

**Universidad de las Ciencias Informáticas**

**Facultad 6**



**Sistema de Información de Gobierno. Mercado de datos Servicios de comercio exterior**

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

**Autor:**

Jorge Gómez Crespo.

**Tutores:**

Ing. Yunier Santana Aldana.

Ing. Yosvany Arrastia Machín.

**La Habana, junio 2012**

**“Año 54 de la Revolución”**



**"Seamos realistas y hagamos lo imposible."**

**Ché**

## DECLARACIÓN DE AUTORÍA

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

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

---

Jorge Gómez Crespo  
Autor

---

Ing. Yunier Santana Aldana  
Tutor

---

Ing. Yosvany Arrastia Machin  
Tutor

## DATOS DE CONTACTO

Tutor: Ing. Yunier Santana Aldana

Email: ysaldana@uci.cu

Graduado en la Universidad de las Ciencias Informáticas en 2009.

Tutor: Ing. Yosvany Arrastia Machin

Email: yarrastia@uci.cu

Graduado en la Universidad de las Ciencias Informáticas en 2010.

## AGRADECIMIENTOS

A mi familia en general, en especial a mis padres y a mi hermana por ser la luz que me guía en el camino de la vida, por brindarme amor y dedicación en todo los momentos difíciles y por demostrarme que un sueño puede hacerse realidad.

A mi novia Meilin por su compañía y amor.

A mis tutores Tito y Yosvany que hicieron posible que el trabajo de diploma pudiera terminarse en tiempo a pesar de las enormes dificultades enfrentadas en todo el transcurso del curso, por demostrarme tranquilidad, apoyo y consejo en todo momento.

A todos los profesores del departamento que en un momento u otro contribuyeron a mi formación como ingeniero.

Al tribunal por corregirme cada detalle y por sus innumerables recomendaciones.

A mis compañeros del grupo 6105 que durante todos estos años a pesar de no estar juntos me han acompañado y ayudado.

A todos mis amigos por apoyarme cada día y ser parte de este sueño: Andry, Ernesto, Yusliel, Yosbel, Joanni, Osmar, Yan, Sergio, Yeilen, Alejandro, Yeney, Adriana.

A todas mis amistades del grupo 6507.

A todas las personas que de una forma u otra fueron parte de este resultado.

## DEDICATORIA

A mi familia en general, en especial a mis padres por su eterno amor, perseverancia y preocupación. Por guiarme en los senderos de la vida y por encontrar siempre en ellos amor y confianza.

A mi hermana por su espíritu de tranquilidad y ternura.

A mi novia Meilin por todo su amor y apoyo incondicional, que sin duda ha sido de vital importancia en este año para hacer este sueño realidad.

## RESUMEN

La presente investigación surge por la necesidad de construir un mercado de datos para el área de Servicios de comercio exterior perteneciente a la Oficina Nacional de Estadística e Información (ONEI), que facilite el proceso de la toma de decisiones. Para el desarrollo de la solución de este trabajo se hizo uso de varias herramientas de la suite de Pentaho entre otras tecnologías, que conjuntamente con la utilización de la metodología para el desarrollo de almacenes de datos propuesta por el Centro de Tecnologías de Gestión de Datos sirvió de guía en el proceso de desarrollo. Una vez concluido el proceso de desarrollo se obtuvo como resultado un mercado de datos poblado y funcional aprobado por los clientes de la ONEI con un total de doce reportes agrupados en un libro de trabajo.

Palabras claves: ONEI, mercado de datos, toma de decisiones.

## TABLA DE CONTENIDOS

|   |     |
|---|-----|
| AGRADECIMIENTOS.....  | I   |
| DEDICATORIA .....   | II  |
| RESUMEN.....  | III |
| INTRODUCCIÓN.....   | 1   |
| CAPÍTULO 1: FUNDAMENTOS TEÓRICOS .....  | 4   |
| 1.1 Almacenes de datos.....   | 4   |
| 1.1.1 Arquitectura de los almacenes de datos .....                                    | 5   |
| 1.1.2 Experiencias del uso de los almacenes de datos en el mundo.....                 | 6   |
| 1.1.3 Beneficios e inconvenientes de un almacén de datos.....                         | 7   |
| 1.2 Mercado de datos .....  | 8   |
| 1.2.1 Características de los mercados de datos .....                                  | 8   |
| 1.2.2 Beneficios de los mercados de datos .....                                       | 8   |
| 1.3 Integración de datos .....  | 9   |
| 1.3.1 Principales tecnologías .....   | 9   |
| 1.4 Inteligencia de negocio .....   | 10  |
| 1.4.1 Definición de BI.....   | 11  |
| 1.4.2 Características de las herramientas de BI.....                                  | 11  |
| 1.4.3 Beneficios de BI.....   | 11  |
| 1.5 Modos de almacenamiento.....  | 12  |
| 1.6 Tipos de modelados de un almacén de datos .....                                   | 13  |
| 1.7 Metodologías para el desarrollo de los almacenes de datos.....                    | 15  |
| 1.7.1 Fases de la metodología Kimball .....   | 16  |
| 1.7.2 Propuesta de metodología para el desarrollo de Almacenes de Datos en DATEC..... | 16  |
| 1.8 Herramientas utilizadas.....  | 18  |
| 1.8.1 Herramienta de modelado .....   | 18  |
| 1.8.2 Sistema gestor de base de datos.....  | 19  |
| 1.8.3 Herramientas para la limpieza de los datos .....                                | 20  |
| 1.8.4 Herramientas para los procesos de Extracción, Transformación y Carga.....       | 21  |
| 1.8.5 Herramientas utilizadas en Inteligencia de negocio .....                        | 21  |
| CAPÍTULO 2: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS .....                              | 24  |



|   |   |    |
|---|---|----|
| 2.1   | Análisis de la solución .....                                       | 24 |
| 2.1.1   | Definición del negocio .....  | 24 |
| 2.1.2   | Tema de análisis identificado .....                                 | 24 |
| 2.1.3   | Reglas del negocio.....   | 24 |
| 2.1.4   | Necesidades de los usuarios .....                                   | 25 |
| 2.1.5   | Requerimientos.....   | 25 |
|   | Requisitos de información identificados.....                        | 25 |
|   | Requisitos funcionales .....  | 26 |
|   | Requisitos no funcionales .....                                     | 26 |
| 2.1.6   | Actores del sistema.....  | 27 |
| 2.1.7   | Diagrama de casos de usos del sistema .....                         | 27 |
| 2.1.8   | Descripción textual de los casos de uso .....                       | 28 |
|   | CUI 1: Presentar informe de indicadores de comercio exterior.....   | 28 |
| 2.2   | Arquitectura del mercado de datos.....                              | 30 |
| 2.3   | Diseño .....  | 31 |
| 2.3.1   | Subsistema de almacenamiento.....                                   | 31 |
| 2.3.2   | Subsistema de integración .....                                     | 34 |
| 2.3.3   | Subsistema de visualización.....                                    | 36 |
| 2.3.4   | Roles y permisos .....  | 37 |
| CAPÍTULO 3: IMPLEMENTACIÓN DEL MERCADO DE DATOS ..... |   | 39 |
| 3.1   | Implementación del subsistema de almacenamiento.....                | 39 |
| 3.2   | Implementación del subsistema de integración de los datos.....      | 39 |
| 3.2.1   | Implementación de los procesos de ETL.....                          | 40 |
| 3.3   | Implementación del subsistema de visualización .....                | 41 |
| 3.3.1   | Cubos OLAP .....  | 41 |
| 3.3.2   | Creación y administración de los reportes.....                      | 42 |
| CAPÍTULO 4: PRUEBA.....                               |   | 44 |
| 4.1   | Pruebas.....  | 44 |
| 4.1.1   | Elaboración, evaluación y aplicación de las Listas de chequeo ..... | 46 |
| 4.1.2   | Diseño de Casos de pruebas .....                                    | 49 |
| 4.1.3   | Resultados y discusión de las pruebas.....                          | 50 |
| CONCLUSIONES.....                                     |   | 52 |
| RECOMENDACIONES .....                                 |   | 53 |

|                                 |    |
|---------------------------------|----|
| REFERENCIAS BIBLIOGRÁFICAS..... | 54 |
| BIBLIOGRAFÍA.....               | 56 |
| ANEXOS.....                     | 59 |
| GLOSARIO.....                   | 64 |

## INTRODUCCIÓN

El creciente y constante avance de las tecnologías de la información y las comunicaciones (TIC), han venido a desempeñar un papel muy importante en el mundo empresarial, permitiendo recopilar y almacenar gran cantidad de información de todo tipo de procesos: industriales, empresariales, bancarios, de marketing, entre otros; en un mundo en que las tecnologías, la información y las estadísticas marcan el ritmo del progreso y las pautas de la vida.

El análisis de los grandes volúmenes de información que manejan las entidades y las diferentes organizaciones ha permitido incidir de manera directa, sobre los niveles de eficiencia y eficacia de las organizaciones. El adecuado acceso y uso de la información constituirá una condición necesaria, para que cualquier entidad pueda llegar a alcanzar un posicionamiento ventajoso en el mercado, en una época en que la información es la clave para obtener una ventaja competitiva en el mundo de los negocios. Es por esto por lo que las organizaciones buscan cada día más soluciones tecnológicas que les brinden mayor eficiencia, flexibilidad, seguridad y rapidez en el manejo de sus datos.

Cuba a pesar de ser un país subdesarrollado no ha quedado al margen de estos avances informáticos, sino que ha perfeccionado e impulsado el desarrollo del conocimiento en aras de ajustarse a los cambios que se desarrollan en la nueva era de la informatización. El firme progreso en centros como CIMEX, ETECSA y en áreas de la agricultura, las comunicaciones y la medicina, ha propiciado el incremento de estudios y aplicaciones que generan una amplia gama de información, convirtiéndose en muchos casos en registros almacenados muy grandes o complejos de analizar.

El órgano rector de la estadística en Cuba la Oficina Nacional de Estadística e Información (ONEI) es una de las entidades que conjuntamente con el Centro de Tecnologías de Gestión de Datos (DATEC) de la Universidad de las Ciencias Informáticas (UCI) se han trazado estrategias para informatizar sus procesos. La ONEI cuenta con diferentes modelos estadísticos enfocados a todos los sectores de la economía y la sociedad cubana. Actualmente debido a los avances de la tecnología, dicha información crece cada día más provocando que el proceso de toma de decisiones sea más difícil y que el acceso a la información recopilada requiera mayor rapidez y precisión. Una de las áreas de la ONEI es la de Servicios de comercio exterior. De la misma se procesan todos los datos pertenecientes a las relaciones de comercio del país con empresas foráneas que solicitan o prestan servicios al país, sin embargo, controlar todo el proceso de organización de los datos es complejo, debido a que la información se encuentra almacenada en diferentes formatos, como Excel, archivos de texto, archivos Dbf, formato duro (papeles), documentos Word y otros, lo que requiere de un alto nivel de conocimiento tanto informático como del negocio para entender la información que contienen los

ficheros fuentes, los cuales son generados de manera trimestral, y que deben ser procesados para obtener los principales reportes, indicadores, tasas y porcentajes.

