disponga de la máquina virtual Java. Es mantenido y desarrollado por miembros de la Apache Software Foundation y voluntarios independientes. Los usuarios disponen de libre acceso a su código fuente y a su forma binaria en los términos establecidos en la *Apache Software Licence*. (30)

1.11.4 Pentaho BI Server 3.6.0

La aplicación más conocida de la Plataforma de BI es Pentaho BI Server que funciona como una red basada en un sistema de gestión de información, el servidor de integración de aplicaciones y un motor de flujo de trabajo ligero (secuencia de acción). Está diseñado para integrarse fácilmente en cualquier proceso de negocio. (31)

1.12 Conclusiones

- A partir del estudio de los sistemas para el control estadístico del Sector externo que existen en el mundo, se decidió la realización de un mercado de datos para mejorar el proceso de toma de decisiones en el área del Sector externo en la ONE, pues ninguno de dichos sistemas aporta funcionalidades para su utilización en esta área.
- Para guiar el proceso de desarrollo del mercado de datos se escogió el Modelo para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio en DATEC. Se seleccionó como SGBD PostgreSQL 8.4 y como herramienta de modelado Visual Paradigm 6.4. En la integración de datos se utilizará Pentaho Data Integration 4.0.1 para limpiar e integrar los datos. También se seleccionaron como herramientas de inteligencia de negocio Pentaho Schema Workbench 3.2.0 para modelar los cubos OLAP, crear esquemas y dimensiones que soporta Mondrian OLAP Server 3.0.4. Se utilizará como servidor web Apache Tomcat en su versión 6.0 y para visualizar los datos finales el Pentaho BI Server en su versión 3.6.0.

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

2.1 Introducción

En este capítulo se definen los temas de análisis, el negocio, los casos de uso del sistema y aspectos del diseño. Entre los contenidos a los que se hace referencia se encuentran los permisos, los requisitos de información, multidimensional, funcional y no funcional. Se definen las dimensiones del mercado de datos, los hechos y las medidas asociadas a los hechos definidos. Se estructura la matriz bus, el modelo dimensional y se realiza la elaboración tanto del diseño lógico como físico para dicho mercado de datos. Además se diseña el subsistema de integración, el de visualización y los casos de prueba.

2.2 Análisis

2.2.1 Estudio preliminar del negocio

La ONE es el órgano rector de la estadística en Cuba, por la cual es responsable de llevar el control de grandes volúmenes de datos. Esta organización también se encarga de recopilar la información de sus departamentos y oficinas, entre ellos el Sector externo.

Con este fin la ONE cuenta con diferentes modelos que recogen la información según su tipo. Estos modelos se diferencian entre sí por la frecuencia con que se recogen (anual, trimestral, etc.), por los datos que se miden y por la estructura de la plantilla, entre otros aspectos. Los modelos que recogen los indicadores de las series históricas del Sector externo son enviados anualmente a la Asociación Latinoamericana de Integración (ALADI), y es divulgada como información pública.

La ALADI es un organismo intergubernamental que promueve la expansión de la integración de la región, a fin de asegurar su desarrollo económico y social, y tiene como objetivo final el establecimiento de un mercado común. Es el mayor grupo latinoamericano de integración.

La República de Cuba fue aceptada como país miembro de la ALADI el 26 de agosto de 1999, luego de haber cumplido las formalidades pertinentes.

Esta organización cuenta con doce países miembros: Argentina, Bolivia, Brasil, Chile, Colombia, Cuba, Ecuador, México, Paraguay, Perú, Uruguay y Venezuela, representando en conjunto 20 millones de kilómetros cuadrados y más de 500 millones de habitantes. Estos países pueden firmar acuerdos con otros países latinoamericanos o en vías de desarrollo. Esta flexibilidad es un principio fundamental del tratado y busca la convergencia, o sea, el agrupamiento progresivo de los acuerdos parciales para alcanzar una gran área de preferencias y luego un mercado común. Hay unos 35 acuerdos suscritos

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

con países no miembros como Costa Rica, Guatemala, Honduras, Nicaragua, El Salvador, Panamá, Trinidad y Tobago, Guyana, entre otros.

La ALADI recoge una serie de datos estadísticos con periodicidad anual, conocidas como series históricas, de todos los bienes y servicios importados y exportados de sus países miembros, esta información es de acceso permitido a cualquier usuario que lo desee. En nuestro país estos datos son almacenados en documentos *Excel*, lo que obstaculiza el proceso de toma de decisiones.

2.2.2 Tema de análisis

Los temas de análisis son áreas de una entidad o empresa a un alto nivel de información, que poseen objetivos o metas propias y son de gran importancia para el desarrollo de un mercado de datos. El cumplimiento de metas trazadas, como son: la fiabilidad, la utilidad y el éxito de las estructuras, está dirigido hacia la realización completa y satisfactoria del mercado de datos. La propuesta de solución se enfoca en función de los cortes de información que comúnmente se realizan en la ONE. La ONE maneja toda la información relacionada con las exportaciones y las importaciones del área del Sector externo, es por ello que se han identificado a los indicadores para el Sector externo como el principal tema de información o de análisis. Dentro de este y como anteriormente se ha descrito surgen tres tipos de información, estos son:

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

2.2.3 Necesidades de información

Las necesidades de información son las publicaciones de información desde diferentes fuentes, donde el usuario además de consultarla puede realizar su análisis. En la ONE los especialistas del área Sector externo se enfocan en el análisis de las exportaciones e importaciones efectuadas en Cuba. Todo este análisis está destinado a satisfacer necesidades gubernamentales. De estas necesidades de información se derivan los requisitos de información, multidimensionales, funcionales y no funcionales del sistema.

Requisitos de información y multidimensionales

Estos requisitos describen qué información debe almacenar el sistema para satisfacer las necesidades de clientes y usuarios. Identifican los conceptos relevantes sobre los que se debe almacenar información y los datos específicos que son de interés. Constituyen una entrada fundamental para el

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

proceso de BI y para futuros reportes. A continuación se enumeran los requisitos de información identificados:

RI 1. Obtener intercambio total y saldo comercial en un tiempo dado.

RI 2. Obtener intercambio comercial de mercancías por país origen y áreas geográficas en un tiempo dado.

RI 3. Obtener exportaciones según país destino de las mercancías y áreas geográficas en un tiempo dado.

RI 4. Obtener importaciones según país de origen de las mercancías y áreas geográficas en un tiempo dado.

RI 5. Obtener importaciones según país de procedencia de las mercancías y áreas geográficas en un tiempo dado.

RI 6. Obtener importaciones de mercancías agrupadas en grandes categorías económicas en un tiempo dado.

RI 7. Obtener exportaciones de mercancías según secciones de la Clasificación Uniforme para el Comercio Internacional (CUCI) en un tiempo dado.

RI 8. Obtener importaciones de mercancías según secciones de la Clasificación Uniforme para el Comercio Internacional (CUCI) en un tiempo dado.

RI 9. Obtener importaciones según lugar de entrada de las mercancías en un tiempo dado.

RI 10. Obtener importaciones de mercancías según tipo de comercio en un tiempo dado.

Los requisitos multidimensionales muestran las variables tanto de entrada como de salida de cada uno de los requerimientos de información. A continuación se enumeran los requisitos multidimensionales identificados:

Variables de Entrada:

- Tiempo
- Subpartida
- País
- Régimen

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

- Grandes Categorías Económicas (GCE)
- Clasificación Uniforme para el Comercio Internacional (CUCI)
- Unidad de Medida
- Aduana

Variables de Salida:

- Intercambio
- Saldo comercial
- Valor fob¹⁰

Requisitos funcionales

Los requisitos funcionales están orientados a las necesidades de información de los usuarios, son las capacidades o condiciones que el sistema debe cumplir. Deben estar orientados a las necesidades de los usuarios finales. Para el desarrollo del mercado de datos se identificaron los siguientes requisitos funcionales:

- 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.** Realizar la extracción de los clasificadores.
- RF 9.** Realizar la transformación y carga de los clasificadores.
- RF 10.** Configurar vista de análisis OLAP.
- RF 11.** Editar consulta MDX.
- RF 12.** Suprimir filas y columnas vacías.

¹⁰ VALOR FOB: término de comercialización internacional que indica el precio de la mercancía a bordo de la nave o aeronave. Esto no incluye fletes, seguros y otros gastos de manipulación después de embarcada la mercancía.

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

Requisitos no funcionales

Se definen como las propiedades o cualidades que el producto debe tener. Debe pensarse como las características que hacen al producto atractivo, usable, rápido o confiable. Para el desarrollo del mercado de datos se identificaron 12 requisitos no funcionales. Estos están clasificados en usabilidad, fiabilidad, restricciones de diseño, requerimientos para la documentación de usuarios y ayuda del sistema, requerimientos de seguridad, requerimientos de software y requerimientos de hardware. Los requisitos no funcionales identificados se muestran a continuación:

Usabilidad

RNF 1. Cumplir con las pautas de diseño de las interfaces.

El sistema debe tener una interfaz gráfica uniforme a través del mismo incluyendo pantallas, menús y opciones. Las pautas de diseño se realizarán siguiendo los lineamientos de la arquitectura de información.

RNF 2. Mostrar los títulos de los componentes que aparezcan en la interfaz del sistema en el idioma español.

Los títulos de los componentes de la interfaz para interactuar con los usuarios deben ser en idioma español, además deben tener una apariencia uniforme en todo el sistema.

RNF 3. Diseño y navegabilidad de los reportes del AD de manera sencilla.

Un usuario con conocimientos básicos del sistema podrá diseñar un reporte de manera sencilla sin necesidad de ser un experto en las herramientas requeridas para ello. Además, podrá realizar agrupaciones de variables en los reportes de manera dinámica en la misma pantalla de trabajo, agilizando la navegabilidad de los usuarios en un reporte.

Fiabilidad

RNF 4. Garantizar la persistencia de la información.

Para garantizar la persistencia de la información se realizará un respaldo total de los datos del almacén con una frecuencia anual. Toda esta información se almacenará en el área de la dirección de informática en un banco de datos especial. Esta información se almacenará en el edificio correspondiente a la oficina de estadísticas de La Habana y será responsabilidad del grupo de administración de redes de la ONE.

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

Restricciones de diseño

RNF 5. Lograr que los elementos definidos en el almacén tengan una estructura homogénea.

Las estructuras del AD se nombrarán de una manera estándar teniendo en cuenta el tipo de estructura que se maneje. Se definen convenciones de nombrado (Ver documento Codificación y componentes de una solución de DWH) con el objetivo de manejar un vocabulario común en el AD que permita un entendimiento claro y conciso de las estructuras por parte de los desarrolladores que interactúen con el AD.

RNF 6. Utilizar el lenguaje de programación definido durante la investigación.

Para la programación en el AD se utilizará PL/pgSQL como lenguaje dentro del Sistema Gestor de Base de Datos y para la implementación de los procesos de integración de datos el lenguaje JavaScript. Además se utilizará el lenguaje MDX para realizar las consultas.

Requerimientos para la documentación de usuarios y ayuda del sistema

RNF 7. Confeccionar manual de usuario.

El sistema debe estar acompañado de un documento que guiará la ejecución del usuario teniendo en cuenta cada funcionalidad.

Requerimientos de seguridad

RNF 8. Sesión de usuarios.

Los permisos correspondientes al usuario autenticado se activarán una vez que éste se autentique y en caso de cambiar, tendrá acceso sólo a la información que le compete de acuerdo con sus privilegios.

Requerimientos de software

RNF 9. Utilizar el Sistema Gestor de Base de datos definido durante la investigación.

El Gestor de Base de Datos para implementar el AD es el PostgreSQL en su versión 8.4. Como Interfaz de Administración del Gestor de Base de Datos se usará el PgAdmin.

RNF 10. Utilizar la herramienta de integración de datos definida durante la investigación.

Para el proceso de integración se usará la herramienta *Pentaho Data Integrator*.

RNF 11. Utilizar las herramientas para la implementación de la capa de inteligencia de negocios definidas durante la investigación.

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

De la suite de *Pentaho*, se usarán los siguientes componentes:

- *Schema Workbench*: Es la herramienta gráfica que se utiliza para construir el esquema multidimensional que soportará la creación de los reportes multidimensionales.
- *Pentaho BI Server*: Es el servidor de visualización de la suite, que se encarga de visualizar los reportes, tableros de control digital, controlar el acceso a la información y unificar en una solución de inteligencia de negocios el uso de las demás herramientas que componen la suite.
- *Pentaho Administrator Console*: Es la herramienta a usar para administrar el *Pentaho BI Server*, que permite la administración de las conexiones a las bases de datos, tareas programadas así como los roles y usuarios.

Las estaciones de trabajo deben contar además con:

- *Firefox 2.0* o superior.
- *Java Virtual Machine 6.0*.

Requerimientos de hardware

RNF 12. Garantizar características mínimas de hardware.

- 1 GB de RAM o superior.
- Procesador Intel Pentium 4 o superior.
- Almacenamiento en disco para las transformaciones de 1 GB como mínimo.

2.2.4 Casos de Uso del Sistema

Los casos de uso del sistema son una secuencia de acciones que el sistema puede llevar a cabo interactuando con sus actores. A través de estos los usuarios pueden consultar la información deseada. Para una mejor comprensión de los mismos en la Figura 5 se hace una representación gráfica que contiene los casos de uso y la relación que mantiene con los usuarios que interactúan con él. Para la realización del mercado de datos fue necesario identificar los casos de uso de información y funcionales.