Esto trae consigo que haya datos no integrados, incompatibles, múltiples versiones de los mismos, carencia de reportes flexibles y dificultad en el análisis de la información acumulada en el tiempo, dificultando así el proceso de la toma de decisiones.

Por lo anteriormente expuesto se plantea como problema de la investigación: ¿Cómo contribuir a la toma de decisiones en el área de Servicios de comercio exterior del Sistema de Información de Gobierno?

Definiendo como objeto de estudio: los almacenes de datos, delimitando su campo de acción en el mercado de datos Servicios de comercio exterior para el Sistema de Información de Gobierno.

Para solucionar el problema planteado, se identifica como objetivo general de la investigación: Desarrollar el mercado de datos Servicios de comercio exterior para el Sistema de Información de Gobierno, que contribuya a la toma de decisiones.

En correspondencia con el mismo, se definen los siguientes objetivos específicos:

1. Fundamentar la selección de las herramientas, metodologías y tecnologías para el desarrollo de almacenes de datos.
2. Realizar el análisis y diseño del mercado de datos Servicios de comercio exterior para el Sistema de Información de Gobierno.
3. Implementar el mercado de datos Servicios de comercio exterior para el Sistema de Información de Gobierno.
4. Realizar pruebas al mercado de datos Servicios de comercio exterior desarrollados.

Para darle cumplimiento a los objetivos específicos planteados, se propone la realización de las siguientes tareas de investigación:

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

9. Diseño de los casos de pruebas.
10. Implementación del modelo de datos.
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 prueba.

Con el objetivo de obtener los siguientes resultados esperados:

- Mercado de datos poblado.
- Capa de visualización de los datos.

El presente documento está compuesto por 4 capítulos estructurados de la siguiente manera:

Capítulo 1: Fundamentos teóricos para el desarrollo de un mercado de datos

Está referido al análisis del estado del arte de los almacenes de datos y mercados de datos, con sus principales características, metas y elementos que lo componen. Se define cómo se realiza el proceso de gestión de la información en el área Servicios de comercio exterior y se caracterizan las metodologías, técnicas y herramientas a utilizar para el desarrollo de la solución.

Capítulo 2: Análisis y diseño del mercado de datos

Se abordan aspectos referentes al levantamiento de requisitos, a la definición de las reglas del negocio, al modelado de los datos con sus respectivos elementos tales como las dimensiones, los hechos y medidas. Además, se describen los casos de usos identificados con sus relaciones.

Capítulo 3: Implementación del mercado de datos

Se hace referencia a la implementación de la solución, abordando cómo se realizan los procesos de extracción, transformación y carga (ETL), y la implementación del subsistema de visualización para el área Servicios de comercio exterior, teniendo en cuenta los requerimientos y necesidades del negocio.

Capítulo 4: Prueba

Este se refiere a las pruebas realizadas al mercado de datos con sus resultados respectivos, mediante la utilización de las listas de chequeo, los casos de pruebas entre otras pruebas que verificarán si el sistema cumple con las necesidades del cliente.

## CAPÍTULO 1: FUNDAMENTOS TEÓRICOS

En el presente capítulo se definen temas relacionados con el desarrollo de los almacenes de datos y los mercados de datos, se describen los conceptos fundamentales que serán tratados a lo largo de la investigación, así como la selección de la metodología, técnicas y herramientas que serán utilizadas para el desarrollo de la solución.

### 1.1 Almacenes de datos

Los almacenes de datos, también llamados data warehouse, son almacenes de información temática, orientadas a cubrir las necesidades a las organizaciones con el fin de obtener un sistema de soporte para la gestión, control y apoyo de la toma de decisiones y así una ventaja comercial. Es considerado como la solución para acceder a los datos almacenados en sistemas no relacionales.

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

- **Clasificados por temas**, ya que están organizados, de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.
- **Integrados**, ya que las bases de datos contienen datos de todos los sistemas operacionales de la organización y deben ser consistentes.
- **Variables en el tiempo**, los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.
- **No Volátil**, ya que los datos no se modifican ni se eliminan, una vez almacenado un dato, éste se convierte en información de solo lectura y se mantiene para futuras consultas, en lugar de sustituir la información ya existente.

Por tanto, el objetivo de los almacenes de datos es integrar todos los datos en un único almacén en el cual los usuarios puedan ejecutar consultas con facilidad, generar informes y realizar análisis, por tanto, es una tecnología de gestión y de análisis de datos. [1]

Mientras que Ralph Kimball otra de las personalidades destacadas en el tema, define a un almacén de datos como: "una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis". También fue Kimball quien determinó que un almacén de datos no era más que: "la unión de todos los mercados de datos de una entidad".

### 1.1.1 Arquitectura de los almacenes de datos

Un almacén de datos consta de un conjunto de componentes que se fusionan entre sí con un fin común, realizar análisis sobre información histórica y bien estructurada. A continuación se muestra una figura que contiene la presentación de dicha arquitectura:

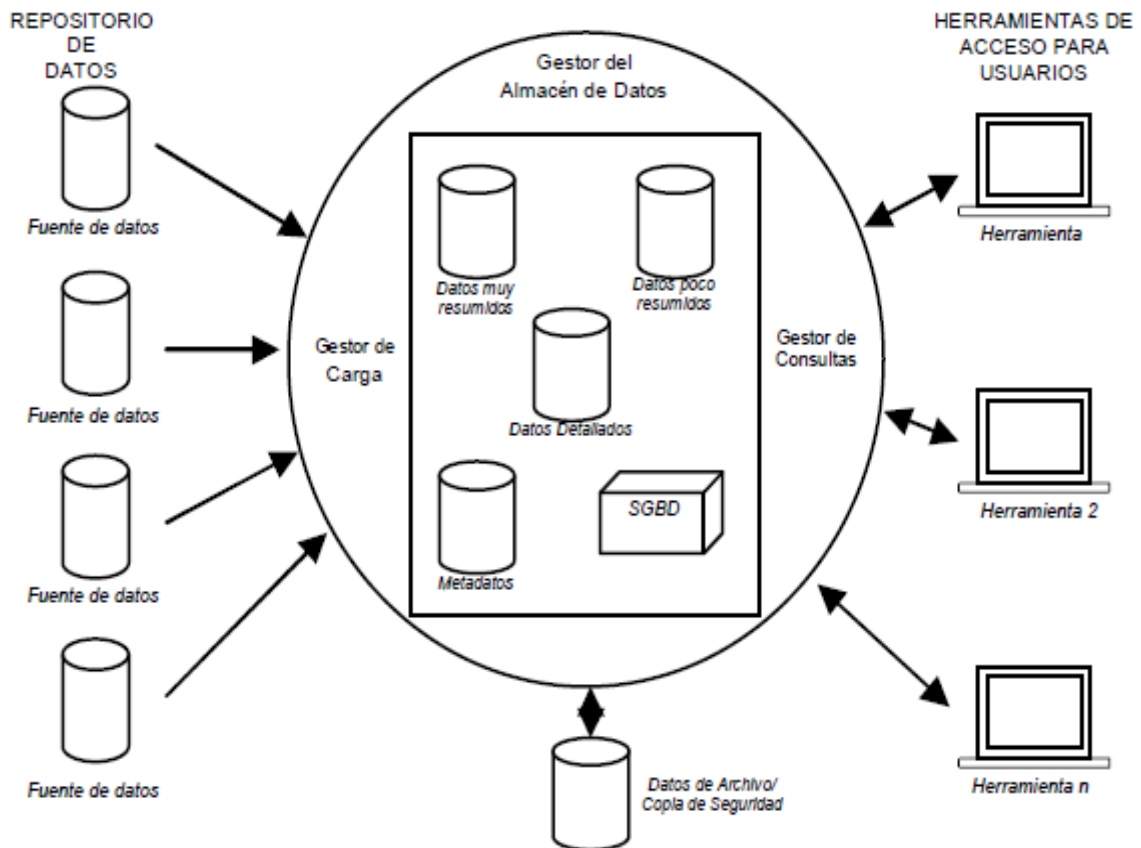


Figura 1: Arquitectura de los almacenes de datos.

#### Repositorio de Datos operacionales

El repositorio de datos operacionales es la fuente donde se encuentran los datos primitivos, actuales e integrados, por lo tanto, es el encargado de suministrar datos al sistema, estos datos operacionales pueden ser:

- Datos de estaciones de trabajo o servidores privados.
- Sistemas externos como las bases de datos comerciales, de proveedores o clientes, o incluso de Internet.
- Datos departamentales almacenados en Sistemas propietarios.

#### Gestor de Carga

Es el encargado de la extracción y carga de los datos del repositorio de datos, además de realizar algunas transformaciones simples a los datos con el fin de que estén adaptados a las necesidades del almacén de datos.

### **Gestor del almacén de datos**

Realiza las operaciones relacionadas con la gestión de los datos dentro del almacén utilizando herramientas específicas que realizan operaciones como la transformación de datos para la incorporación de estos en las tablas del almacén de datos, la creación de índices y vistas de las tablas base, creación de copias de seguridad y archivado de datos, además del análisis de los datos para garantizar la coherencia de los mismos.

### **Tipos de Datos**

En esta área del almacén de datos es donde se encuentran todos los datos contenidos en la base de datos, estos pueden ser:

- **Datos detallados:** Son los que se obtienen directamente del procesado de los datos, no se encuentran almacenados en línea, sino que se puede acceder a ellos con un nivel más bajo de detalle. Se almacenan en disco ocupando mucho espacio, sin embargo, así se facilita el acceso.
- **Datos poco resumidos y muy resumidos:** Contienen los datos que el gestor del almacén de datos ha establecido como resúmenes predefinidos, con el propósito de acelerar las consultas. Pueden verse modificados constantemente con el fin de responder a las variaciones de las consultas. Los datos poco resumidos corresponden al primer nivel de agregación de los datos detallados actualmente, o lo que es lo mismo a las consultas que se realizan habitualmente, siendo estos almacenados en disco. Mientras que los muy resumidos corresponden con el nivel más alto de agregación, o lo que es lo mismo a las consultas que se realizan muy a menudo y que se deben obtener muy rápidamente.
- **Datos de archivo/ copia de seguridad:** Es el área donde se guardan los datos detallados y resumidos con el fin de mantener un archivo y copias de seguridad de los datos.
- **Metadatos:** Describen la estructura de los datos que se contienen en el almacén y pueden ser utilizados por los gestores de carga, del almacén de datos y de consultas. [1]

#### **1.1.2 Experiencias del uso de los almacenes de datos en el mundo**

El ambiente competitivo en las empresas de los años 90 así como el avance tecnológico en materia de sistemas de información, han provocado un nuevo enfoque en el tratamiento y proceso de la información ejecutiva, la cual es un elemento vital hoy en día como soporte en el proceso de toma de



decisiones. El concepto de almacenes de datos, o el proceso de contar con la información más importante de la empresa (incluyendo la histórica), en un solo lugar, ha logrado convertirse en una valiosa herramienta y clave desde el punto de vista tecnológico.

Poco a poco las empresas fueron almacenando un gran número de información en diferentes fuentes de datos (archivos, documentos de texto, bases de datos), y los directivos de las empresas se dieron cuenta de que esta, podría ser útil pues reflejaba la mayoría de las operaciones diarias del negocio.

En 1994 el 90 % de las empresas, según la revista Fortune 2000, planeaba implementar un almacén de datos entre 1994 y 1996. En 1996 el 90 % de las grandes corporaciones consideraba adoptar la tecnología del almacén de datos. En la actualidad empresas como Nike, Adidas, Coca Cola y Google implementan esta tecnología, para almacenar y sacar provecho del enorme volumen de datos que se recogen a diario. De la información recopilada se realizan estudios en cuanto a preferencias y gustos de los clientes que consumen sus productos, brindándole resultados de como se ha ido comportando su negocio a lo largo del tiempo. La implementación de esta tecnología ha permitido a sus directivos tomar decisiones valiosas en momentos definitivos en el negocio, tener una estructura organizada de sus datos, además de enormes ganancias que les hicieron ocupar un lugar ventajoso en el mercado.

En Cuba empresas como ETECSA y CIMEX obtienen experiencias en la implementación de los almacenes de datos. Debido a los grandes volúmenes de información generados por las unidades de las empresas, por la diversidad de los productos en ventas y por la necesidad de mantenerlos correctamente actualizados, sus directivos han decidido utilizar dicha tecnología. Esto ha permitido realizar estudios de los índices de los resultados de las ventas y de los niveles de aceptación de sus productos por parte de sus clientes, convirtiéndose por ejemplo ETECSA en una de las empresas de mayores resultados económicos en el país.

### **1.1.3 Beneficios e inconvenientes de un almacén de datos**

La implantación y uso de un almacén de datos puede beneficiar a una organización en muchos aspectos por ejemplo:

- Los almacenes de datos pueden trabajar en conjunto y, por lo tanto, aumentar el valor operacional de las aplicaciones empresariales, en especial la gestión de relaciones con clientes.
- Mejoran la productividad de los responsables en la toma de decisiones de la organización debido a:
  - ✓ Los almacenes de datos hacen más fácil el acceso a una gran variedad de datos.
  - ✓ Se obtiene una base de datos clasificada por temas e histórica.

- ✓ Integración de información procedente de múltiples sistemas.
- ✓ Mayor organización de los datos.

Sin embargo, utilizar almacenes de datos también plantea algunos problemas como:

- Problemas con los sistemas de origen de los datos.
- Pueden suponer altos gastos. Además de los gastos de mantenimiento que son muy elevados.
- Pueden quedarse obsoletos relativamente pronto si los usuarios incrementan sus necesidades.
- La construcción de un almacén de datos puede requerir de mucho tiempo.
- La integración de las herramientas de almacén de datos, para conseguir un beneficio en la organización, es muy complejo. [1]

## **1.2 Mercado de datos**

Es un subconjunto de datos de un almacén relativos a los requisitos de un departamento o área de negocio concretos. Este subconjunto de datos puede funcionar de forma autónoma, o bien enlazado al almacén de datos. El motivo por el cual se crean mercados de datos es el crecimiento que tiene el almacén y así facilitar su construcción y utilización.

### **1.2.1 Características de los mercados de datos**

Las características de los mercados de datos son:

- Se centra en los requisitos de los usuarios asociados a un departamento o área de negocio concretos.
- Son más sencillos a la hora de utilizarlos y comprender sus datos, debido a que la cantidad de información que contienen es mucho menor que en los almacenes de datos. [1]

### **1.2.2 Beneficios de los mercados de datos**

Beneficios más importantes de los mercados de datos:

- Permiten manejar información confidencial.
- Son simples de implementar y fácil de crear.
- Fácil acceso a los datos que se necesitan de forma frecuente.
- Crea vista colectiva para grupos de usuarios.
- Mejora el tiempo de respuesta al usuario final.
- Conllevan poco tiempo de construcción y puesta en marcha.
- Costo inferior de la aplicación de un completo almacén de datos.

### 1.3 Integración de datos

La mayoría de las organizaciones poseen un alto número de aplicaciones de misión crítica de las cuales dependen para su normal funcionamiento, esas aplicaciones están lógicamente interconectadas, mantienen datos vigentes (poca o ninguna historia) y están optimizadas para una función específica con información fragmentada en diferentes repositorios. En ocasiones se requiere una integración de datos operacionales vigentes de diferentes fuentes. Ejemplo: “¿Cuál es el saldo total de todos los productos de un cliente?” o la información más estratégica requiere integrar datos históricos de diferentes fuentes. Ejemplo: “Predecir comportamiento y valor futuro de un cliente”.

La integración de los datos es uno de los procesos más importantes en la que hay que tener presente en la construcción de un almacén de datos, ya que a partir de esta se garantiza la solidez y calidad de los datos, para después realizar un buen análisis. El proceso de integración de los datos es la unificación de ellos para ser cargados en el almacén, de manera tal que sean entendidos por el sistema que va a utilizarlos. Esta unión puede efectuarse ya sea mediante una replicación de datos o a través de los procesos de ETL. Una integración cuidadosa ayuda a evitar y reducir las redundancias y las inconsistencias.



*Figura 2: Proceso de Integración de datos.*

#### 1.3.1 Principales tecnologías

Entre las principales técnicas para la integración de datos se encuentran:

- Herramientas de integración funcional (EAI: Enterprise Application Integration).
- Bodegas de Datos y herramientas de extracción, transformación y carga (ETL).
- Herramientas de captura, transformación y flujo (CTF: Capture, Transform and Flow).
- Sistemas para manejo de datos federados (EII: Enterprise Information Integration).

#### Extracción, Transformación y Carga

El proceso ETL organiza el flujo de datos entre diferentes sistemas y aporta los métodos y herramientas necesarias para mover los datos de múltiples fuentes, manipularlos, integrarlos,

transformarlos y posteriormente cargar los resultados obtenidos en una base de datos ya sea un almacén de datos o un mercado de datos. El proceso de ETL fortalece los datos para la construcción de bases de datos permanentes dedicadas al análisis o la generación de informes, también es utilizada para conversiones entre bases de datos de un formato a otro diferente. Las funciones específicas del proceso ETL son: Extracción, Transformación y Carga.

### **Extracción**

La primera parte del proceso ETL consiste en extraer los datos necesarios de las fuentes existentes, basándose en las necesidades y requisitos de los usuarios. Las fuentes pueden encontrarse sobre arquitecturas o estructuras heterogéneas, cada sistema puede usar una organización diferente de los datos o diferentes formatos. Por otro lado, los formatos de las fuentes normalmente se encuentran en bases de datos relacionales o ficheros planos, aunque se pueden incluir bases de datos no relacionales u otras estructuras diferentes. Permite almacenar y gestionar los metadatos que se generan en los procesos ETL y facilita la integración de las diversas fuentes de datos. Las herramientas utilizadas en la extracción de datos deben ser adaptables, extensibles y capaces de filtrar los datos relevantes de las fuentes, permitiendo la compresión, descompresión y encriptación de datos. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación.

### **Transformación**

Esta fase es la encargada de convertir aquellos datos inconsistentes en un conjunto de datos compatibles y congruentes, para que puedan ser cargados en el almacén de datos. Estas acciones se llevan a cabo, debido a que pueden existir diferentes fuentes de información, y es vital conciliar un formato y forma única, definiendo estándares, para que todos los datos que ingresarán al almacén de datos estén integrados.

### **Carga**

Esta función es la responsable de cargar la estructura del almacén de datos mediante la transferencia de aquellos datos que han sido transformados y residen en el almacenamiento intermedio, y los datos de los Procesamientos de Transacciones en Línea (OLTP) que tienen correspondencia directa con el depósito de datos. Es preciso tener en cuenta que los datos antes de ser movidos deben de ser analizados con el objetivo de asegurar la calidad.

## **1.4 Inteligencia de negocio**

La inteligencia de negocio (BI- del inglés business intelligence) 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, esto ha provocado

que en la actualidad muchas empresas que trabajan con grandes volúmenes de datos utilicen herramientas de BI.

#### **1.4.1 Definición de BI**

La adquisición y utilización de conocimiento basado en los hechos para mejorar la estrategia del negocio y las ventajas tácticas del mercado. [14]

#### **1.4.2 Características de las herramientas de BI**

- **Accesibilidad a la información:** Los datos son la fuente principal de este concepto. Lo primero que deben garantizar este tipo de herramientas y técnicas será el acceso de los usuarios a los datos con independencia de la procedencia de estos.
- **Apoyo en la toma de decisiones:** Se busca ir más allá en la presentación de la información, de manera que los usuarios tengan acceso a herramientas de análisis que les permitan seleccionar y manipular solo aquellos datos que les interesen.
- **Orientación al usuario final:** Se busca independencia entre los conocimientos técnicos de los usuarios y su capacidad para utilizar estas herramientas.

#### **1.4.3 Beneficios de BI**

Con frecuencia, la transformación y el análisis de la información y los datos que las compañías generan se convierten en un verdadero problema y, por lo tanto, la toma de decisiones se vuelve muy lenta. Las tecnologías de BI intentan revertir esa situación, ayudando a comprender mejor y más rápidamente los datos. De este modo, quienes toman las decisiones pueden hacerlo de una manera más rápida y efectiva para, posteriormente mejorar sus acciones hacia el futuro. El uso apropiado de las herramientas de BI puede marcar la diferencia entre el estancamiento y el crecimiento de las empresas, entre el servicio de mala calidad y el servicio al cliente personalizado, entre la relación óptima con los proveedores y la pérdida de los beneficios de trabajar con ellos y con otros socios de negocios, pueden ayudar a las organizaciones a reducir los ciclos de desarrollo de productos, agilizar operaciones, afinar campañas de marketing y mejorar relaciones con clientes y proveedores. Todo lo anterior significa menores costos y mayores márgenes de utilidad. Sin embargo, es importante tener en cuenta algunos aspectos que pueden derivar en la falla de los proyectos que utilizan la inteligencia de negocio que son detallados a continuación:

- Las empresas fracasan en reconocer que los proyectos de BI son iniciativas de negocios interorganizacionales y que por lo tanto, difieren de las típicas soluciones independientes.