Los casos de uso de información se agrupan por el tipo de información que maneja la ONE, específicamente para las series históricas del Sector externo, y en dependencia de las necesidades de información de los usuarios. Estos se muestran a continuación:

- CUS Consultar información de importaciones.

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

- CUS Consultar información de exportaciones.
- CUS Consultar información de intercambio comercial.

Los casos de uso funcionales identificados están basados en la realización de las operaciones de ETL (Extracción, Transformación y Carga) que se le realizarán a los datos fuentes, en la gestión de usuarios para la aplicación y en la realización de las consultas en la base de datos. Estos se muestran a continuación:

- CUS Realizar la extracción de los clasificadores.
- CUS Realizar la transformación y carga de los clasificadores.
- CUS Autenticar usuario.
- CUS Administrar usuario.
- CUS Administrar rol.
- CUS Administrar reporte.
- CUS Modificar reporte.

La especificación de los casos de usos del sistema identificados se encuentra en el expediente de proyecto, en el artefacto “Modelo de Casos de Uso”. La Figura 5 muestra el diseño del diagrama de caso de uso del sistema.

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

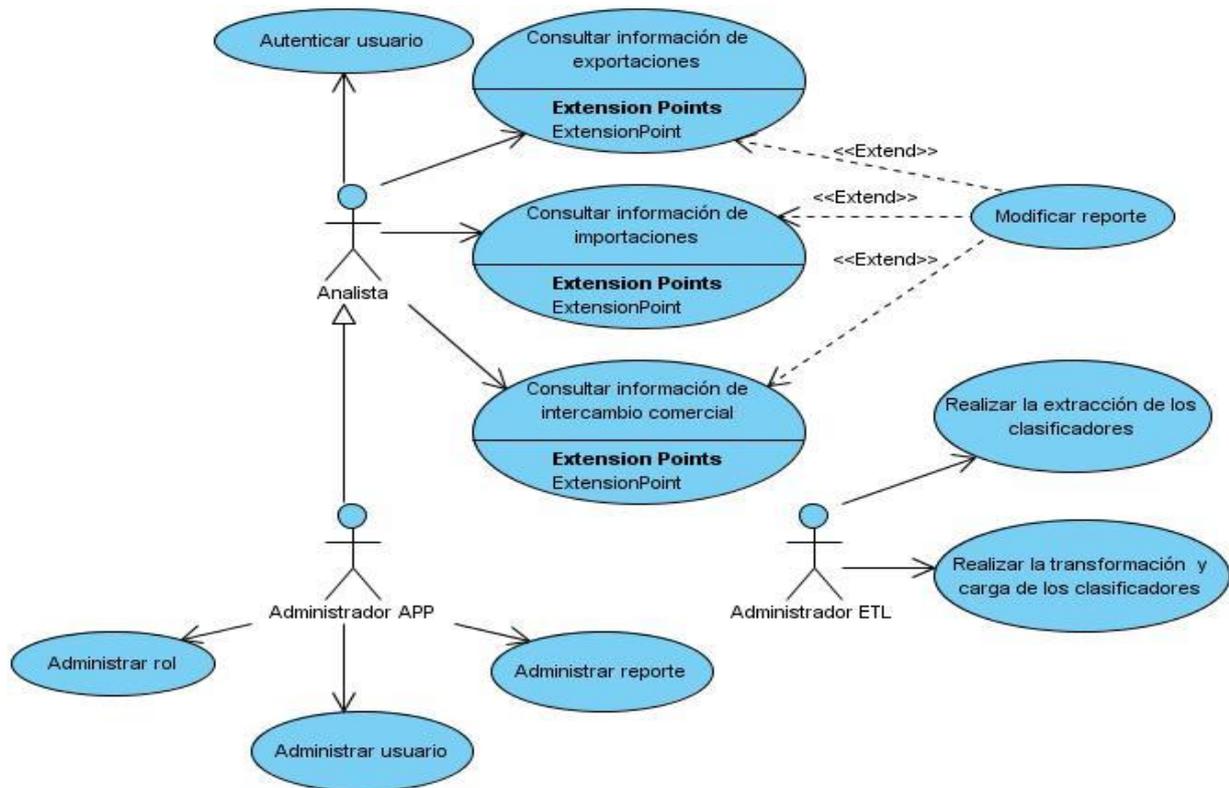


Figura 5 Diagrama de Casos de Uso del Sistema.

2.2.5 Reglas de Negocio

Las reglas del negocio describen las políticas, normas, operaciones, definiciones y restricciones presentes en una organización. En el proceso de almacenamiento de los indicadores del Sector externo se guarda la información de las exportaciones y las importaciones 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 las mismas:

Clasificador Subpartida

RN 1: Describe a 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 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 SACLAP¹¹.

¹¹ SACLAP: 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.

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

Clasificador Grandes Categorías Económicas (GCE)

RN 2: Divide los productos en grandes categorías económicas. Posee un código que puede ser desde uno hasta tres dígitos. EL 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.

RN 4: Intercambio: Es el resultado de la suma de las importaciones y las exportaciones realizadas en un período determinado.

Intercambio = valor_fob (imp) + valor_fob (exp)

RN 5: Saldo comercial: Es el resultado de la resta de las exportaciones y las importaciones realizadas en un período determinado.

Saldo = valor_fob (exp) - valor_fob (imp)

RN 6: En el mercado de datos no pueden existir campos nulos.

RN 7: El código de los atributos en cada una de las dimensiones no pueden tomar valores repetidos.

2.3 Diseño

2.3.1 Identificación de dimensiones, hechos y medidas

Dimensiones Identificadas

Las tablas de dimensiones describen los objetos relevantes para la organización, generalmente son de tipo alfanuméricos. Durante la fase de diseño del mercado de datos se identificaron 8 dimensiones que se describen a continuación:

Tabla 1 Dimensiones identificadas

Dimensión	Descripción
dim_regimen	Esta dimensión contiene los diferentes tipos de comercio (Importación, Exportación, Reimportación, Reexportación, Reimportación de bienes devueltos, Reexportación de bienes devueltos) que se efectúan con los productos.

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

dim_subpartida	<p>Subgrupos por los que están conformadas las mercancías de una partida, está compuesto por 8 dígitos.</p> <p><u>Jerarquía:</u> código SACLAP→ subpartida→partida→ capítulo→sección</p>
dim_pais	<p>Esta dimensión contiene el nombre y descripción del país, nombre del área geográfica a la pertenecen, así como los diferentes códigos ISO (código numérico, alpha 2 y alpha 3) de cada uno de los países de los cuales provienen las importaciones y a los cuales se realizan las exportaciones.</p> <p><u>Jerarquía:</u> área geográfica → país</p>
dim_temporal_mes	<p>Se agrupan una serie de atributos relacionado con la fecha en que se realiza la actividad comercial.</p> <p><u>Jerarquía:</u> año→ semestre→ trimestre→ mes</p>
dim_aduana	<p>Se refleja el nombre y código de la aduana donde se presentan los documentos para su autorización, registro y concesión del levante.</p>
dim_gran_cat_econ	<p>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.</p> <p><u>Jerarquía:</u> categorías básicas→subcategoría→ categoría</p>
dim_cuci	<p>Se especifica el código y la descripción de los productos según la Clasificación Uniforme para el Comercio Internacional (CUCI).</p> <p><u>Jerarquía:</u> subgrupo→ grupo→ capítulo→sección</p>
dim_um	<p>Recoge el código y la descripción de las unidades de medidas por la cual se miden los productos.</p>

Tablas de hechos identificadas

Las tablas de hechos contienen datos sobre las actividades básicas de la organización, estos datos son generalmente de tipo numérico. Durante la fase de diseño del mercado de datos se identificaron 2 hechos que se describen a continuación:

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

Tabla 2 Hechos identificados

Hecho	Descripción
hech_importaciones	Representa la ocurrencia del proceso de importaciones dentro del negocio.
hech_exportaciones	Representa la ocurrencia del proceso de exportaciones dentro del negocio.

Medidas identificadas

Las medidas son los valores, atributos numéricos o indicadores de análisis de un hecho. Durante la fase de diseño del mercado de datos se identificaron 7 medidas que se describen a continuación:

Tabla 3 Medidas identificadas

Medidas	Descripción
peso_bruto	Es la unidad de medida de la mercancía expresada en kilogramos, incluyendo el embalaje.
peso_netto	Es la unidad de medida de la mercancía expresada en kilogramos, sin incluir el embalaje.
cant_unidades	Se reflejará la cantidad de artículos correspondiente a la mercancía declarada luego de su conversión a la unidad de medida comercial.
valor_fob	Se reflejará el valor (CIF ¹² para las importaciones y FOB para las exportaciones) de las mercancías incluyendo todos los elementos de costos.
costo_seguro	Se indicará el importe del seguro en dólar estadounidense que corresponde al producto.
costo_flete	Indica el importe del flete ¹³ en dólar estadounidense que corresponde al producto.

Relación entre los hechos y las medidas.

Tabla 4 Relación entre los hechos y las medidas identificadas

Medidas/Hechos	hech_importaciones	hech_exportaciones
peso_bruto	X	X
peso_netto	X	X

¹² Valor CIF: término de comercialización internacional que indica el precio de la mercancía incluyendo el costo, seguro y fletes.

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.

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

cant_unidades	X	X
valor_fob	X	X
costo_seguro	X	
costo_flete	X	

2.3.2 Matriz Bus o dimensional.

La matriz BUS representa la relación existente entre cada uno de los hechos con sus respectivas dimensiones. A continuación se muestra la matriz BUS:

- D1. dim_temporal_mes
- D2. dim_regimen
- D3. dim_pais
- D4. dim_subpartida
- D5. dim_aduana
- D6. dim_gran_cat_econ
- D7. dim_cuci
- D8. dim_um

Tabla 5 Matriz Bus o dimensional

Hecho\Dimensión	D1	D2	D3	D4	D5	D6	D7	D8
hech_importaciones	X	X	X	X	X	X	X	X
hech_exportaciones	X	X	X	X	X	X	X	X

2.3.3 Modelo de datos

El modelo de datos es un diseño lógico utilizado para modelar bases de datos simples y entendibles al usuario final. La idea fundamental es que este visualice fácilmente la relación que existe entre los distintos componentes del modelo. La Figura 6 muestra el modelo de datos diseñado para el mercado de datos, este se modeló utilizando un esquema constelación de hechos, debido a que existe más de un hecho y estos comparten dimensiones.

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

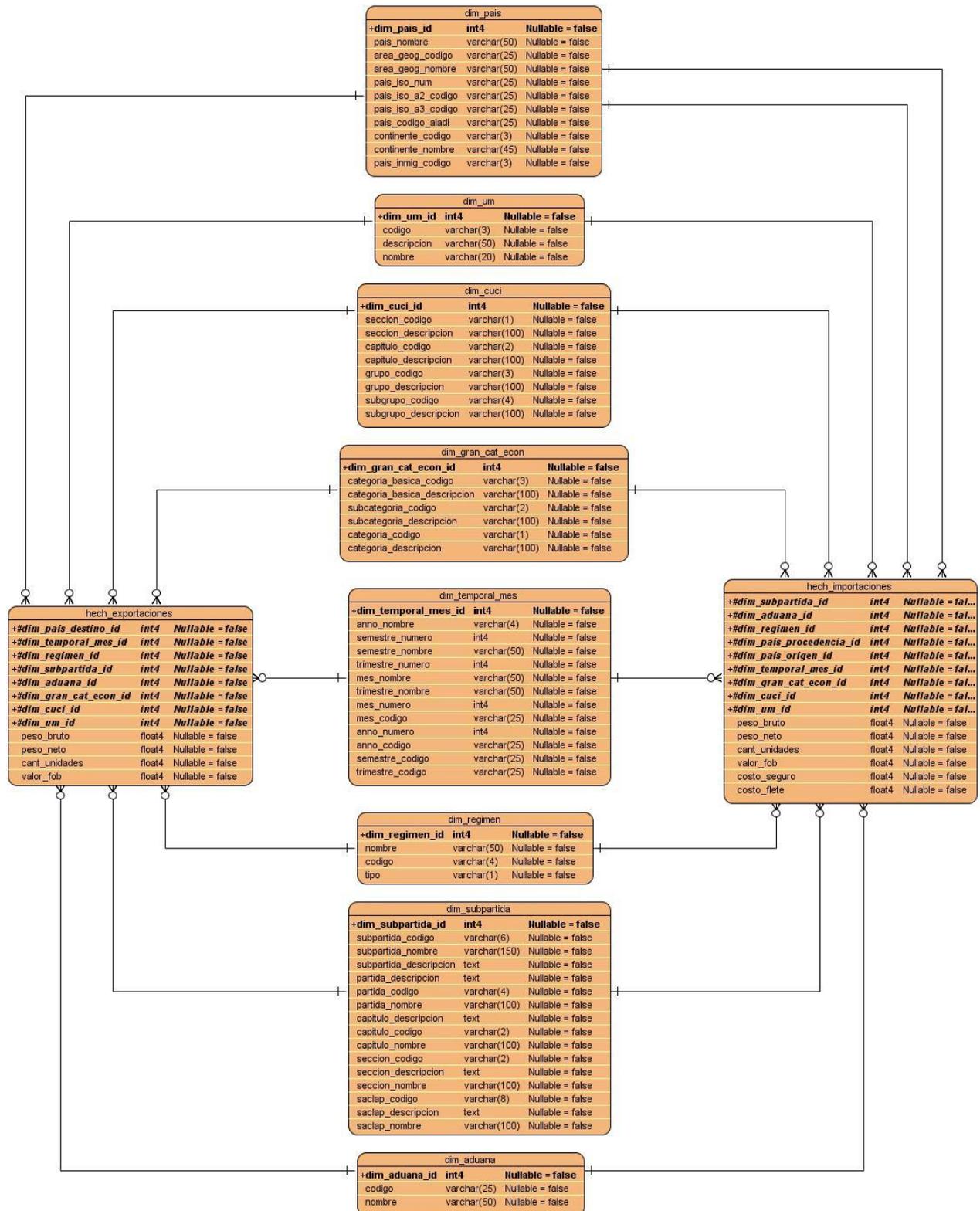


Figura 6 Modelo de datos

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

2.3.4 Diseño del subsistema de integración de datos

Los procesos ETL son los componentes más importantes de una infraestructura de BI. Aunque pueden ser invisibles por los usuarios de la plataforma de BI, los procesos ETL recuperan los datos de todos los sistemas operativos y les pre-elaboran para las herramientas de análisis y de reportes. La exactitud de la plataforma BI entera depende de los procesos ETL.