- Falta de compromiso por parte de los sponsors (aquellos que tienen autoridad en la empresa).
- Se tiene poca disponibilidad de los representantes de negocios.
- Hay ausencia de un personal disponible y hábil.
- Existe un mal concepto del software de BI.
- No existe un análisis del negocio detallado.
- No se entiende la necesidad del uso de un metadatos.
- Métodos y herramientas no alienadas.

## 1.5 Modos de almacenamiento

### Modo de almacenamiento OLAP

En el procesamiento analítico en línea u OLAP los datos son clasificados en diferentes dimensiones y pueden ser vistas unas con otras en diferentes combinaciones para obtener diferentes análisis de los datos que contienen. [15] OLAP es un componente clave en los procesos de inteligencia de negocios.

Las principales características de OLAP son:

- **Rápido:** la primera regla se refiere a que el sistema debe ser capaz de responder de una forma rápida y ágil a la información que le sea solicitada por el usuario.
- **Análisis:** significa que el sistema debe poder reflejar cualquier lógica del negocio para poder responder a las preguntas específicas y necesidades empresariales.
- **Compartido:** el sistema deberá proporcionar herramientas que garanticen la confidencialidad de los datos y la seguridad de acceso por perfiles de los usuarios.
- **Multidimensional:** la herramienta deberá proporcionar soporte a cada una de las múltiples jerarquías que puedan existir dentro de la organización de información.
- **Información:** son todos los datos e información derivada de este proceso de análisis, la cual permitirá la toma de decisiones.

Existen diferentes tipos de sistemas OLAP, entre los que se encuentran:

1. MOLAP (Procesamiento Analítico Multidimensional en Línea).
2. ROLAP (Procesamiento Analítico Relacional en Línea).
3. HOLAP (Procesamiento analítico híbrido en línea).

## **ROLAP**

Relational On-line Analytical Process (ROLAP): Son herramientas OLAP que crean vistas multidimensionales extrayendo los datos de bases de datos SQL relacionales. Estas herramientas simulan los datos multidimensionales usando sofisticadas técnicas de indexación, caches y metadatos. [9]

## **MOLAP**

Multidimensional On-line Analytical Process (MOLAP): Son herramientas que acceden a datos que no están almacenados en registros de tablas, sino que almacenan los datos en arreglos de varias dimensiones, llamados cubos. Estos cubos utilizan índices para optimizar el acceso a los datos. [9]

## **HOLAP**

Hybrid Online Analytical Process, procesamiento analítico en línea híbrido (HOLAP): Permite un análisis híbrido de la información, es decir, que une lo mejor de los dos tipos anteriores. El análisis HOLAP ayudará a reducir costos de hardware ya que necesitará menos capacidad de disco que en las bases de datos relacionales.

Entre los sistemas OLAP mencionados anteriormente se utilizará el sistema ROLAP por las ventajas que brinda, entre las que se encuentran:

- ROLAP brinda seguridad e integridad en las bases de datos.
- Es escalable para grandes volúmenes de datos.
- Los datos de los sistemas ROLAP pueden ser compartidos con aplicaciones SQL.

A diferencia de MOLAP Y HOLAP, los datos de ROLAP son accedidos directamente desde el almacén de datos u otra fuente de datos relacional y no son almacenados por separados.

### **1.6 Tipos de modelados de un almacén de datos**

**Esquema estrella:** es aquel que tiene una única tabla de hechos que contiene los datos de análisis, rodeada de las dimensiones. Este aspecto, de tabla de hechos (o central) más grande rodeada de radios o tablas más pequeñas es lo que se asemeja con una estrella. Este esquema es ideal por su simplicidad y velocidad para ser usado en los mercados de datos. Permite acceder tanto a datos agregados como de detalle. Además, permite reducir el número de joins entre tablas (relaciones entre las tablas) y deja a los usuarios establecer jerarquías y niveles entre las dimensiones. Finalmente, es la opción con mejor rendimiento y velocidad pues permite indexar las dimensiones de forma individualizada sin que repercuta en el rendimiento de la base de datos en su conjunto.

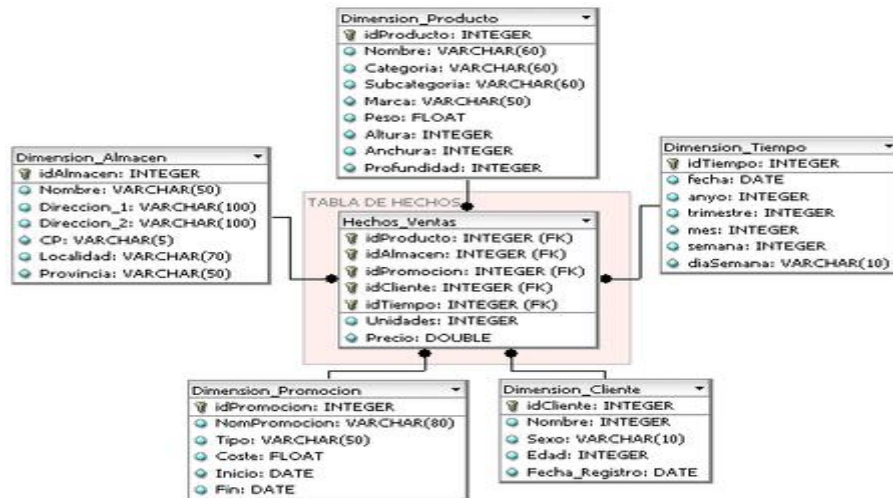


Figura 3: Esquema Estrella.

**Esquema Copo de Nieve:** es una estructura más compleja que el esquema en estrella. Se da cuando las dimensiones se implementan con más de una tabla de datos. Aunque puede reducir espacio por la mínima redundancia de datos, tiene la contrapartida de peores rendimientos al tener que crear más tablas de dimensiones y más joins lo que tiene un impacto directo sobre el rendimiento.

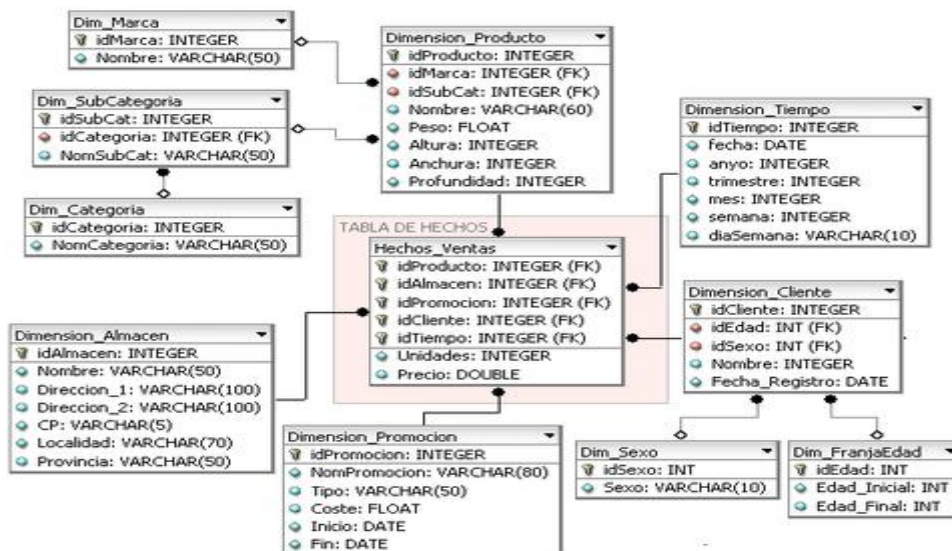


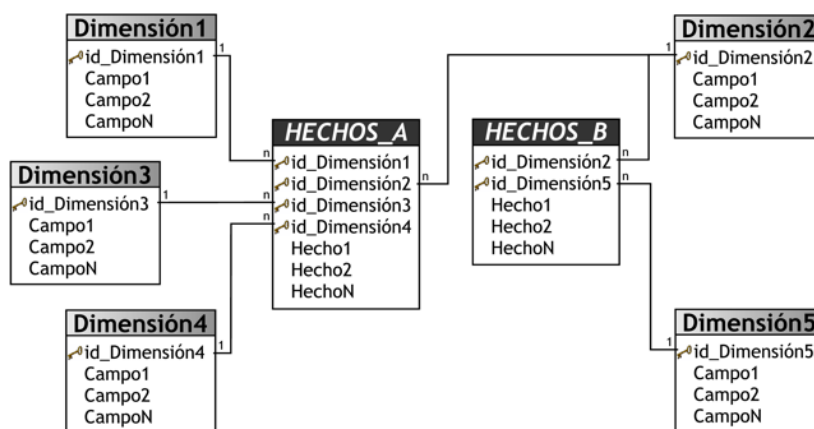
Figura 4: Esquema Copo de Nieve.

**Esquema Constelación de Hechos:** está compuesto por una serie de esquemas de estrella, es decir, una tabla de hechos central con una o más tablas de hechos auxiliares y sus respectivas tablas de dimensiones.



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

- Permite tener más de una tabla de hechos, por lo cual se podrán analizar más aspectos claves del negocio con un mínimo esfuerzo adicional de diseño.
- Contribuye a la reutilización de dimensiones, ya que una misma dimensión puede utilizarse para varias tablas de hechos.



*Figura 5: Esquema Constelación de hechos.*

### 1.7 Metodologías para el desarrollo de los almacenes de datos

Son diversas las metodologías que existen en la actualidad para el desarrollo de los almacenes de datos entre las que se encuentran la metodología de Kimball y Bill Inmon, personas destacadas en el tema de almacenes de datos.

Inmon afirma que la creación de una base de datos relacional con una leve normalización necesita ser la base para los mercados de datos. Por lo que no se crean los mercados de datos directamente desde los Sistemas de Procesamiento de Transacciones en Línea (OLTP) a través de un área de ensayo. En lugar de ello, se crean a partir de la arquitectura relacional de los datos corporativos, a diferencia de la anterior, la propuesta de Kimball se basa en dividir el mundo de inteligencia de negocio entre los hechos y las dimensiones, esta propuesta conduce a una solución completa con un costo menor de tiempo en los procesos de desarrollo, además tiene abundante documentación y se puede consultar dudas a través del sitio web creado por los autores de la metodología. Una de sus características principales es su arquitectura ascendente (bottom-up), plantea que se debe crear el almacén de datos

a partir de la construcción de los mercados de datos centralizados en un área o departamento específico.