A la hora de diseñar un proceso ETL es importante realizar previamente un buen diseño y análisis de los datos. Al igual que en todos los procesos relacionados con la informática. El análisis previo, y los sistemas de detección y gestión de errores son dos fases del desarrollo a las que no se les suele dar mucha importancia y si que la tienen. Un buen análisis de los datos permitirá tener datos limpios, ordenados y cualificados. Una buena gestión de los errores permitirá saber rápidamente por que ha fallado el proceso ETL y como poder subsanar el error para poder reanudar el trabajo inmediatamente.

En la Figura 7 se muestra el flujo de una de las transformaciones realizadas en el proceso de ETL para el mercado de datos.

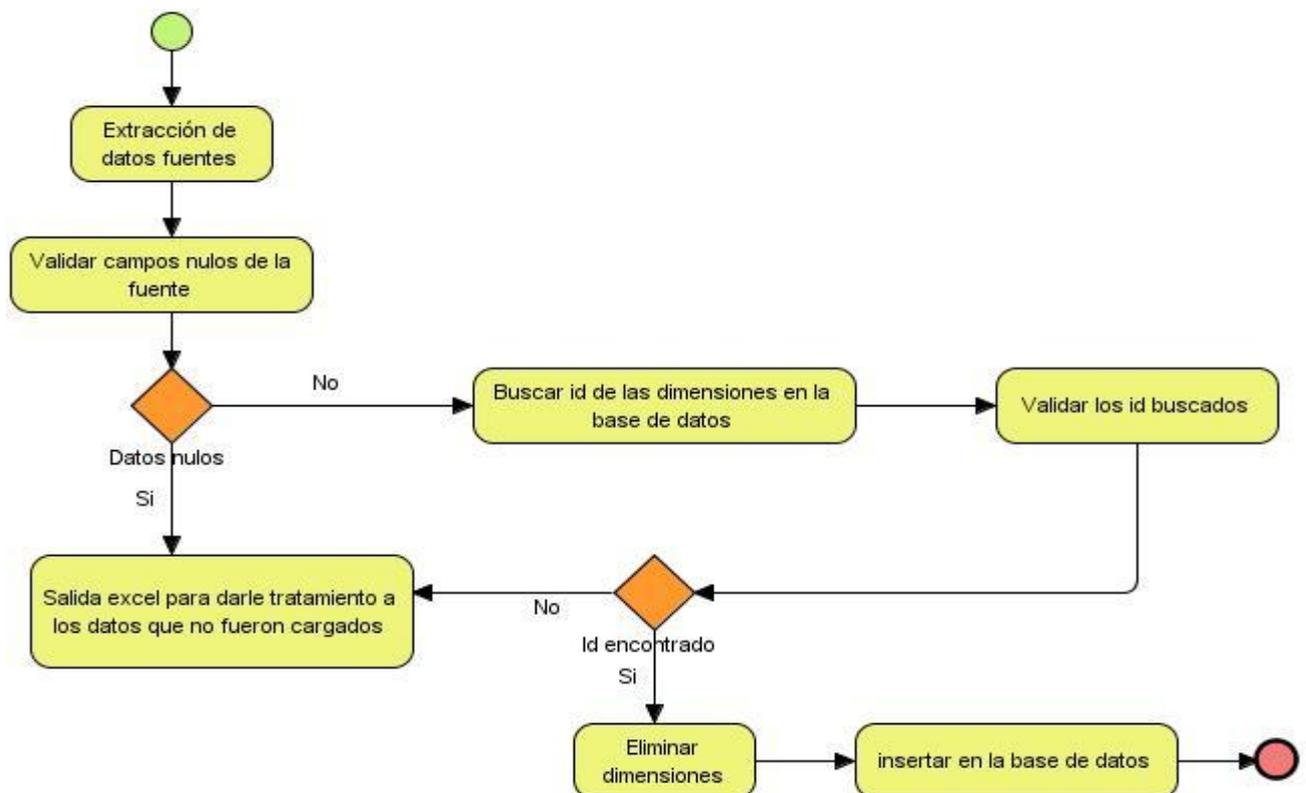


Figura 7 Diseño de transformación

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

2.3.5 Diseño del subsistema de visualización de datos

La BI está evolucionando hacia una nueva manera más potente y visual de integrar, analizar, filtrar y reportar los datos. Usuarios sin conocimientos informáticos pueden manejar grandes tablas de datos con sus imágenes, mapas y enlaces; pueden crear presentaciones dinámicas e interactivas que sorprenderán por su impacto visual y la capacidad de análisis que proporcionan. Esto significa soporte a la toma de decisiones y análisis avanzado de los datos de la empresa, mediante técnicas de reportes, representación gráfica y visualización avanzada.

Las Áreas de Análisis (AA) son definidas usando la información contenida dentro de la capa de usuario final. Con ella se forman grupos temáticos de la información del AD, agrupando las necesidades de los usuarios finales en carpetas y ficheros, que pueden estar o no relacionados, y que para mantener su seguridad se le puede asignar su acceso a un usuario o a un rol determinado.

A continuación se detallan los elementos que componen las estructuras de navegación de la información que será presentada en la capa de visualización del mercado de datos para las series históricas del Sector externo:

- SIGOB: Área de Análisis General (AAG) que agrupa las áreas de análisis de SIGOB.
- A.A Sector Externo: agrupa la información referente a las importaciones, exportaciones e intercambio comercial efectuada en Cuba, concentra reportes estadísticos que apoyan la concepción de medidas y estrategias encaminadas a realizar un mejor control en el proceso de toma de decisiones del área Sector externo.
- L.T Importación: libro de trabajo contenido dentro del área de análisis Sector externo, contiene 6 reportes que permiten realizar un análisis de datos correspondiente a las importaciones efectuadas en Cuba.
- L.T Exportación: libro de trabajo contenido dentro del área de análisis Sector externo, contiene 2 reportes que permiten realizar un análisis de datos correspondiente a las exportaciones efectuadas en Cuba.
- L.T Intercambio: libro de trabajo contenido dentro del área de análisis Sector externo, contiene 2 reportes que permiten realizar un análisis de datos correspondiente al intercambio comercial de bienes y servicios efectuados en Cuba.
- Reportes correspondientes a cada Libro de Trabajo.

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

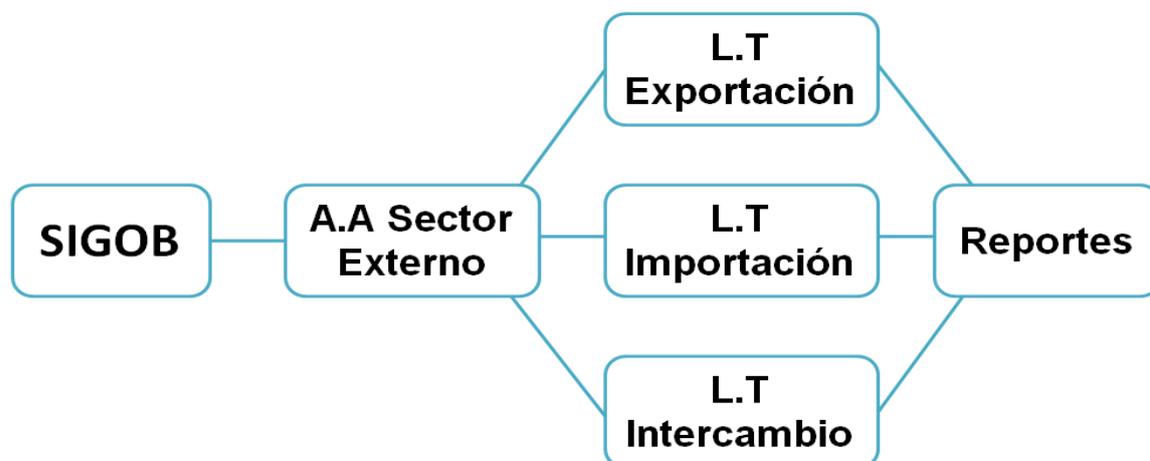


Figura 8 Diseño del subsistema de visualización de datos

2.3.6 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 en el mercado de datos para las series históricas del Sector externo 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

Para la interacción de los usuarios con la base de datos se definieron los siguientes usuarios:

Tabla 6 Usuarios en la base de datos

Actor	Descripción
Administrador ETL	Realiza los procesos de Extracción, Transformación y Carga (ETL).
Administrador APP	Tiene total acceso a la base de datos del sistema.
Analista	Tiene acceso de lectura a la base de datos del sistema.

Seguridad en la aplicación.

Debido al creciente incremento que se evidencia actualmente de las aplicaciones desplegadas en el Servidor de BI 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:

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

Tabla 7 Roles y permisos en la aplicación

Roles	Permisos
Administrador APP	<ul style="list-style-type: none">• Gestiona “Sistema de Información de Gobierno”.• Tiene acceso total a todas las Áreas de Análisis (AA), libros de trabajos y reportes.• Visualiza los reportes.
Analista	<ul style="list-style-type: none">• Tiene acceso de solo lectura al AA Sector externo.• Visualiza los reportes.

Además de los roles antes definidos para interactuar con la aplicación, la misma trae su propia seguridad interna, la cual está basada en la infraestructura proporcionada por el sistema de seguridad Acegi. El mismo en aras de presentar una mejor segmentación de la seguridad y para lograr un mayor rendimiento, se divide en cuatro áreas fundamentales, estas son:

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 AD 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.

Autorización a objetos del dominio: en el sistema los únicos objetos del dominio protegidos por la plataforma son los objetos de repositorio otorgados al usuario autenticado. Es responsabilidad de los objetos del dominio autorizar las operaciones solicitadas por este. (32)

2.4 Política de respaldo y recuperación

Las políticas de respaldo y recuperación que se emplean en el mercado de datos se rigen por los 3 puntos que a continuación se muestran:

Capítulo 2: Análisis y diseño del Mercado de datos para las series históricas del Sector externo

- **Periodicidad de salvos del sistema:** estas se harán anuales a toda la información contenida en el mercado de datos, verificando siempre la existencia de una copia de toda la información almacenada.
- **Tablas involucradas:** las tablas que se involucran en la realización son: hech_importaciones y hech_exportaciones.
- **Salvos existentes:** aunque no existen salvos en esta área se prevé la realización de reemplazos de *backups* cada un año. Además de esto se realizan dos salvos obligatorias, una en el mes de julio y otra en diciembre, de estas una copia se almacena en la ONE y la otra es guardada en repositorio externo.

2.5 Conclusiones

- Se identificaron 10 requisitos de información, 12 requisitos funcionales y 12 requisitos no funcionales. Para describir las políticas y normas presentes en una organización se conformaron 7 reglas del negocio.
- Se definieron 8 dimensiones, 2 tablas de hechos y 6 medidas, los cuales integraron el diseño del modelo lógico que estructura el mercado de datos. También quedó diseñado el subsistema de integración y el de visualización.

Capítulo 3: Implementación del Mercado de datos para las series históricas del Sector externo

Capítulo 3: Implementación del Mercado de datos para las series históricas del Sector externo

3.1 Introducción

Los temas que aborda este capítulo están relacionados con todos los elementos de extracción, transformación y carga de los datos. Se implementa el modelo de datos, los flujos de transformación, los cubos multidimensionales y los reportes candidatos. Además incluye la configuración de la seguridad de los usuarios.

3.2 Implementación de la Base de Datos

Modelo de Datos Físico

Los modelos físicos proporcionan conceptos que describen los detalles de cómo se almacenan los datos en el ordenador: el formato de los registros, la estructura de los ficheros (desordenados, ordenados) y los métodos de acceso utilizados (índices). Los conceptos de los modelos físicos están dirigidos al personal informático, no a los usuarios finales. Estos modelos proporcionan las características reales de almacenamiento físico de los datos; su posición y formato, la configuración de los registros físicos, las relaciones entre registros de datos y los métodos de acceso. (33)

Esquemas

Los esquemas en la base de datos es una forma de organizar los datos en la misma, estos pueden tener funciones, operadores y tipos de datos. Los usuarios pueden acceder a ellos siempre que tengan los permisos necesarios. En la implementación del modelo de datos se definieron dos esquemas:

dimensiones: se encuentran las dimensiones que son comunes en el almacén central de la ONE.

mart_comercio_ext_series: se encuentran los hechos, los que son específicos del mercado.

Tablas

La base de datos de la solución propuesta cuenta con diez tablas en total, ocho de dimensiones y dos hechos distribuidas en los esquemas como se muestra en la siguiente tabla.

Tabla 8 Esquemas y tablas del mercado de datos

Esquema	Tablas
dimensiones	dim_regimen
dimensiones	dim_subpartida

Capítulo 3: Implementación del Mercado de datos para las series históricas del Sector externo

dimensiones	dim_pais
dimensiones	dim_temporal_mes
dimensiones	dim_aduana
dimensiones	dim_gran_cat_econ
dimensiones	dim_cuci
dimensiones	dim_um
mart_comercio_ext_series	hech_importaciones
mart_comercio_ext_series	hech_exportaciones

En la Figura 3 se muestra como queda conformada la base de datos.

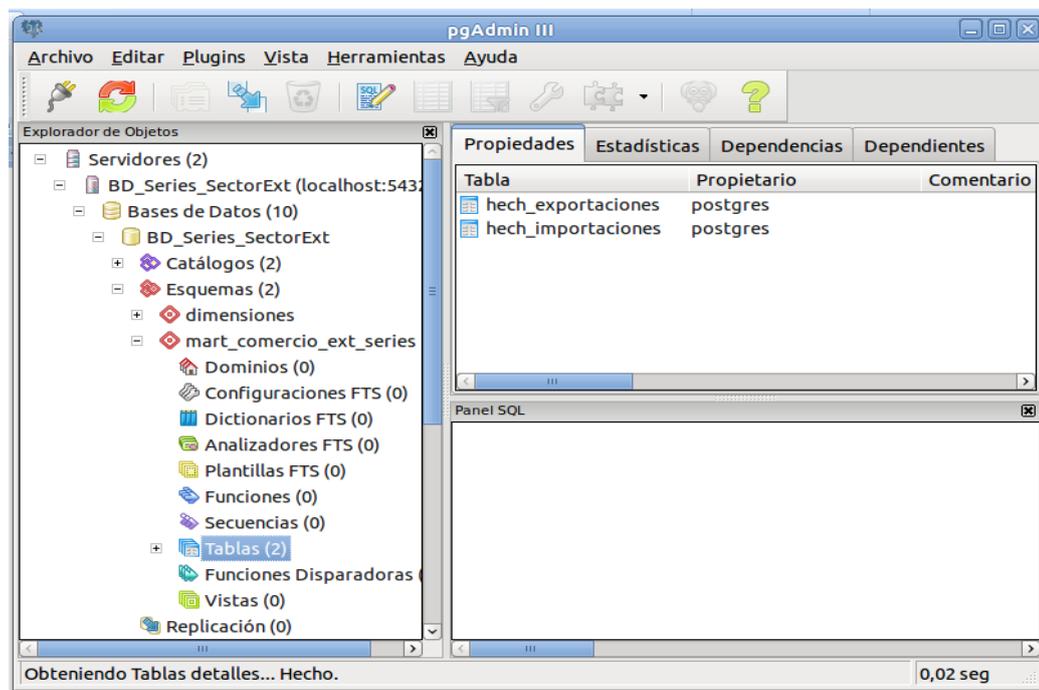


Figura 9 Estructura física de la BD

3.3 Implementación de los procesos de ETL

En la implementación del subsistema de integración es donde se desarrollan las tareas relacionadas con el proceso de ETL. Este proceso permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos, limpiarlos, y cargarlos en otra base de datos, mercado de datos, o AD para analizar, o en otro sistema operacional para apoyar un proceso de negocio. El objetivo principal de

Capítulo 3: Implementación del Mercado de datos para las series históricas del Sector externo

ETL es transferir datos desde las aplicaciones de producción a los sistemas de BI, este proceso consta de tres etapas:

- Extracción de datos
- Transformación de datos
- Carga de datos

3.3.1 Arquitectura del subsistema 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 *Excel*. 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 prepararan 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 AD.

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

La arquitectura del proceso queda representada de la siguiente manera:



Figura 10 Arquitectura del subsistema de integración

3.3.2 Implementación de los flujos de transformación

Extracción de los datos

El primer proceso de ETL consiste en extraer los datos desde los sistemas de origen, que pueden ser provenientes de diferentes fuentes. Las fuentes establecen desde dónde se extraerán los datos para

Capítulo 3: Implementación del Mercado de datos para las series históricas del Sector externo

analizarlos. Cada sistema puede usar formatos distintos o una organización diferente de los datos. El formato de la fuente de esta investigación son los clasificadores que recogen toda la información que se procesa en el área del Sector externo en la ONE. Este proceso convierte los datos a un formato preparado para iniciar el proceso de transformación.

Implementación de las transformaciones

Una transformación es una unidad de ejecución compuesta por pasos enlazados por medio de saltos. Estos pasos y saltos conforman caminos por donde los datos fluyen: entran, se transforman y salen. Por eso se dice que una transformación está orientada al flujo de datos. Los pasos son los elementos más pequeños dentro de las transformaciones. Los saltos son el medio por donde fluye la información entre los diferentes pasos.

La fase de transformación aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. La salida final del flujo son las tabla de la base de datos donde se van a cargar los datos del mercado. Después de realizada la extracción de los datos el sistema se encuentra listo para la etapa de transformación. Terminado este proceso, los datos podrán ser cargados.

Carga de datos

La carga es el último subproceso dentro de los procesos de ETL, el cual consiste en cargar todos los datos que ya han sido transformados satisfactoriamente. La fase de carga es el momento en el cual los datos de la fase anterior (transformación) son cargados en el sistema de destino.

Para este proceso se tiene que realizar primero la carga de las tablas que no tienen llaves foráneas, o sea, que no dependen de la información de otras tablas. Con estas características, para el mercado de datos se encuentran las tablas del esquema dimensiones, y luego se cargarán las tablas que tienen alguna dependencia de otras, que son las correspondientes a los hechos que se encuentran en el esquema `mart_comercio_ext_series`.

La Figura 11 describe el proceso explicado anteriormente, en este caso es la transformación para cargar el hecho exportaciones.

Capítulo 3: Implementación del Mercado de datos para las series históricas del Sector externo

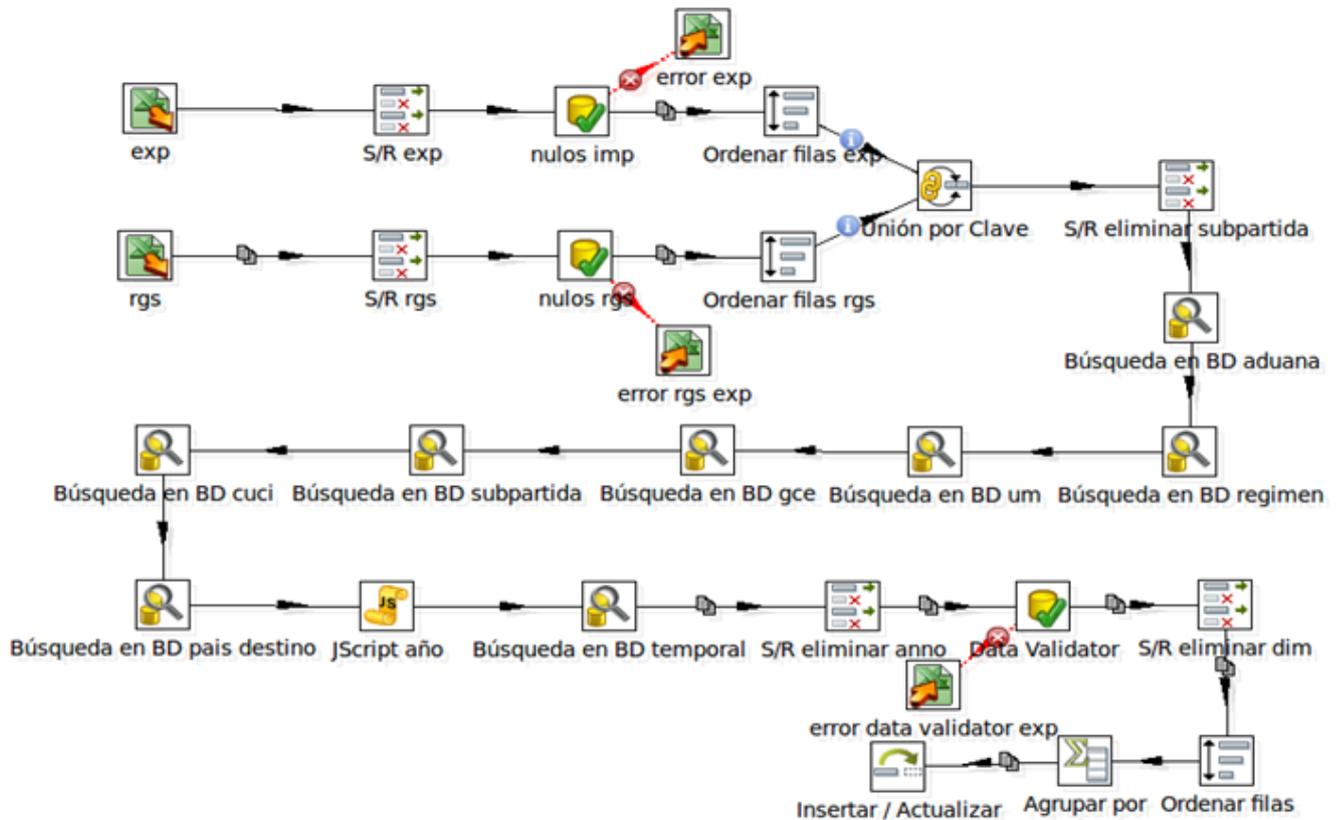


Figura 11 Carga del hecho exportaciones

Implementación de los Trabajos

Un Trabajo (*Job*) es un componente compuesto por entradas de trabajo enlazadas por medio de saltos. En los trabajos se utilizan pasos específicos que son diferentes a los disponibles en las transformaciones. Permite ejecutar una o varias transformaciones de las diseñadas siguiendo una secuencia de ejecución. Los trabajos estarían en un nivel superior a las transformaciones.

Después que se les ha realizado las transformaciones a los datos se hace necesario organizar la carga de las tablas estableciendo un orden. Para el mercado de datos se carga primero las transformaciones de las tablas de dimensiones, y luego las tablas que tienen alguna dependencia de otras, que son las correspondientes a los hechos. La carga en este orden es importante porque se evita cargar llaves nulas que podrían pertenecer a otras tablas que no han sido cargadas. En la Figura 12 se muestra el trabajo realizado para cargar los datos del mercado.

Capítulo 3: Implementación del Mercado de datos para las series históricas del Sector externo

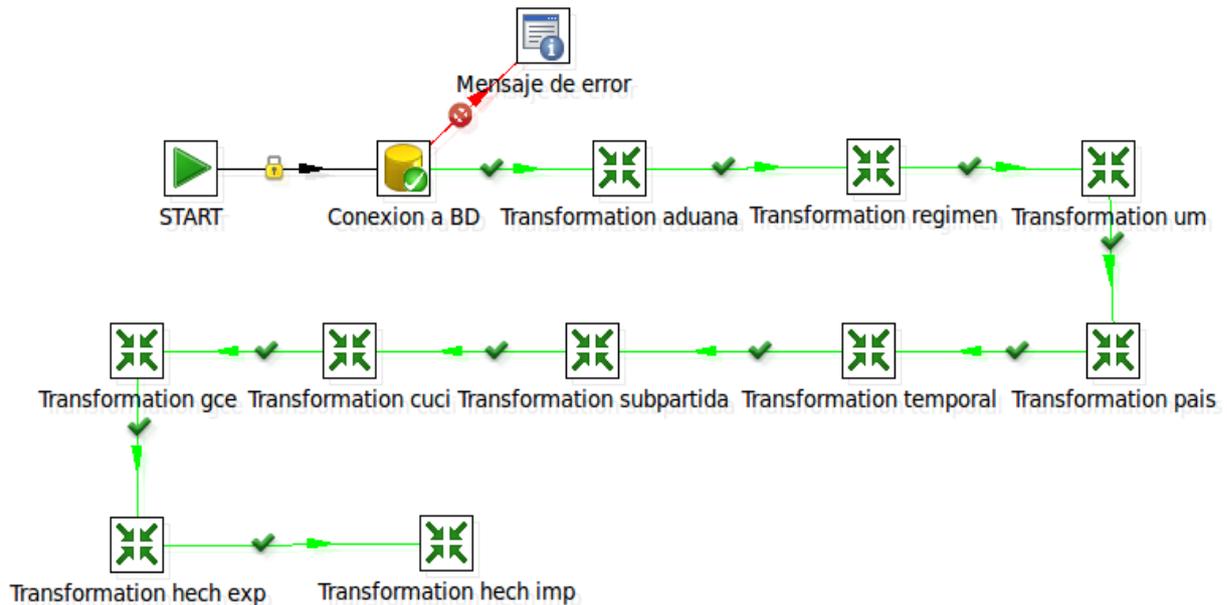


Figura 12 Trabajo

3.4 Implementación del subsistema de visualización de datos (BI)

Las soluciones de BI permiten implementar proyectos para agilizar y mejorar la calidad en la gestión de la información. Se debe concebir como una solución que permite mapear y sistematizar la información generada en los distintos procesos de la empresa, por los diferentes departamentos de la misma. Las cuales, a su vez, brindan la información necesaria para tomar decisiones adecuadas que se orienten a reducir costos operativos y financieros y optimizar los sistemas de inventarios.