### **1.7.1 Fases de la metodología Kimball**

#### **1-Requerimientos y Gestión del Proyecto.**

- Definición del Proyecto.
- Planeación y gestión del Proyecto.
- Definición de los requerimientos de usuario.

#### **2-Arquitectura.**

- Diseño Técnico de la Arquitectura.
- Medidas Tácticas de Seguridad.
- Plan Estratégico de Seguridad.
- Selección e Instalación de Productos.

#### **3-Diseño e Implementación.**

- Análisis Multidimensional (Lógico y físico).
- Análisis de Fuentes de Datos.
- Diseño y Implementación del Área Temporal.
- Popular y Validar Data bases.
- Optimización del Rendimiento.
- Especificación y Desarrollo de Aplicaciones de Usuario Final.

#### **4-Implantación & Operaciones**

- Plan de Implantación.
- Pruebas.
- Implantación (Alpha, Beta & Product Iterations).
- Optimización del Rendimiento.
- Mantenimiento.
- Crecimiento.
- Capacitación y Transferencia Tecnológica.

### **1.7.2 Propuesta de metodología para el desarrollo de Almacenes de Datos en DATEC**

La UCI tiene entre sus principales objetivos desarrollar software con un alto nivel de calidad, para esto ha estandarizado un grupo de aspectos entre los que se encuentra la creación del Expediente de Proyecto creado por el centro de CALISOFT para documentar el proceso de desarrollo de software,

este expediente sigue normas y lineamientos que siguen estándares internacionales de calidad de software como el Modelo de Madurez de la Capacidad Integrado (Capability Maturity Model for Integration) CMMI.

El Centro de Tecnología de Gestión de Datos (DATEC) propone una metodología para este tipo de soluciones, adaptada a su proceso de desarrollo, el cual está basado en Líneas de Productos de Software, y a los lineamientos de calidad exigidos por la UCI, realizando un estudio del Expediente de Proyecto y de cada uno de los artefactos propuestos en el expediente oficial para establecer cuáles eran las adaptaciones que debían realizarse. Esta propuesta también buscaba entre sus retos adaptarse al modo de producción establecido por el departamento DATEC. Esta metodología debía organizar los roles y las actividades por grupos de trabajos estableciendo un balance entre cada grupo y las etapas de desarrollo, de manera que pudieran reutilizarse los recursos humanos en los diferentes proyectos desarrollados por el departamento.

Para definir la metodología de desarrollo a utilizar en el departamento de almacenes de datos de DATEC, se tomó como base la Metodología de Kimball por los siguientes elementos:

- Crea los conceptos de Hechos y Dimensiones, lo que proporciona mayor agilidad en el proceso de desarrollo.
- Propone ir construyendo el almacén de datos a través de la construcción de los mercados de datos departamentales, lo que constituye una buena estrategia y coincide con la división lógica de las empresas, entidades, organismos, etc. Además, permite ir presentando resultados parciales a los clientes en cortos plazos.
- Existe abundante documentación sobre la misma y se puede consultar la web a través de los servicios que brindan el grupo creador de la metodología.

A pesar de todas las ventajas que ofrece la utilización de la Metodología de Kimball, esta no era totalmente adaptable a las características del centro y de la producción en la UCI, por lo que solo se decidió utilizarla como guía en el proceso de confección de metodología de desarrollo del departamento de almacenes de datos. Entre sus principales desventajas se encuentran:

- No tiene definido un criterio que permita estimar los costos de desarrollo de un almacén de datos, basándose en las características de la construcción del mismo.
- Presenta un grupo de roles, pero no explica claramente cuáles son las competencias y responsabilidades de cada uno dentro del proyecto. Por la cantidad de roles que propone se necesita de grupos grandes para su desarrollo.

- Propone un gran número de actividades y artefactos que pueden extender los tiempos de desarrollo si se cuenta con pocos recursos humanos, además no se especifica cómo deben realizarse estos artefactos.
- Está estructurada para el desarrollo de proyectos – productos, donde un proyecto desarrolla un producto determinado.
- No establece el análisis de diferentes criterios de diseño en el levantamiento de requisitos que permita la construcción más adecuada del almacén, teniendo en cuenta las metas de la organización, las necesidades de los usuarios y la disponibilidad de las fuentes operaciones.

Por tales motivos se decidió definir una metodología que permita mitigar las desventajas identificadas en la Metodología de Kimball y ajustada a las condiciones y características de producción de DATEC y de la UCI. Además, la nueva metodología integra las mejores prácticas identificadas en las metodologías estudiadas que permiten un adecuado desarrollo de soluciones de los almacenes de datos denominándose esta: Propuesta de metodología para el desarrollo de almacenes de datos en DATEC.

Con la definición de una metodología de desarrollo para almacenes de datos en el Departamento de Almacenes de Datos de DATEC, se contribuyó a:

- Organizar el proceso de desarrollo; se definieron fases, ciclo de vida, grupos de trabajo, artefactos, actividades, roles, entre otros aspectos que contribuyen positivamente al desarrollo de los proyectos y ajustados al modelo de producción de DATEC y de la UCI.
- Se formaron recursos humanos especializados en cada uno de los grupos de trabajo.
- Se definió un expediente de proyecto con clasificación de artefactos específicos para este tipo de soluciones.
- Se ganó en rapidez, organización y calidad del desarrollo de soluciones de almacenes de datos. [3]

## **1.8 Herramientas utilizadas**

### **1.8.1 Herramienta de modelado**

#### **Visual Paradigm 8.0**

Visual Paradigm for UML (VP-UML) es una herramienta de diseño Unified Modeling Language (UML) y herramienta CASE UML diseñado para ayudar al desarrollo de software. VP-UML soporta los principales estándares de la industria tales como Lenguaje de Modelado Unificado (UML), SysML, BPMN y XMI. Ofrece un completo conjunto de herramientas de los equipos de desarrollo de software

necesario para la captura de requisitos, software de planificación, la planificación de controles, el modelado de clases y el modelado de datos. [5]

#### **Otras características:**

- Soporte de UML versión 2.1.
- Diagramas de Procesos de Negocio - Proceso, Decisión, Actor de negocio, Documento.
- Ingeniería de ida y vuelta.
- Ingeniería inversa - Código a modelo, código a diagrama.
- Generación de código - Modelo a código, diagrama a código.
- Editor de Detalles de Casos de Uso: Entorno todo en uno para la especificación de los detalles de los casos de uso, incluyendo la especificación del modelo general y de las descripciones de los casos de uso.
- Generación de bases de datos: Transformación de diagramas de Entidad-Relación en tablas de base de datos.
- Ingeniería inversa de bases de datos: Desde Sistemas Gestores de Bases de Datos (DBMS) existentes a diagramas de Entidad-Relación.
- Generador de informes para generación de documentación.
- Distribución automática de diagramas: Reorganización de las figuras y conectores de los diagramas UML.
- Importación y exportación de ficheros XML.
- Editor de figuras.

### **1.8.2 Sistema gestor de base de datos**

#### **PostgreSQL 9.1**

PostgreSQL es un sistema de gestión de bases de datos objeto-relacional con más de 15 años de vida, distribuido bajo licencia BSD (Berkeley Software Distribution) y con su código fuente disponible libremente. Es el sistema de gestión de bases de datos de código abierto más potente del mercado y en sus últimas versiones no tiene nada que envidiarle a otras bases de datos comerciales.

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.

A continuación se enumeran las principales características de este gestor de bases de datos:

- Soporta distintos tipos de datos: además del soporte para los tipos base, también soporta datos de tipo fecha, monetarios, elementos gráficos y datos sobre redes (MAC, IP).
- Incorpora una estructura de datos de arreglo.
- Soporta el uso de índices, reglas y vistas.
- Incluye herencia entre tablas, por lo que a este gestor de bases de datos se le incluye entre los gestores objeto-relacionales.
- Permite la gestión de diferentes usuarios, como también los permisos asignados a cada uno de ellos. [7]

Esta versión presenta mejoras como:

- Permite realizar búsquedas más potentes implementando un método de indexado de los vecinos más próximos los cuales mejoran la velocidad.
- Mejora la seguridad.
- Ha añadido las tablas inlogged para incrementar el rendimiento.

### **PgAdmin 1.14.0**

PgAdmin 3 es una herramienta de código abierto para la administración de bases de datos PostgreSQL y derivados (EnterpriseDB Postgres Plus Advanced Server y Greenplum Database).

Incluye:

- Interfaz administrativa gráfica.
- Editor de la sintaxis SQL.
- Editor de código del lado del servidor.

PgAdmin se diseña para responder a las necesidades de la mayoría de los usuarios, desde escribir simples consultas SQL hasta desarrollar bases de datos complejas.

La interface gráfica soporta todas las características de PostgreSQL y hace simple la administración. Está disponible en más de una docena de lenguajes y para varios sistemas operativos, incluyendo Microsoft Windows, Linux, FreeBSD, Mac OSX y Solaris. PgAdmin III soporta versiones de servidores 7.3 y superiores. Versiones anteriores a 7.3 deben usar pgAdmin II. [8]

### **1.8.3 Herramientas para la limpieza de los datos**

#### **Data Cleaner 1.5.3**

DataCleaner es una aplicación de análisis de calidad de los datos y una plataforma de soluciones de DQ (calidad de los datos). Su núcleo es un motor de datos de perfiles de fuerte, que es extensible y por lo tanto, añade la limpieza de datos, las transformaciones, el enriquecimiento, la adecuación y la fusión. [10]

Características:

- Compatibilidad con diferentes tipos de bases de datos: Oracle, MySQL, PostgreSQL, Firebird, SQLite.
- Puede detectar duplicados utilizando la lógica difusa y los pesos y umbrales configurables.
- Escribir datos a bases de datos relacionales, archivos CSV, hojas de cálculo Excel o bases de datos MongoDB.

#### **1.8.4 Herramientas para los procesos de Extracción, Transformación y Carga**

##### **Pentaho Data Integration 4.2.1**

Pentaho Data Integration proporciona la extracción de gran alcance, la transformación y la capacidad de carga (ETL), utilizando un innovador enfoque basado en los metadatos. Con una interfaz intuitiva, gráfica de arrastrar y una arquitectura probada, escalable y basada en estándares, Pentaho Data Integration es cada vez más la elección de las organizaciones tradicionales. [16]

Para el desarrollo del proceso ETL se propone el Pentaho Data Integration debido a que es una herramienta de código abierto adoptado por Pentaho BI, la cual incluye cuatro herramientas:

- Spoon: para diseñar transformaciones ETL usando el entorno gráfico.
- PAN: para ejecutar transformaciones diseñadas con Spoon.
- CHEF: para crear trabajos.
- Kitchen: para ejecutar trabajos.