3.4.1 Implementación de los cubos OLAP

Un cubo OLAP proporciona sólo los resultados resumidos de los informes, en vez de una gran cantidad de registros individuales. Para la implementación del subsistema de visualización, es necesaria la creación de los cubos multidimensionales en los cuales se definen las dimensiones, los niveles de jerarquía de las dimensiones y las medidas, esto se realiza a través de la herramienta *Pentaho Schema Workbench*, la misma genera un fichero *XML*, el cual contendrá todos los cubos OLAP.

En la realización del diseño se modelaron 3 cubos, uno de los cubos es llamado importación, otro exportación, y el último intercambio. A continuación se muestra una imagen con el diseño de estos cubos.

Capítulo 3: Implementación del Mercado de datos para las series históricas del Sector externo

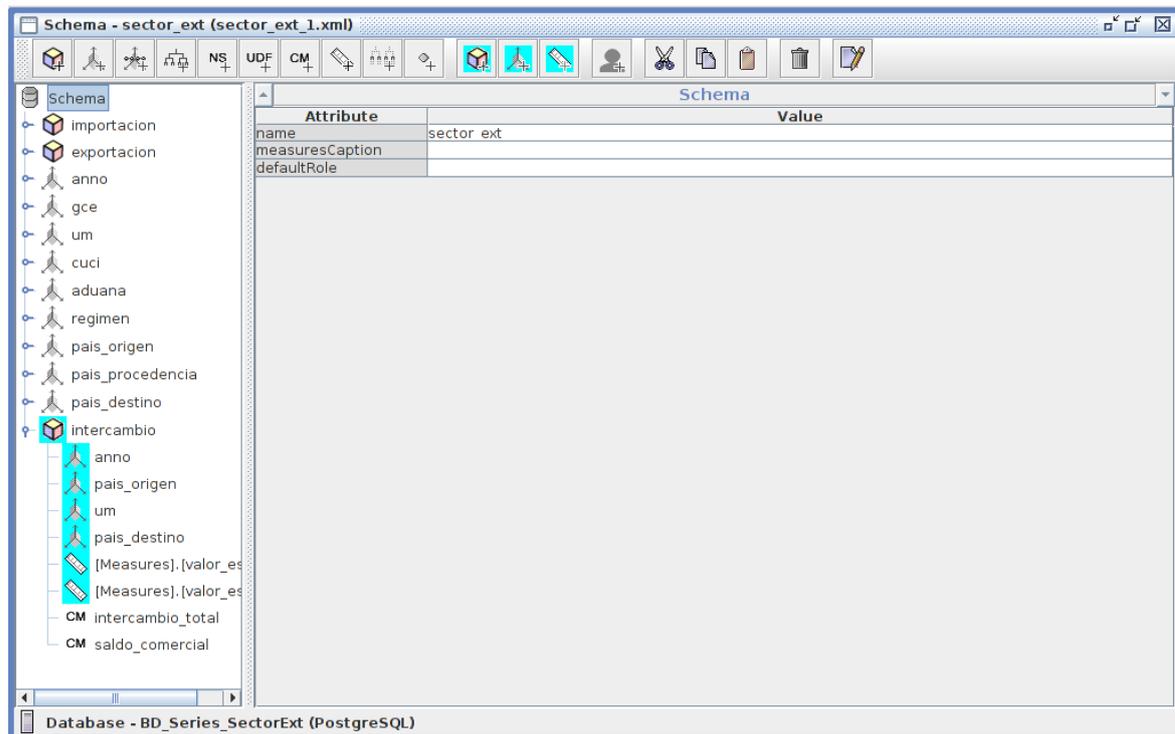


Figura 13 Implementación de los cubos OLAP

3.4.2 Implementación de los reportes candidatos

La implementación de los reportes se realizó utilizando consultas *MDX*, la complejidad de cada una de estas consultas están en dependencia de la complejidad de cada una de los reportes que muestra la aplicación.

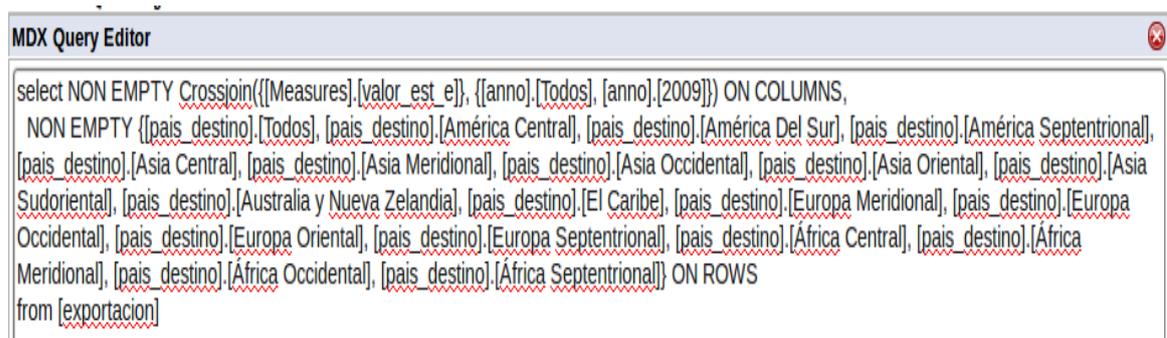


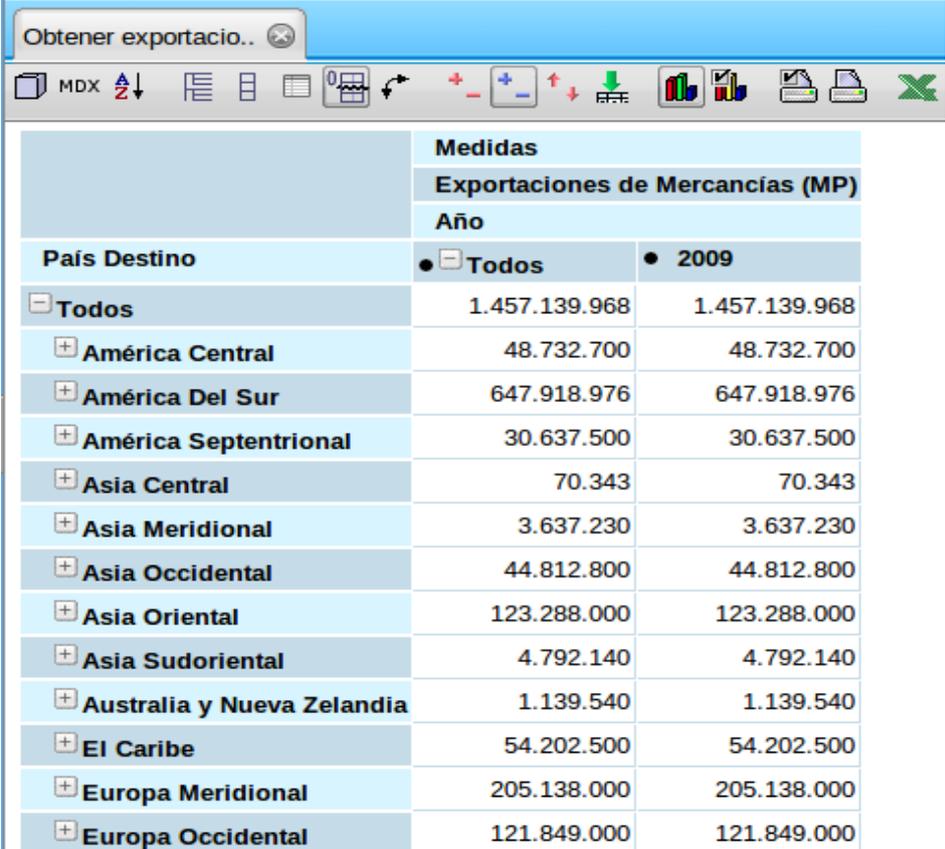
Figura 14 Consulta MDX Obtener exportaciones según país destino de las mercancías y áreas geográficas en un tiempo dado

3.4.3 Arquitectura de información

El mercado de datos está compuesto por 10 reportes en total, los cuales fueron distribuidos dentro de los 3 libros de trabajo descritos en el epígrafe 2.3.5. Para la realización de estos se utilizaron como herramientas: el Mondrian para la visualización de los reportes, utilizando el Tomcat como servidor

Capítulo 3: Implementación del Mercado de datos para las series históricas del Sector externo

web y el Pentaho BI Server donde queda estructurado finalmente el mapa de navegación junto con todos los reportes, la Figura 15 muestra uno de estos reportes.



The screenshot shows a web browser window with a report titled "Obtener exportacio..". The report displays a table with the following data:

País Destino	Medidas	
	Exportaciones de Mercancías (MP)	
	Año	
	● <input type="checkbox"/> Todos	● <input checked="" type="checkbox"/> 2009
<input type="checkbox"/> Todos	1.457.139.968	1.457.139.968
<input checked="" type="checkbox"/> América Central	48.732.700	48.732.700
<input checked="" type="checkbox"/> América Del Sur	647.918.976	647.918.976
<input checked="" type="checkbox"/> América Septentrional	30.637.500	30.637.500
<input checked="" type="checkbox"/> Asia Central	70.343	70.343
<input checked="" type="checkbox"/> Asia Meridional	3.637.230	3.637.230
<input checked="" type="checkbox"/> Asia Occidental	44.812.800	44.812.800
<input checked="" type="checkbox"/> Asia Oriental	123.288.000	123.288.000
<input checked="" type="checkbox"/> Asia Sudoriental	4.792.140	4.792.140
<input checked="" type="checkbox"/> Australia y Nueva Zelandia	1.139.540	1.139.540
<input checked="" type="checkbox"/> El Caribe	54.202.500	54.202.500
<input checked="" type="checkbox"/> Europa Meridional	205.138.000	205.138.000
<input checked="" type="checkbox"/> Europa Occidental	121.849.000	121.849.000

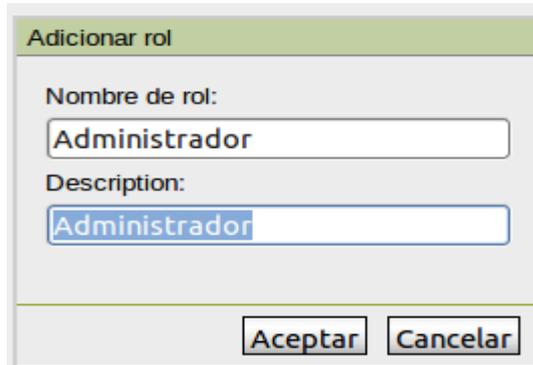
Figura 15 Reporte Obtener exportaciones según país destino de las mercancías y áreas geográficas en un tiempo dado

3.4.4 Configurar la seguridad de los usuarios

Durante la implementación del subsistema de visualización del mercado de datos se crearon 2 usuarios y roles, los cuales tienen diferentes permisos de acceso a la información, proporcionando una mayor seguridad al sistema.

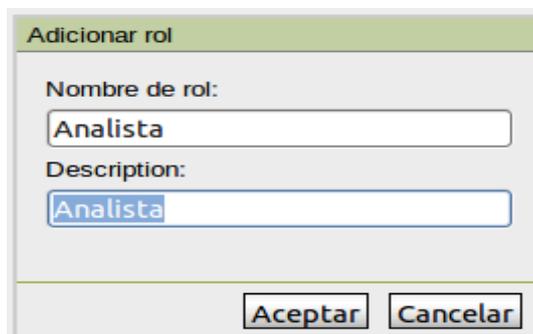
- Rol de Administrador: tiene todos los permisos de la aplicación y se le asigna al usuario administrador del sistema.
- Rol de Analista: tiene permiso de solo lectura y se le asigna al usuario analista del sistema.

Capítulo 3: Implementación del Mercado de datos para las series históricas del Sector externo



The screenshot shows a dialog box titled "Adicionar rol" with a light green header. It contains two text input fields. The first field, labeled "Nombre de rol:", contains the text "Administrador". The second field, labeled "Description:", also contains the text "Administrador". At the bottom of the dialog, there are two buttons: "Aceptar" and "Cancelar".

Figura 16 Adicionar rol administrador



The screenshot shows a dialog box titled "Adicionar rol" with a light green header. It contains two text input fields. The first field, labeled "Nombre de rol:", contains the text "Analista". The second field, labeled "Description:", also contains the text "Analista". At the bottom of the dialog, there are two buttons: "Aceptar" and "Cancelar".

Figura 17 Adicionar rol analista

3.5 Conclusiones

- Se implementó el subsistema de integración donde se realizaron las transformaciones correspondientes para la carga de los datos. A través de la carga de los clasificadores al mercado de datos se logró que la información esté lista para ser procesada y visualizada.
- Quedó definida la estructura de los datos dividida en dos esquemas: esquema dimensiones y esquema mart_comercio_ext_series.
- Se diseñaron los cubos OLAP en dependencia del tipo de información que se maneje en cada uno de estos. A partir del diseño de los cubos se implementó el subsistema de visualización en el que fueron realizados 10 reportes.