Características:

- Alto rendimiento.
- Soporte para grandes datos.
- Amplia conectividad. [17]

#### **1.8.5 Herramientas utilizadas en Inteligencia de negocio**

##### **Schema Workbench 3.2.1**

Es una herramienta para el desarrollo del esquema del modelo multidimensional, el cual se guarda en formato XML, desarrollado en Java. Este programa fue publicado en el 2007, entrega todas las facilidades para poder realizar el modelo lógico del cubo OLAP al cual se le realizarán las consultas.

Este programa se conecta directamente con la BD, para así poder diseñar los cubos OLAP que se requieren, para que el usuario final pueda visualizar los indicadores en el mercado de datos. [11]

### **Mondrian OLAP Server 3.2.1**

Mondrian es una de las aplicaciones más importantes de la plataforma Pentaho BI. Mondrian es un servidor OLAP open source (código abierto) que gestiona comunicación entre una aplicación OLAP (escrita en Java) y la base de datos con los datos fuente.

Es decir, Mondrian actúa como “JDBC para OLAP”. [12]

Entre sus principales características se encuentran:

- Agilizar la consulta de grandes cantidades de datos.
- Alta velocidad de respuesta.
- Permite realizar consultas al mercado de datos.
- Soporta los APIs: Java OLAP (JOLAP) y XML para el análisis de aplicaciones programadas.

### **Pentaho BI Server 3.8.0**

La plataforma Pentaho BI proporciona la arquitectura y la infraestructura necesaria para crear soluciones de BI. El marco de trabajo proporciona los servicios básicos, incluyendo la autenticación, registro, auditoría, servicios web y motores de reglas. La plataforma también incluye un motor de solución que integra información, análisis, cuadros de mando y los componentes de minería de datos. El diseño modular y arquitectura de plugin permite que todo o parte de la plataforma pueda ser embebido en aplicaciones de terceros por los usuarios finales, así como fabricantes de equipos originales. [18]

La aplicación más conocida de la Plataforma de BI es el servidor de BI de Pentaho, que funciona como una red basada en sistema de gestión de informes, integración de aplicaciones y servidor de motor de flujo de trabajo ligero (secuencias de acción) Está diseñado para integrarse fácilmente en los procesos de negocio. Esta versión trae mejoras como:

- Permite como administrador cambiar los temas de diseño.
- Permite crear una nueva conexión o editarla sin necesidad de abrir la consola de administración.
- Proporciona una forma genérica para los plugins para registrar las secuencias de comandos y CCS para la inclusión en otras aéreas de plataforma.

Algunas de sus ventajas son:



- Integración con procesos de negocio.
- Administra y programa reportes.
- Administra seguridad de usuarios.

### **Apache Tomcat 6.0.29**

Apache Tomcat (también llamado Jakarta Tomcat o simplemente Tomcat) funciona como un contenedor de servlets desarrollado bajo el proyecto Jakarta en la Apache Software Foundation. Tomcat implementa las especificaciones de los servlets y de JavaServer Pages (JSP) de Sun Microsystems. Es un servidor web desarrollado en Java, presenta las siguientes características:

1. Software libre.
2. Puede funcionar como servidor HTTP o conectado a otro servidor HTTP como Apache HTTP Server o IIS.
3. Puede ejecutar servicios web mediante Apache Axis.

### **Conclusiones**

Tomando como punto de partida la fundamentación teórica y por la importancia que tiene para la realización de una correcta investigación sobre los AD y MD en el capítulo se arribaron a las siguientes conclusiones:

- Se fundamentaron los principales conceptos relacionados con los AD y los MD permitiendo obtener un mayor conocimiento sobre los AD en general.
- Se seleccionó como metodología la propuesta por DATEC, la cual permitió guiar y agilizar el desarrollo de la solución.
- Durante la investigación detallada de las tecnologías informáticas existentes referente a los almacenes de datos, se decidió utilizar las siguientes herramientas:
  - ✓ Visual Paradigm 8.0
  - ✓ PostgreSQL 9.1
  - ✓ Data Cleaner 1.5.3
  - ✓ Pentaho Data Integration 4.2.1
  - ✓ Schema Workbench 3.2.0
  - ✓ Mondrian OLAP Server 3.2.1
  - ✓ Pentaho BI Server 3.8.0
  - ✓ Apache Tomcat 5.5

## CAPÍTULO 2: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS

### Introducción

En este capítulo se describe el negocio, se realiza el análisis y diseño del MD, se identifican los requerimientos del sistema, las necesidades de los usuarios y las reglas del negocio identificadas. Se especifican los actores, se describen los CU y además se definen las dimensiones y el hecho identificado para el MD. Se diseña la arquitectura del sistema y se definen los permisos que le corresponden a cada uno de los actores que intervienen en el sistema.

### 2.1 Análisis de la solución

#### 2.1.1 Definición del negocio

La ONEI tiene como objetivo fundamental capturar, analizar y difundir las cifras económicas y sociales que son recogidas a lo largo y ancho del país. La ONEI está estructurada por 22 áreas que presentan sucursales en distintas zonas y lugares en todo el país, una de estas áreas es la de Servicios de comercio exterior, que se encarga de recoger la información de una serie de indicadores fundamentales con información relacionada con el comercio exterior de servicios del país con el resto del mundo desglosado por productos y socio comercial. Todo esto genera un gran cúmulo de informaciones que se recogen en el modelo estadístico 0793 que se genera de manera trimestral.

#### 2.1.2 Tema de análisis identificado

Dada el área de trabajo se definió el tema de análisis que permite agrupar las principales necesidades de información. Esto posibilita determinar una organización global de los datos y enfocar la investigación en dominios informativos. Para la construcción de la propuesta se definió como tema de análisis: los indicadores de servicios del comercio exterior.

#### 2.1.3 Reglas del negocio

Las reglas de negocios (o las directivas empresariales) definen y controlan la estructura, el funcionamiento y la estrategia de una organización. Las reglas de negocios pueden estar formalmente definidas en manuales de procedimiento, contratos o acuerdos, o bien pueden existir como conocimiento o experiencia que tienen los empleados. Las reglas de negocios son dinámicas, están sujetas a cambios en el tiempo y pueden encontrarse en todo tipo de aplicaciones. A continuación se muestran las reglas del negocio identificadas para el mercado de datos Servicios de comercio exterior:

**RN 1.** Los identificadores de los indicadores no pueden ser nulos.

**RN 2.** El código de los atributos en cada una de las dimensiones no puede tomar valores repetidos.

**RN 3.** El código de los indicadores son los dos primeros números.

**RN 4.** El código de la descripción de los indicadores son los dos últimos números.

**RN 5.** El código de los servicios solo pueden ser números.

**RN 6.** El nombre de los servicios solo pueden ser letras.

#### **2.1.4 Necesidades de los usuarios**

Cuando se realiza un producto informático es de gran importancia conocer lo que necesitan los usuarios, pues de este conocimiento se logrará el éxito para determinar si el producto es satisfactorio o no y si da respuesta a las exigencias y necesidades planteadas por el cliente. Luego de un análisis que tiene lugar en el área de Servicios de comercio exterior perteneciente a la ONEI se determinó que las necesidades de los clientes estaban enmarcadas en el tema de los indicadores de comercio exterior.

#### **2.1.5 Requerimientos**

Un requerimiento es una condición o capacidad a la que el sistema siendo construido debe conformar. A continuación se muestran los requisitos de información identificados para el mercado de datos Servicios de comercio exterior. [4]

#### **Requisitos de información identificados**

- **RI 1** Obtener el valor de las cuentas por cobrar al inicio del año por servicio, tiempo, país, entidad y tipo de indicador.
- **RI 2** Obtener el valor de los servicios facturados de ingresos por servicio, tiempo, país, entidad y tipo de indicador.
- **RI 3** Obtener el valor de los cobros realizados por servicio, tiempo, país, entidad y tipo de indicador.
- **RI 4** Obtener el valor de los cobros realizados por servicio, tiempo, país, entidad y tipo de indicador.
- **RI 5** Obtener el valor de los servicios donados y compensados entregados por servicio, tiempo, país, entidad y tipo de indicador.
- **RI 6** Obtener el valor de las cuentas por pagar al inicio del año por servicio, tiempo, país, entidad y tipo de indicador.
- **RI 7** Obtener el valor de los servicios facturados de egresos por servicio, tiempo, país, entidad y tipo de indicador.

- **RI 8** Obtener el valor de los pagos realizados por servicio, tiempo, país, entidad y tipo de indicador.
- **RI 9** Obtener el valor de las cuentas por pagar al final por servicio, tiempo, país, entidad y tipo de indicador.
- **RI 10** Obtener el valor de los servicios donados y compensados por servicio, tiempo, país, entidad y tipo de indicador.

## Requisitos funcionales

*Tabla 1: Requisitos funcionales.*

|                                   |   |
|-----------------------------------|---|
| <b>RF 1</b> - Autenticar usuario. | <b>RF 9</b> - Extraer información.              |
| <b>RF 2</b> - Adicionar roles.    | <b>RF 10</b> - Realizar transformación y carga. |
| <b>RF 3</b> - Eliminar roles.     | <b>RF 11</b> - Abrir navegador OLAP.            |
| <b>RF 4</b> - Adicionar usuarios. | <b>RF 12</b> - Mostrar editor MDX.              |
| <b>RF 5</b> - Eliminar usuarios.  | <b>RF 13</b> - Mostrar Padres.                  |
| <b>RF 6</b> - Insertar reportes.  | <b>RF 14</b> - Ocultar repeticiones.            |
| <b>RF 7</b> - Modificar reportes. | <b>RF 15</b> - Intercambiar ejes.               |
| <b>RF 8</b> - Eliminar reportes.  | <b>RF 16</b> - Mostrar gráfico.                 |

## Requisitos no funcionales

Los requerimientos no funcionales hacen relación a las características del sistema que se aplican de manera general en conjunto, más que a rasgos particulares del mismo. Estos requerimientos son adicionales a los requerimientos funcionales que debe cumplir el sistema, y corresponden a aspectos tales como la disponibilidad, mantenibilidad, flexibilidad, seguridad y facilidad de uso entre otros. Para

ver los restantes requisitos no funcionales propuestos para el mercado de datos Servicios de comercio exterior remitirse al artefacto Especificación de Requisitos de Software.

### **Requerimientos legales y de derecho de autor**

**RNF16.** Entregar el sistema a la ONEI.

El sistema debe ser transferido a la ONEI mediante un proceso de transferencia una vez que esté en explotación, incluyendo el código fuente y la documentación correspondiente.