Capítulo 4: Validación del Mercado de datos para las series históricas del Sector externo

Capítulo 4: Validación del Mercado de datos para las series históricas del Sector externo

4.1 Introducción

En este capítulo se hace referencia a la validación de la solución. Se describen pruebas existentes para la validación de una aplicación. Además se detalla el proceso de diseño y aplicación de los casos de pruebas, y queda diseñada la lista de chequeo. También se hace referencia a la validación por parte del cliente mediante las pruebas de aceptación.

4.2 Validación del sistema

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 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.

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

4.2.1 Pruebas aplicadas al Mercado de datos para las series históricas del Sector externo

Lista de chequeo

La lista de chequeo es un documento que tiene un conjunto de parámetros a medir sobre un aspecto determinado, dígame 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 de 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.

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.

Capítulo 4: Validación del Mercado de datos para las series históricas del Sector externo

- **Elementos definidos por la metodología:** abarca todos los indicadores a evaluar según la metodología.
- **Semántica del documento:** contempla todos los indicadores a evaluar respecto a la ortografía, redacción y demás.

Elementos que forman parte de la estructura de la lista de chequeo:

- **Peso:** define si el indicador a evaluar es crítico o no.
- **Indicadores a evaluar:** son los indicadores a evaluar en las secciones **Estructura del documento, Semántica del documento e Indicadores definidos por la etapa.**
- **Evaluación (Eval):** es la forma de evaluar el indicador en cuestión. Se evalúa de 1 en caso de que exista alguna dificultad sobre el indicador y 0 en caso de que el indicador revisado no presente problemas.
- **N.P (No Procede):** se usa para especificar que el indicador no es necesario evaluarlo en ese caso.
- **Cantidad de elementos afectados:** especifica la cantidad de errores encontrados sobre el mismo indicador.
- **Comentario:** especifica los señalamientos o sugerencias que quiera incluir la persona que aplica la lista de chequeo. Pueden o no existir señalamientos o sugerencias.

La lista de cheque diseñada para el mercado puede verse en el Anexo 1.

En la Figura 18 se muestra un gráfico de barras que describe el comportamiento de los indicadores definidos para la lista de chequeo elaborada. De forma general se identificaron 14 indicadores, de ellos 8 críticos y luego de aplicada la lista de chequeo se generaron 4 no conformidades.

Capítulo 4: Validación del Mercado de datos para las series históricas del Sector externo

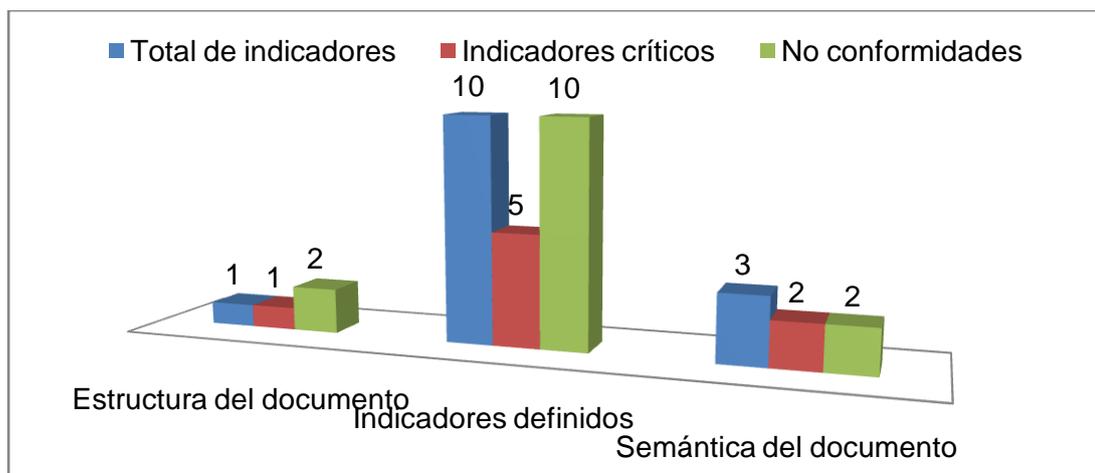


Figura 18 Comportamiento de los indicadores por secciones

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. Se parte de la descripción de los reportes candidatos, como apoyo para las revisiones. Cada planilla de caso de prueba recoge la especificación de un caso de uso de información, detallando las funcionalidades descritas en él y describiendo cada variable que recoge el caso de uso en cuestión, además quedan plasmadas las revisiones realizadas al caso de prueba; así como un registro de todo aquello que no corresponde a la calidad del software.

Para detallar el caso de uso se utiliza una tabla, donde se desglosa esta funcionalidad en secciones (SC) y a su vez estas en escenarios (EC), para hacer más fructífera la ejecución de las pruebas. Esta tabla contiene los campos:

- **Escenario:** se especifica el nombre del escenario.
- **Descripción:** se describe brevemente el escenario.
- **Variables [1, 2,..., n]:** se especifica el nombre de la variable, o el número según la tabla de descripción de variables.
- **Respuesta del sistema:** se escribe el resultado que se esperaba al realizar la prueba.
- **Resultado de la prueba:** se escribe el resultado que se obtuvo al realizar la prueba.
- **Flujo central:** se describe el flujo central del caso de uso al que se le estaba diseñando el caso de prueba.

Capítulo 4: Validación del Mercado de datos para las series históricas del Sector externo

A continuación se muestra la descripción del caso de uso Consultar información de intercambio comercial, de la SC Reportes candidatos:

Tabla 9 Sección Reportes candidatos

Escenario	Descripción	Perfiles de análisis	Indicadores a medir	Respuesta del sistema	Flujo central
EC 1.1: Obtener intercambio comercial de mercancías por país de origen y áreas geográficas en un tiempo dado	Permite visualizar el reporte con las variables presentes en el mismo.	Año País Origen	Intercambio (MP)	Se muestra la tabla con los valores correspondientes al escenario.	Se abre la aplicación Se autentifica Se entra al sistema Se selecciona la opción Nueva Vista de Análisis Se selecciona el esquema comercio_ext_series Se selecciona el cubo Exportación
EC 1.2: Obtener intercambio total y saldo comercial en un tiempo dado	Permite visualizar el reporte con las variables presentes en el mismo.	Año	Exportaciones de Mercancías (MP) Importaciones de Mercancías (MP) Intercambio (MP) Saldo Comercial (MP)	Se muestra la tabla con los valores correspondientes al escenario.	Se selecciona la opción Abrir navegador OLAP en la parte superior izquierda. Se verifica que estén disponibles en las filas y las columnas todos los perfiles de análisis y los indicadores a medir.

En los casos de pruebas también se detallan las variables que son utilizadas, esta tabla contiene los campos:

- **No:** se enumeran todos los campos ó variables descritas en el Caso de Uso.
- **Nombre de campo:** nombre del campo de entrada.
- **Clasificación:** la clasificación es según el componente de diseño utilizado. Ejemplo: campo de texto, lista desplegable o campo de selección.
- **Valor nulo:** se especifica si el campo puede ser nulo o no. Para ello solo se pone Sí o No.
- **Descripción:** una breve descripción de los datos que deben introducirse. Reglas que tiene que cumplir el campo.

Capítulo 4: Validación del Mercado de datos para las series históricas del Sector externo

A continuación se muestra la descripción de las variables del caso de uso Consultar información de intercambio comercial:

Tabla 10 Descripción de las variables

No	Nombre de campo	Clasificación	Valor Nulo	Descripción
1	Año	Lista desplegable	No.	Año en que se realiza la actividad comercial.
2	País Origen	Lista desplegable	No.	Países de origen las mercancías.
3	País Destino	Lista desplegable	No.	Países destino de las mercancías.
4	Unidad de Medida	Lista desplegable	No.	Recoge la descripción de las unidades de medidas por la cual se miden los productos.
5	Exportaciones de Mercancías (MP)	Valor fijo	No.	Precio de la mercancía, no incluye fletes, seguros y otros gastos de manipulación.
6	Importaciones de Mercancías (MP)	Valor fijo	No.	Precio de la mercancía incluyendo el costo, seguro y fletes.
7	Intercambio (MP)	Valor fijo calculable	No.	Exportaciones de Mercancías + Importaciones de Mercancías
8	Saldo Comercial (MP)	Valor fijo calculable	No.	Exportaciones de Mercancías - Importaciones de Mercancías

El mercado fue evaluado por el grupo de calidad de la universidad, haciendo uso de los casos de prueba; esta prueba arrojó en total de 11 no conformidades. A continuación se muestra una imagen con la cantidad de no conformidades por caso de prueba.

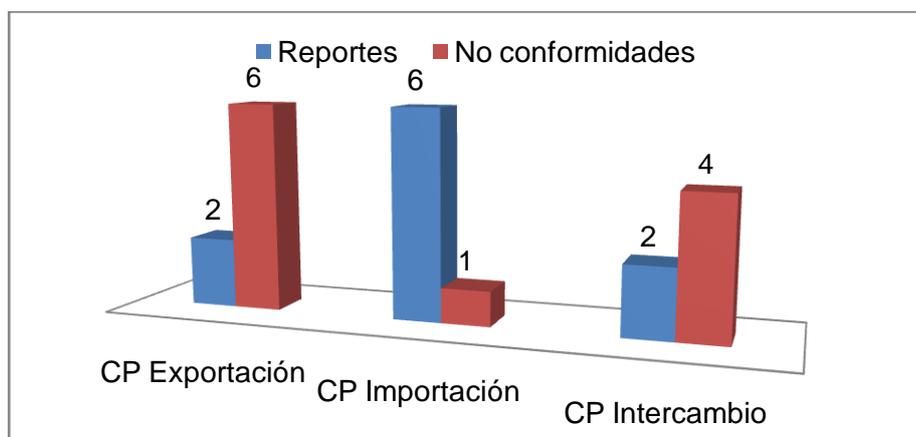


Figura 19 Resultados de los casos de prueba

Capítulo 4: Validación del Mercado de datos para las series históricas del Sector externo

En este epígrafe se hace referencia al caso de prueba para el caso de uso Consultar información de intercambio comercial. Los casos de pruebas para el caso de uso Consultar información de exportaciones y el caso de uso Consultar información de importaciones pueden verse en los Anexos 2 y 3 respectivamente.

Pruebas de aceptación

Durante el desarrollo del mercado de datos se aplicaron las pruebas de aceptación por la representante de la entidad cliente: Elena L. Fernández, intensificándose dicho proceso de pruebas en la última quincena de mayo y el mes de junio. Durante este período se definieron las principales funcionalidades a añadir y unido a esto se le presentaron los resultados obtenidos en el transcurso de los procesos. Durante la revisión de la aplicación el cliente solicitó cambiar la estructura de los reportes de acuerdo al análisis que se realizará posteriormente a la información, la cual fue atendida inmediatamente.

4.3 Conclusiones

- Con la intención de una mejor organización de la ejecución de las pruebas se diseñaron y aplicaron 3 casos de prueba para validar la solución, arrojando 11 no conformidades que fueron solucionadas en un corto plazo de tiempo.
- Se diseñó y aplicó una lista de chequeo para evaluar una serie de indicadores que se encuentran implicados en el proceso de la creación de la capa de integración y visualización del sistema, además de medir la calidad de los artefactos y documentos generados en la realización del producto, arrojando 14 no conformidades que fueron solucionadas en un corto plazo de tiempo.
- Se realizaron las pruebas de aceptación al mercado de datos, arrojando los resultados esperados por el cliente.

Conclusiones

Conclusiones

Luego de concluida toda la investigación que trajo consigo el desarrollo de un mercado de datos para la ONE relacionado con las series históricas del Sector externo se arribó a las siguientes conclusiones:

- Haciendo uso del Modelo para el Desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio en DATEC se identificaron los requisitos de información, los funcionales y no funcionales. Además quedó definido el modelo lógico que estructura el mercado de datos.
- Como resultado de la implementación se obtuvo un mercado de datos poblado y funcional, con información disponible para ser consultada por parte de los usuarios.
- Quedó implementada la capa de visualización de los datos con los reportes disponibles para ser analizada la información, permitiendo apoyar el proceso de toma de decisiones en el área del Sector externo en la ONE específicamente para las series históricas.
- Se realizó la validación del mercado de datos a través las listas de chequeo, los casos de prueba y las pruebas de aceptación, obteniendo resultados satisfactorios.

Recomendaciones

Recomendaciones

- Extender el uso de la aplicación a otras instituciones que recopile y organice los datos estadísticos del Sector externo en Cuba.
- Utilizar este trabajo como guía para todas aquellas personas que deseen aprender a implementar mercados de datos sobre tecnología de Almacenamiento de Datos.
- Integrar el mercado de datos para las series históricas del Sector externo con el Sistema Informático de Gestión Estadística (SIGE).

Referencias Bibliográficas

1. **Grinquévich, Vlad y NOVOSTI, RIA** . BRIC, una comunidad de economías emergentes tras la crisis. *El Economista de Cuba*. [En línea] [Citado el: 1 de Diciembre de 2010.] <http://www.eleconomista.cubaweb.cu/2010/nro377/bric.html>.
2. Informe de Cuba sobre la Resolución 63/7 . *Cuba Misión Permanente*. [En línea] [Citado el: 1 de Diciembre de 2010.] <http://embacuba.cubaminrex.cu/Default.aspx?tabid=11013>.
3. **Febres Arellano, Ana Alejandra**. Conociendo al SIDUNEA - Sistema Aduanero Automatizado. *Gestiopolis*. [En línea] Marzo de 2004. [Citado el: 1 de Diciembre de 2010.] <http://www.gestiopolis.com/canales2/economia/sidunea.htm>.
4. Preguntas Frecuentes ISIDORA. *Servicio Nacional de Aduana*. [En línea] [Citado el: 1 de Diciembre de 2010.] http://www.aduana.cl/prontus_aduana/site/artic/20070228/pags/20070228162102.html.
5. Sistema Estadístico de Comercio Exterior. *Websiex*. [En línea] [Citado el: 1 de Diciembre de 2010.] http://websiex.dian.gov.co/siex/Que_es_Siex_/que_es_siex_.html.
6. **W, K, Inmon**. *Building the Data Warehouse*. s.l. : Wiley: Chichester, 2002. Vol. III.
7. AD. *EcuRed*. [En línea] [Citado el: 20 de Noviembre de 2010.] http://www.ecured.cu/index.php/Almac%C3%A9n_de_Datos#Definici.C3.B3n_de_Ralph_Kimball.
8. **R, Elmasri y S. B, Navathe**. *Fundamentos de sistemas de bases de datos*. s.l. : University of Texas at Arlington, 2002.
9. **Casales Cabrera , María Evelia**. Data Warehouse (Almacenes de Datos). [En línea] Enero de 2009. [Citado el: 30 de Noviembre de 2010.] <http://hp.fciencias.unam.mx/~alg/bd/dwh.pdf>.
10. **CONNOLLY, Thomas M. y BEGG, Carolyn E.** *CONNOLLY, Thomas M.; BEGG, Carolyn E., Sistemas de Bases de datos: Un enfoque práctico para diseño, implementación y gestión*. España : Pearson, 2005.
11. *Redk.net*. [En línea] [Citado el: 30 de Noviembre de 2010.] <http://www.redk.net/procesos-etl.html>.
12. ¿Qué es Business Intelligence? *Sinnexus*. [En línea] [Citado el: 30 de Noviembre de 2010.] http://www.sinnexus.com/business_intelligence/.
13. Apuntes. Ingeniería del software. Sistemas Informaticos. Nivel de madurez software. Informatica Aplicada a la Gestión Pública. *Página profesor Rafael Barzanallana. Universidad de Murcia (España)*. [En línea] 16 de Diciembre de 2006. [Citado el: 1 de Diciembre de 2010.] <http://www.um.es/docencia/barzana/LAGP/lagp2.html>.
14. **Rodríguez Yunta, Luis**. BASES DE DATOS DOCUMENTALES: ESTRUCTURA Y PRINCIPIOS DE USO. *Universidad de Navarra*. [En línea] 2001. [Citado el: 22 de Octubre de 2010.] www.unav.es/dpp/documentacion/proteger/lryunta.pdf.
15. **Mundy, Joy**. Kimball University: Building a Foundation for Smart Applications. *InformationWeek*. [En línea] 1 de Diciembre de 2006. [Citado el: 20 de Octubre de 2010.] http://www.kimballgroup.com/html/articles_search/articles2006/0612IE.html.
16. **Bernabeu R, Dario**. Datawarehouse manager. *Dataprix*. [En línea] 6 de Mayo de 2009. [Citado el: 20 de Octubre de 2010.] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse/34-datawarehouse-manager#x1-500003.4.5.1>.
17. **Gill, Harjinder S**. *Data Warehousing. La integración de información para la mejor toma de decisiones*. México : s.n., 1996.

Referencias Bibliográficas

18. *Sistemas de Bases de datos: Un enfoque práctico para diseño, implementación y gestión*. **M., Thomas y BEGG, Carolyn E.** España : Pearson, 2005, CONNOLLY.
19. Manual para la adquisición de un sistema de Data Warehouse. *Dataprix*. [En línea] [Citado el: 18 de Noviembre de 2010.] <http://www.dataprix.com/olap-rolap-molap>.
20. *Sinnexus*. [En línea] [Citado el: 18 de Noviembre de 2010.] http://www.sinnexus.com/business_intelligence/index.aspx.
21. **SILBERSCHATZ, ABRAHAM, KORTH, HENRY F. y S., SUDARSHAN.** *Fundamentos de bases de datos*. s.l. : MCGRAW-HILL, 1995.
22. Sobre PostgreSQL. *PostgreSQL-es*. [En línea] 2 de Octubre de 2010. [Citado el: 16 de Noviembre de 2010.] http://www.postgresql-es.org/sobre_postgresql.
23. Herramientas Case. [En línea] [Citado el: 23 de Noviembre de 2010.] <http://www1.inei.gob.pe/biblioineipub/bancopub/Inf/Lib5103/Libro.pdf>.
24. **Pressman, Roger S.** *Ingeniería de Software, un enfoque práctico*. *Ecured*. [En línea] 2002. [Citado el: 18 de Noviembre de 2010.] http://www.ecured.cu/index.php/Visual_Paradigm.
25. Criterios para evaluar una plataformas ETL o de integración de datos. *CMS-Spain.com*. [En línea] 6 de Abril de 2009. [Citado el: 24 de Noviembre de 2010.] <http://www.ecm-spain.com/interior.asp?IdItem=4813>.
26. PENTAHO. *Gravitar*. [En línea] [Citado el: 20 de Noviembre de 2010.] <http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.
27. Herramientas BI. *Gravitar*. [En línea] [Citado el: 30 de Noviembre de 2010.] <http://www.gravitar.biz/index.php/herramientas-bi/>.
28. **Conesa Caralt, Jordi y Curto Díaz, Josep.** *Introducción al Bussines Intelligence*. s.l. : UOC - EDIUOC, 2010.
29. El servidor OLAP Mondrian . *Portada sobre la plataforma Pentaho Open Source Business Intelligence* . [En línea] [Citado el: 29 de Noviembre de 2010.] <http://pentaho.almacen-datos.com/mondrian.html>.
30. Apache Tomcat. *The Apache Software Foundation* . [En línea] [Citado el: 30 de Noviembre de 2010.] <http://tomcat.apache.org/>.
31. Pentaho BI Platform y servidor. *Pentaho*. [En línea] [Citado el: 30 de Noviembre de 2010.] http://community.pentaho.com/projects/bi_platform/.
32. *Pentaho open source business intelligence*. [En línea] [Citado el: 22 de Enero de 2011.] <http://www.pentaho.com>.
33. **Marqués Andrés, María Mercedes.** *Modelos de datos*. *Universitat Jaume-I*. [En línea] 2 de Diciembre de 2001. [Citado el: 5 de Abril de 2011.] <http://www3.uji.es/~mmarques/f47/apun/node32.html>.

Bibliografía

- ¿Qué es Business Intelligence? *Sinnexus*. [En línea] [Citado el: 30 de Noviembre de 2010.] http://www.sinnexus.com/business_intelligence/.
- **Alcocer, Alberto**. Dimension Tecnologías de la Información . *Dimension Tecnologías de la Información* . [En línea] 2003. [Citado el: 26 de 03 de 2010.] <http://www.dimensionti.com/prod051.htm>.
- Almacén de Datos. *EcuRed*. [En línea] [Citado el: 20 de Noviembre de 2010.] http://www.ecured.cu/index.php/Almac%C3%A9n_de_Datos#Definici.C3.B3n_de_Ralph_Kimbal.
- **Alvarez, Sara**. Modelo entidad-relación. *DesarrolloWeb.com*. [En línea] 12 de Septiembre de 2007. [Citado el: 21 de Noviembre de 2010.] <http://www.desarrolloweb.com/articulos/modelo-entidad-relacion.html>.
- Apache Tomcat. *The Apache Software Foundation* . [En línea] [Citado el: 30 de Noviembre de 2010.] <http://tomcat.apache.org/>.
- Apuntes. Ingeniería del software. Sistemas Informaticos. Nivel de madurez software. Informatica Aplicada a la Gestión Pública. *Página profesor Rafael Barzanallana. Universidad de Murcia (España)*. [En línea] 16 de Diciembre de 2006. [Citado el: 1 de Diciembre de 2010.] <http://www.um.es/docencia/barzana/IAGP/lagp2.html>.
- **Bernabeu R, Dario**. Datawarehouse manager. *Dataprix*. [En línea] 6 de Mayo de 2009. [Citado el: 20 de Octubre de 2010.] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse/34-datawarehouse-manager#x1-500003.4.5.1>.
- Business Intelligence - Almacenes de Datos - ETL. *ETL-Tools.Info*. [En línea] [Citado el: 30 de Noviembre de 2010.] http://etl-tools.info/es/bi/proceso_etl.htm.
- **Casales Cabrera , María Evelia**. Data Warehouse (Almacenes de Datos). [En línea] Enero de 2009. [Citado el: 30 de Noviembre de 2010.] <http://hp.fciencias.unam.mx/~alg/bd/dwh.pdf>.
- **Catrin, Franco**. PostgreSQL 8.4 mantiene distancia con sus rivales. *FayerWayer*. [En línea] [Citado el: 1 de Diciembre de 2010.] <http://www.fayerwayer.com/2009/07/postgresql-84-mantiene-distancia-con-sus-rivales/>.
- **Cavazos, Eliezer** . Metodología para Business Intelligence. *Gravitar*. [En línea] 30 de Septiembre de 2008. [Citado el: 2 de Noviembre de 2010.] <http://www.gravitar.biz/index.php/bi/metodologia-business-intelligence/>.
- **Conesa Caralt, Jordi y Curto Díaz, Josep**. *Introducción al Bussines Intelligence*. s.l. : UOC - EDIUOC, 2010.
- **CONNOLLY, Thomas M. y BEGG, Carolyn E.** *CONNOLLY, Thomas M.; BEGG, Carolyn E., Sistemas de Bases de datos: Un enfoque práctico para diseño, implementación y gestión*. España : Pearson, 2005.
- Criterios para evaluar una plataformas ETL o de integración de datos. *CMS-Spain.com*. [En línea] 6 de Abril de 2009. [Citado el: 24 de Noviembre de 2010.] <http://www.ecm-spain.com/interior.asp?IdItem=4813>.
- Data Warehousing and Business Intelligence. *1keydata*. [En línea] [Citado el: 22 de Noviembre de 2010.] <http://www.1keydata.com/datawarehousing/datawarehouse.html>.
- **Díaz, Néstor A**. Tiendo Linux. *Tiend Linux*. [En línea] 2003. [Citado el: 22 de 03 de 2010.] http://soporte.tiendalinux.com/portal/Portfolio/postgresql_ventajas_html.
- El servidor OLAP Mondrian . *Portada sobre la plataforma Pentaho Open Source Business Intelligence* . [En línea] [Citado el: 29 de Noviembre de 2010.] <http://pentaho.almacen-datos.com/mondrian.html>.

Bibliografía

- *Embarcadero ER*. [En línea] [Citado el: 18 de Noviembre de 2010.] <http://www.scribd.com/doc/16626706/Embarcadero-ER>.
- **Errecalde, Marcelo Luis**. Preparación de los datos en el proceso KDD. [En línea] Diciembre de 2007. [Citado el: 21 de Noviembre de 2010.]
- **Febres Arellano, Ana Alejandra**. Conociendo al SIDUNEA - Sistema Aduanero Automatizado. *Gestiopolis*. [En línea] Marzo de 2004. [Citado el: 1 de Diciembre de 2010.] <http://www.gestiopolis.com/canales2/economia/sidunea.htm>.
- *Fundamento de bases de datos*. s.l. : 2da Edición, 1995.
- **Gill, Harjinder S**. *Data Warehousing. La integración de información para la mejor toma de decisiones*. México : s.n., 1996.
- **Grinquevich, Vlad y NOVOSTI, RIA** . BRIC, una comunidad de economías emergentes tras la crisis. *El Economista de Cuba*. [En línea] [Citado el: 1 de Diciembre de 2010.] <http://www.eleconomista.cubaweb.cu/2010/nro377/bric.html>.
- **Group, Work**. ETL-Tools.Info. *ETL-Tools.Info*. [En línea] 2006. [Citado el: 13 de 04 de 2010.] http://etl-tools.info/es/bi/almacenedatos_esquema-estrella.htm.
- Herramientas BI. *Gravitar*. [En línea] [Citado el: 30 de Noviembre de 2010.] <http://www.gravitar.biz/index.php/herramientas-bi/>.
- Herramientas Case. [En línea] [Citado el: 23 de Noviembre de 2010.] <http://www1.inei.gob.pe/biblioineipub/bancopub/Inf/Lib5103/Libro.pdf>.
- Informe de Cuba sobre la Resolución 63/7 . *Cuba Misión Permanente*. [En línea] [Citado el: 1 de Diciembre de 2010.] <http://embacuba.cubaminrex.cu/Default.aspx?tabid=11013>.
- **Intercom, Grupo**. Softonic. *Softonic*. [En línea] INTERSHARE, S.L., 2004. [Citado el: 06 de 02 de 2010.] <http://estadisticas-baloncesto.softonic.com/>.
- **Jiménez, Mariana**. PENTAHO: Creación de un cubo con Mondrian. *Gravital*. [En línea] 22 de Julio de 2008. [Citado el: 26 de Noviembre de 2010.] <http://www.gravitar.biz/index.php/bi/pentaho-ejemplo-cubo-mondrian/>.
- **Kimball, R., Reeves, L., Ross, M., & Thornthwaite, W**. *The data warehouse lifecicle toolkit*.
- Manual para la adquisición de un sistema de Data Warehouse. *Dataprix*. [En línea] [Citado el: 18 de Noviembre de 2010.] <http://www.dataprix.com/olap-rolap-molap>.
- **Marqués Andrés, María Mercedes**. Modelos de datos. *Universitat Jaume-I*. [En línea] 2 de Diciembre de 2001. [Citado el: 5 de Abril de 2011.] <http://www3.uji.es/~mmarques/f47/apun/node32.html>.
- **Martha Julieta Reyna, Bibiana Prado, Roberto Taboada, Agustin Copca**. Universidad Autónoma del Estado de Morelos. *Universidad Autónoma del Estado de Morelos*. [En línea] 2009. [Citado el: 02 de 02 de 2010.] www.uaem.mx/posgrado/mcruz/cursos/miic/oracle3.ppt.
- **Mendez, A., Mártire, A., Britos, P. Y Garcia-Martínez, R**. Fundamentos de Data Warehouse. [En línea] 2003. [Citado el: 22 de Noviembre de 2010.] <http://artemisa.unicauca.edu.co/~ecaldon/docs/bd/fundamentosdedatawarehouse.pdf>.
- **Mundy, Joy**. Kimball University: Building a Foundation for Smart Applications. *InformationWeek*. [En línea] 1 de Diciembre de 2006. [Citado el: 20 de Octubre de 2010.] http://www.kimballgroup.com/html/articles_search/articles2006/0612IE.html.
- **ONE**. Oficina Nacional de Estadísticas. *Oficina Nacional de Estadísticas*. [En línea] ONE, 2006. [Citado el: 20 de 01 de 2010.] <http://www.one.cu/>.
- **Pecos, Daniel**. PostGreSQL vs. MySQL. *PostGreSQL vs. MySQL*. [En línea] 2010. [Citado el: 11 de 02 de 2010.] http://danielpecos.com/docs/mysql_postgres/x57.html.
- Pentaho BI Platform y servidor. *Pentaho*. [En línea] [Citado el: 30 de Noviembre de 2010.] http://community.pentaho.com/projects/bi_platform/.