**RNF17.** Requerimientos legales y de derecho de autor.

Se firmará un acuerdo de colaboración entre la UCI y la ONEI donde quedarán establecidas todas las pautas que se deben cumplir durante el desarrollo de la solución. Se reconoce la UCI con derecho de autor y se firmará un proyecto técnico por ambas partes que recoge entre otros elementos, el cronograma de ejecución del producto a desarrollar.

#### **2.1.6 Actores del sistema**

Los actores representan los usuarios del sistema u otras aplicaciones que interactúan con él, es decir, representan terceros fuera del sistema que se relacionan con los casos de uso.

El sistema contará con tres actores:

1. Especialista que se encargará de analizar las necesidades de información del sistema.
2. Administrador que se encargará de gestionar los roles, permisos y reportes además de tener las mismas funciones que el especialista.
3. Administrador de ETL que se encargará de realizar todas las funciones de los procesos de extracción, transformación y carga (ETL).

#### **2.1.7 Diagrama de casos de usos del sistema**

Para la confección del diagrama de caso de uso se agruparon los 25 requisitos funcionales del sistema en casos de uso, y se definieron las relaciones existentes entre ellos y los actores del sistema.

### **Diagrama de Casos de Uso del Sistema**

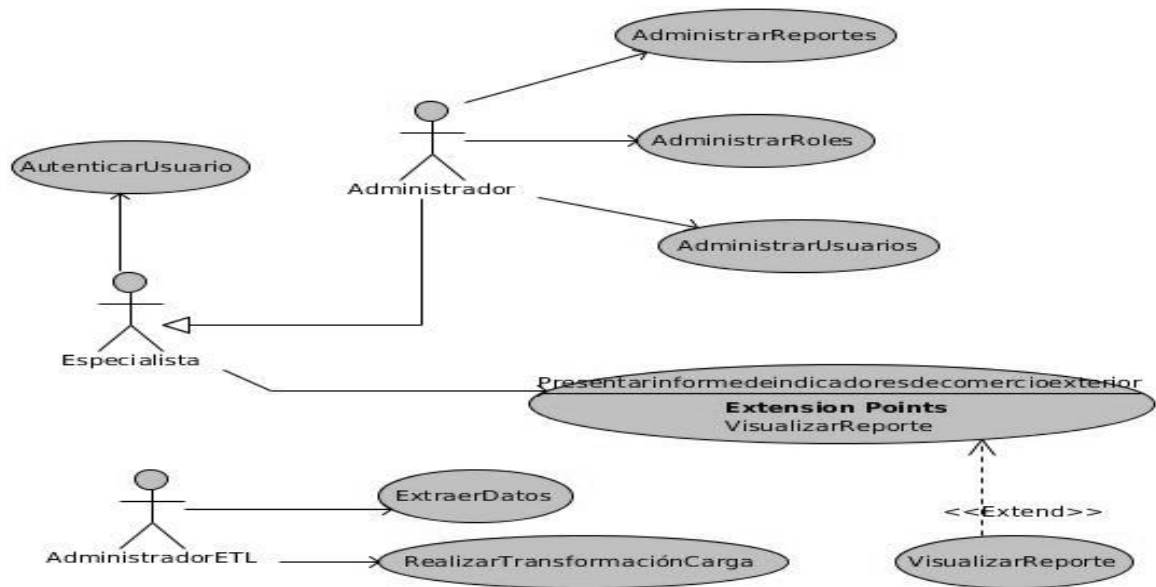


Figura 6: Diagrama de Casos de Uso.

### 2.1.8 Descripción textual de los casos de uso

A continuación se muestra la descripción textual del caso de uso de información: Presentar información de indicadores de comercio exterior, para ver las restantes descripciones de los otros casos de uso que pertenecen al mercado de datos Servicios de comercio exterior remitirse al artefacto Especificación de caso de uso.

#### CUI 1: Presentar informe de indicadores de comercio exterior.

|                       |  |
|-----------------------|--|
| <b>Objetivo</b>       | Presentar información relacionada con los servicios.   |
| <b>Actores</b>        | Especialista, Administrador.   |
| <b>Resumen</b>        | El caso de uso inicia cuando el actor desea hacer un análisis de la información relacionada con los indicadores relacionados con los servicios. El actor selecciona el reporte que desea ver, el sistema muestra la información contenida en él y las opciones de los posibles cambios que le puede hacer al reporte. El caso de uso finaliza cuando el actor termina el análisis de la información relacionada con los servicios. |
| <b>Complejidad</b>    | Alta   |
| <b>Prioridad</b>      | Media  |
| <b>Precondiciones</b> | El usuario se autenticó correctamente.   |

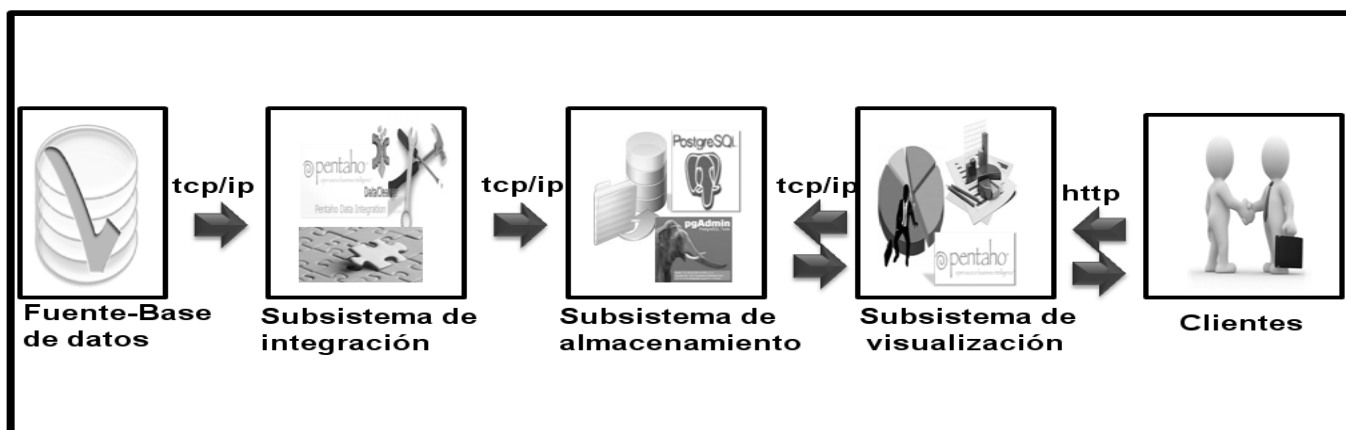
|  |  |   |
|--|--|---|
|  | Los datos correspondientes fueron cargados en el mercado de datos.<br>Los reportes relacionados con los visitantes fueron creados. |   |
| <b>Postcondiciones</b>   | Los reportes correspondientes al caso de uso fueron consultados.   |   |
| <b>Flujo de eventos</b>  |  |   |
| <b>Flujo básico Presentar información relacionada con los servicios</b>                                |  |   |
|  | <b>Actor</b>   | <b>Sistema</b>  |
| 1.   | Selecciona el área de análisis general A.A.G SIGOB.  |   |
| 2.   |  | Muestra las áreas de análisis contenidas en el A.A.G SIGOB.   |
| 3.   | Selecciona el área de análisis A.A. Servicios de comercio exterior.  |   |
| 4.   |  | Muestra el libro de trabajo contenido en el A.A. Servicios de comercio exterior.  |
| 5.   | Selecciona el libro de trabajo L.T Indicadores de comercio exterior.   |   |
| 6.   |  | Muestra los reportes contenidos en el L.T Indicadores de comercio exterior.   |
| 7.   | Selecciona el reporte que desea analizar.  |   |
| 8.   |  | Muestra la información contenida en el reporte seleccionado y brinda opciones al actor para hacer cambios al reporte durante su análisis. Ir al CU Visualizar reportes. Finaliza el CU. |
| <b>Flujos alternos</b>   |  |   |
|  |  |   |
| <b>Opciones de reportes de Presentar información relacionada con los servicios</b>                     |  |   |
| <b>Perspectivas de análisis</b>  | <b>Posibles resultados</b>   |   |
|  | <b>Medidas</b>   | <b>Periodicidad</b>   |
| Variables de entrada relacionadas con el CU Presentar información de indicadores de comercio exterior: | Variables de salida disponibles en el hecho<br>Viviendas:  | Rango de tiempo en que se solicitan las variables de  |

|   |   |  |
|---|---|--|
| <ul style="list-style-type: none"> <li>• Temporal_trimestre</li> <li>• País</li> <li>• Entidad</li> <li>• Servicio</li> <li>• Indicador_servicio</li> </ul> | <ul style="list-style-type: none"> <li>• Valor_servicio</li> </ul>  | salida: <ul style="list-style-type: none"> <li>• Trimestral</li> </ul>   |
| <b>Prototipo de interfaz de usuario:</b>  |   |  |
| <b>Relaciones</b>   | <b>CU Incluidos</b>   | No aplica.   |
|   | <b>CU Extendidos</b>  | Visualizar reporte: Paso 10 del Flujo Básico.<br>Visualizar reporte en el CUI Presentar información de indicadores de comercio exterior. |
| <b>Requisitos no funcionales</b>  | RNF1, RNF2, RNF3, RNF4, RNF5, RNF6, RNF7, RNF8, RNF9, RNF10, RNF11, RNF12, RNF13, RNF14, RNF15, RNF16, RNF17. |  |
| <b>Asuntos pendientes</b>   | [Posibles mejoras al caso de uso.]  |  |

## 2.2 Arquitectura del mercado de datos

El desarrollo de un sistema de software requiere que este sea visto desde diferentes perspectivas. Diferentes usuarios (usuario final, analistas, desarrolladores, integradores y jefes de proyecto) siguen diferentes actividades en diferentes momentos del ciclo de vida del proyecto, lo que da lugar a las diferentes vistas del proyecto, dependiendo de qué interés más en cada instante de tiempo. La arquitectura es el conjunto de decisiones significativas sobre la organización del sistema, selección de elementos estructurales y sus interfaces a través de los cuales se constituye el sistema, composición de los elementos estructurales y de comportamiento en subsistemas. A continuación se muestra en la figura 7 los diferentes subsistemas que conforman la arquitectura del mercado de datos Servicios de comercio exterior.





*Figura 7: Diseño de la arquitectura.*

## 2.3 Diseño

La arquitectura del mercado de datos Servicios de comercio exterior va a estar conformada por tres subsistemas y el sistema fuente los cuales desempeñan responsabilidades independientes.

Subsistema de almacenamiento: en este subsistema están contenidas las dimensiones, hechos y medidas identificadas en el mercado de datos.

Subsistema de integración: es el subsistema encargado de integrar, estandarizar y limpiar la información que se va a extraer de las fuentes.