Bibliografía

- *Pentaho open source business intelligence*. [En línea] [Citado el: 22 de Enero de 2011.] <http://www.pentaho.com>.
- PENTAHO. *Gravitar*. [En línea] [Citado el: 20 de Noviembre de 2010.] <http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.
- **Pmdalmeria**. Patrimonio municipal de Deportes de Almería. . *Patrimonio municipal de Deportes de Almería*. . [En línea] 2005. [Citado el: 20 de 03 de 2010.] <http://www.pmdalmeria.com/documentos/subvenciones/anexo2.pdf>.
- Preguntas Frecuentes ISIDORA. *Servicio Nacional de Aduana*. [En línea] [Citado el: 1 de Diciembre de 2010.] http://www.aduana.cl/prontus_aduana/site/artic/20070228/pags/20070228162102.html.
- **Pressman, Roger S.** Ingeniería de Software, un enfoque práctico. *Ecured*. [En línea] 2002. [Citado el: 18 de Noviembre de 2010.] http://www.ecured.cu/index.php/Visual_Paradigm.
- **R, Elmasri y S. B, Navathe**. *Fundamentos de sistemas de bases de datos*. s.l. : University of Texas at Arlington, 2002.
- *Redk.net*. [En línea] [Citado el: 30 de Noviembre de 2010.] <http://www.redk.net/procesos-etl.html>.
- **Rodríguez Yunta, Luis**. BASES DE DATOS DOCUMENTALES: ESTRUCTURA Y PRINCIPIOS DE USO. *Universidad de Navarra*. [En línea] 2001. [Citado el: 22 de Octubre de 2010.] www.unav.es/dpp/documentacion/proteger/lryunta.pdf.
- **SILBERSCHATZ, ABRAHAM, KORTH, HENRY F. y S., SUDARSHAN**. *Fundamentos de bases de datos*. s.l. : MCGRAW-HILL, 1995.
- **Sinnexus**. [En línea] [Citado el: 18 de Noviembre de 2010.] http://www.sinnexus.com/business_intelligence/index.aspx.
- **SIRCAS**. Sistema Interactivo de Referenciación Ambiental Sectorial. *Sistema Interactivo de Referenciación Ambiental Sectorial*. [En línea] [Citado el: 12 de 02 de 2007.] <http://www.sirac.info/Curtiembres/html/indicadores.asp>.
- **Sistema Estadístico de Comercio Exterior**. *Websiex*. [En línea] [Citado el: 1 de Diciembre de 2010.] http://websiex.dian.gov.co/siex/Que_es_Siex_/que_es_siex_.html.
- **Sistemas de Bases de datos: Un enfoque práctico para diseño, implementación y gestión. M., Thomas y BEGG, Carolyn E.** España : Pearson, 2005, CONNOLLY.
- **Sobre PostgreSQL**. *PostgreSQL-es*. [En línea] 2 de Octubre de 2010. [Citado el: 16 de Noviembre de 2010.] http://www.postgresql-es.org/sobre_postgresql.
- **Soto**. Mercados De Datos Data Mart. *Mercados De Datos Data Mart*. [En línea] 2007. [Citado el: 20 de 03 de 2010.] <http://www.mitecnologico.com/Main/MercadosDeDatosDataMart..>
- **Trabajo, Colectivo de**. *definicion.de*. [En línea] 2008. [Citado el: 16 de 04 de 2010.] <http://definicion.de/metodologia/>.
- **Trabajo, Grupo de**. MPM-Software. *MPM-Software*. [En línea] 2009. [Citado el: 24 de 04 de 2010.] <http://www.mpm.es/sites/spain/productos/integraciondatos/default.aspx>.
- **Velasco, Roberto Hernando**. Almacenes de datos (Datawarehouse). *Almacenes de datos (Datawarehouse)*. [En línea] 2009. [Citado el: 20 de 03 de 2010.] <http://www.rhernando.net/modules/tutorials/doc/bd/dw.html..>
- **W, K, Inmon**. *Building the Data Warehouse*. s.l. : Wiley: Chichester, 2002. Vol. III.

Anexos

Anexos

Anexo 1: Lista de Chequeo.

Tabla 11 Lista de Chequeo

Estructura del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
crítico	1. ¿Los entregables contienen las secciones obligatorias de la plantilla estándar definidas para un expediente de proyecto? (portada, control de versiones, reglas de confidencialidad, tabla de contenidos y contenido) (ver expediente de proyecto)	1		2	
Indicadores definidos en el desarrollo					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
	1. ¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?	0		0	
crítico	2. ¿Los reportes son configurables a través de la interfaz del sistema?	0		0	
	3. ¿La interfaz está orientada a facilitar el uso de las funciones del sistema por parte de los usuarios?	1		10	Los nombres de los reportes deben ser entendibles para los usuarios.
crítico	4. ¿No existen restricciones para construir cubos OLAP con dimensiones y niveles de agregación ilimitados?	0		0	
crítico	5. ¿Los usuarios son capaces de manipular los resultados de manera que se ajusten a sus necesidades,	0		0	

Anexos

	conformando nuevos reportes?				
	6. ¿El sistema responde de una forma rápida a la información que le sea solicitada por el usuario?	0		0	
	7. ¿El sistema refleja cualquier lógica del negocio para poder responder a preguntas específicas?	0		0	
crítico	8. ¿El sistema garantiza la confidencialidad y seguridad de acceso a los datos por rol de los usuarios?	0		0	
	9. ¿Los datos e información derivados del proceso de análisis realizado mediante la aplicación, apoyan la toma de decisiones en la Institución?	0		0	
crítico	10. ¿Los cambios en los datos se reflejan automáticamente en los reportes de forma instantánea?	0		0	
Semántica del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
crítico	1. ¿Se han identificado errores ortográficos en los entregables?	1		2	
crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?	0		0	
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?	0		0	

Anexo 2: Caso de prueba para el caso de uso Consultar información de exportaciones.

Tabla 12 Caso de prueba para el caso de uso Consultar información de exportaciones

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

Anexos

<p>EC 1.1: Obtener exportaciones de mercancías según secciones de la CUCI en un tiempo dado</p>	<p>Permite visualizar el reporte con las variables presentes en el mismo.</p>	<p>Año Secciones CUCI</p>	<p>Exportaciones de Mercancías (MP)</p>	<p>Se muestra la tabla con los valores correspondientes al escenario.</p>	<p>Se abre la aplicación Se autentifica Se entra al sistema Se selecciona la opción Nueva Vista de Análisis Se selecciona el esquema comercio_ext_series</p>
<p>EC 1.2: Obtener exportaciones de mercancías según secciones de la CUCI en un tiempo dado</p>	<p>Permite visualizar el reporte con las variables presentes en el mismo.</p>	<p>Año País Destino</p>	<p>Exportaciones de Mercancías (MP)</p>	<p>Se muestra la tabla con los valores correspondientes al escenario.</p>	<p>Se selecciona el cubo Exportación Se selecciona la opción Abrir navegador OLAP en la parte superior izquierda. Se verifica que estén disponibles en las filas y las columnas todos los perfiles de análisis y los indicadores a medir.</p>

Anexo 3: Caso de prueba para el caso de uso Consultar información de importaciones.

Tabla 13 Caso de prueba para el caso de uso Consultar información de importaciones

Escenario	Descripción	Perfiles de Análisis	Indicadores a medir	Respuesta del sistema	Flujo central
<p>EC 1.1: Obtener importaciones de mercancías agrupadas en grandes categorías económicas en un tiempo dado</p>	<p>Permite visualizar el reporte con las variables presentes en el mismo.</p>	<p>Año Categoría</p>	<p>Importaciones de Mercancías (MP)</p>	<p>Se muestra la tabla con los valores correspondientes al escenario.</p>	<p>Se abre la aplicación Se autentifica Se entra al sistema Se selecciona la opción Nueva Vista de Análisis Se selecciona el esquema comercio_ext_series</p>
<p>EC 1.2: Obtener importaciones de mercancías</p>	<p>Permite visualizar el reporte con las variables</p>	<p>Año Secciones CUCI</p>	<p>Importaciones de Mercancías (MP)</p>	<p></p>	<p>Se selecciona el cubo Exportación Se selecciona la</p>

Anexos

según secciones de la CUCI en un tiempo dado	presentes en el mismo.				opción	Abrir navegador OLAP en la parte superior izquierda.
EC Obtener importaciones de mercancías según tipo de comercio en un tiempo dado	1.3: Permite visualizar el reporte con las variables presentes en el mismo.	Año Régimen	Importaciones de Mercancías (MP)	Se muestra la tabla con los valores correspondientes al escenario.		Se verifica que estén disponibles en las filas y las columnas todos los perfiles de análisis y los indicadores a medir.
EC Obtener importaciones según lugar de entrada de las mercancías en un tiempo dado	1.4: Permite visualizar el reporte con las variables presentes en el mismo.	Año Aduana	Importaciones de Mercancías (MP)	Se muestra la tabla con los valores correspondientes al escenario.		
EC Obtener importaciones según país de origen de las mercancías y áreas geográficas en un tiempo dado	1.5: Permite visualizar el reporte con las variables presentes en el mismo.	Año País Origen	Importaciones de Mercancías (MP)	Se muestra la tabla con los valores correspondientes al escenario.		
EC Obtener importaciones según país de procedencia de las mercancías y áreas geográficas en un tiempo dado	1.6: Permite visualizar el reporte con las variables presentes en el mismo.	Año País Procedencia	Importaciones de Mercancías (MP)	Se muestra la tabla con los valores correspondientes al escenario.		

Glosario

Automatizar: aplicar procedimientos automáticos a un aparato, proceso o sistema: Han automatizado la biblioteca universitaria.

Base de Datos: es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso.

Clasificador: es un instrumento que asigna un código a elementos ya definidos por otras vías.

Digitalizar: transformar una información a un sistema de dígitos para su tratamiento informático: digitalizar una fotografía.

Estadísticas: ciencia con base matemática referente a la recolección, análisis e interpretación de datos, que busca explicar condiciones regulares en fenómenos de tipo aleatorio.

Funcionalidad: conjunto de características que hacen que algo sea práctico y utilitario: en el diseño de este vehículo se ha buscado la funcionalidad.

Implementación: poner en funcionamiento, aplicar los métodos y medidas necesarios para llevar algo a cabo: implementar un algoritmo.

Indicador: se dice de la variable que puede tomar un valor de una determinada unidad de medida y de un determinado tipo de datos (generalmente numérico).

Informática: ciencia aplicada que abarca el estudio y aplicación del tratamiento automático de la información, utilizando dispositivos electrónicos y sistemas computacionales. También está definida como el procesamiento automático de la información.

Metodología: es el conjunto de procedimientos basados en principios lógicos, utilizados para alcanzar una gama de objetivos que rigen en una investigación científica o en una exposición doctrinal

Modelo: prototipo digno de ser imitado que se toma como pauta a seguir.

Multidimensional: tiene varias dimensiones: espacio multidimensional.

Procesamiento: tratamiento de la información: procesamiento de ficheros.

Sistema: conjunto de elementos dinámicamente relacionados formando una actividad para alcanzar un objetivo operando sobre datos, energía o materia para proveer información.

Tecnología: es el conjunto de saberes, habilidades, destrezas y medios necesarios para llegar a un fin predeterminado mediante el uso de objetos artificiales o artefactos.