Subsistema de visualización: es el encargado de consultar la información contenida. Las funcionalidades de dicho subsistema están encaminadas a generar reportes, gráficos en diferentes perspectivas, vistas de análisis, con el objetivo de arribar a conclusiones correctas con mayor prontitud.

### 2.3.1 Subsistema de almacenamiento

#### Dimensiones, hechos y medidas identificadas

Se definieron 6 dimensiones:

- Entidad (dim\_entidad): recoge la información relacionada con todas las entidades de Cuba.
- Temporal Trimestre (dim\_temporal\_trimestre): almacena la periodicidad con la que se recoge la información contenida en el modelo 0793.
- Servicio (dim\_servicio): almacena el valor asignado a los servicios prestados/recibidos de manera gratuita mediante donaciones otorgadas y/o el valor de los servicios prestados/recibidos mediante trueque, conocidos como compensados.

- País (dim\_pais): recoge toda la información de todos los países que solicitan o presentan un servicio.
- Indicador (dim\_tipo\_indicador): recoge toda la información relacionada con las Exportaciones (Ingresos)/Importaciones (Egresos) de los servicios que presta/recibe un servicio a Unidad de una Unidad no Residente.

### Tabla de hecho

Al analizar el modelo 0793 se definió una tabla de hecho:

Servicios de comercio exterior (hech\_serv\_com\_ext): almacena la información de los indicadores de los servicios de comercio exterior recogida en el modelo 0793.

### Medidas

- valor\_servicio: se recogen todos los valores de los indicadores de los servicios.

### Matriz bus

La matriz bus es la representación de la relación que existe entre las dimensiones y los hechos, cuyo objetivo es crear un nivel de gran abstracción y planificación visionaria en la arquitectura del almacén de datos que permitirá evitar el solapamiento entre los hechos.

*Tabla 2: Matriz Bus*

| Dimensiones            | Hechos            |
|------------------------|-------------------|
|                        | hech_serv_com_ext |
| dim_entidad            | x                 |
| dim_temporal_trimestre | x                 |
| dim_temporal_mes       | x                 |
| dim_pais               | x                 |
| dim_tipo_indicador     | x                 |

### Modelo de datos

Un modelo de datos es un lenguaje orientado a describir una base de datos. Típicamente un modelo de datos permite describir:

- Las estructuras de datos de la base: El tipo de los datos que hay en la base y la forma en que se relacionan.
- Las restricciones de integridad: Un conjunto de condiciones que deben cumplir los datos para reflejar correctamente la realidad deseada.

- Operaciones de manipulación de los datos: típicamente, operaciones de agregado, borrado, modificación y recuperación de los datos de la base. [2]

Una vez diseñado el modelo de datos con las dimensiones y los hechos correspondientes al mercado de datos Servicios de comercio exterior se tomó como modelo dimensional el modelo estrella, por las características ya mencionadas en el capítulo anterior.

### Especificación del modelo físico

El modelo físico está conformado por seis tablas dentro de las cuales cuatro son dimensiones y una es una tabla de hechos.

Para la creación de las llaves primarias se utilizó el estándar de definición `dim_dimension_id`. El tipo de datos de la llave primaria de cada tabla es entero, siendo este no nulo. Para una mejor distribución de las tablas se definieron dos esquemas:

- **dimensiones:** en este esquema se encuentran las dimensiones comunes para todo el almacén.
- **hech\_serv\_com\_ext:** este esquema recoge el hecho, las dimensiones propias identificadas para la construcción del mercado de datos Servicios de comercio exterior y las tablas de metadatos propias también del mercado de datos.

### Descripción de la estructura física

*Tabla 3: Descripción de la estructura física.*

| Número | Nombre de la tabla     | Descripción                           | Llave Primaria         |
|--------|------------------------|---------------------------------------|------------------------|
| 1      | hech_serv_com_ext      | Hecho servicios de comercio exterior. | hech_serv_com_ext_id   |
| 2      | dim_entidad            | Dimensión entidad.                    | dim_entidad_id         |
| 3      | dim_servicio           | Dimensión servicio                    | dim_servicio_id        |
| 4      | dim_temporal_trimestre | Dimension temporal trimestre          | dim_temporal_trimestre |
| 5      | dim_indicador          | Dimension indicador                   | dim_indicador_id       |
| 6      | dim_país               | Dimensión país                        | dim_país_id            |

## Modelo físico

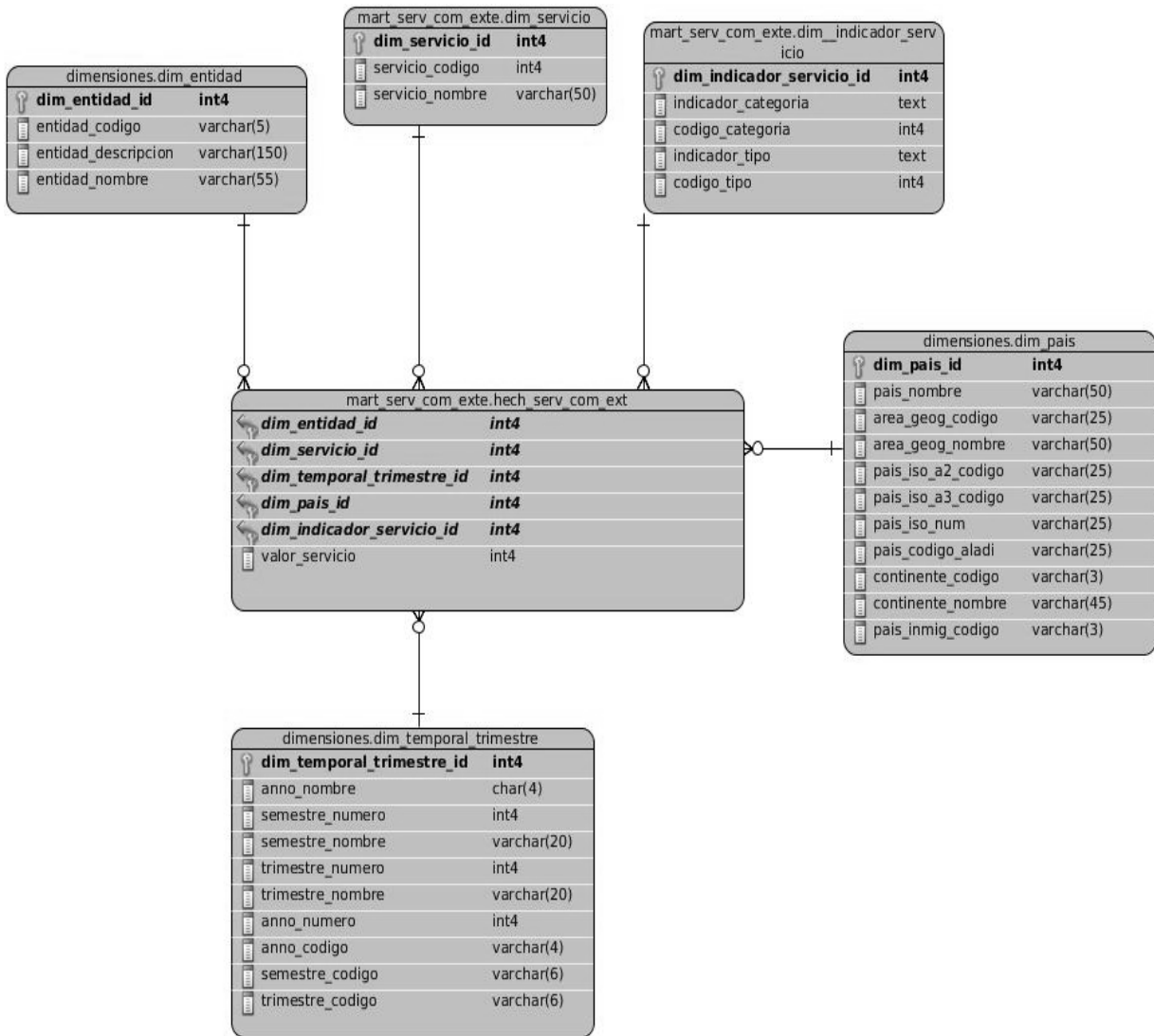


Figura 8: Modelo físico.

### 2.3.2 Subsistema de integración

#### Diseño de las transformaciones

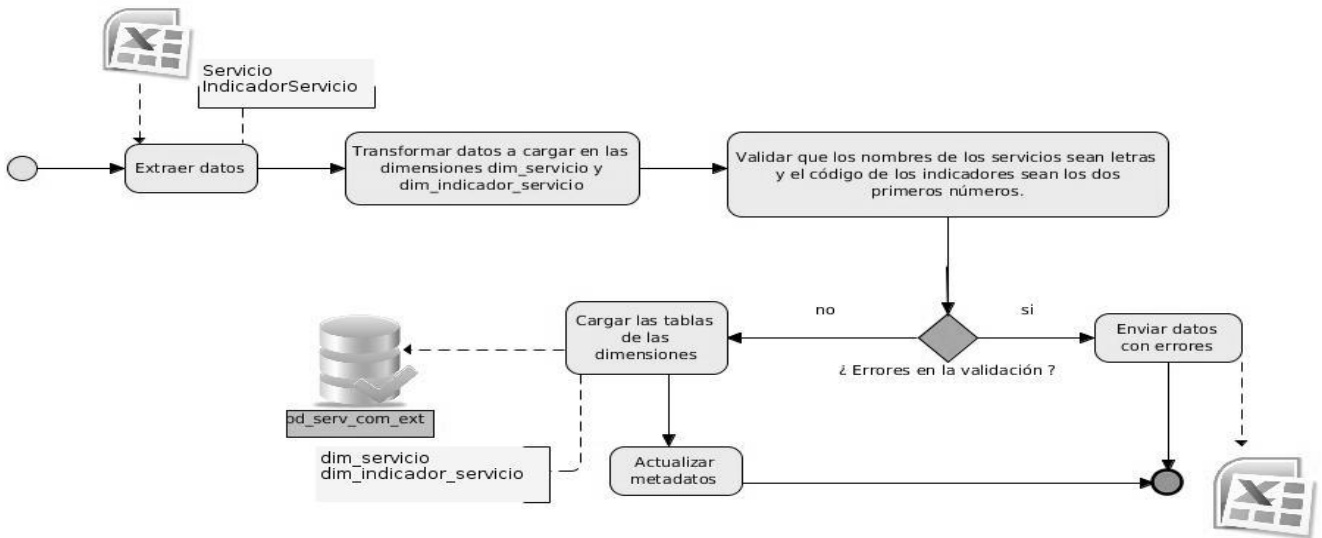


Figura 9: Diseño del subsistema de integración para las dimensiones.

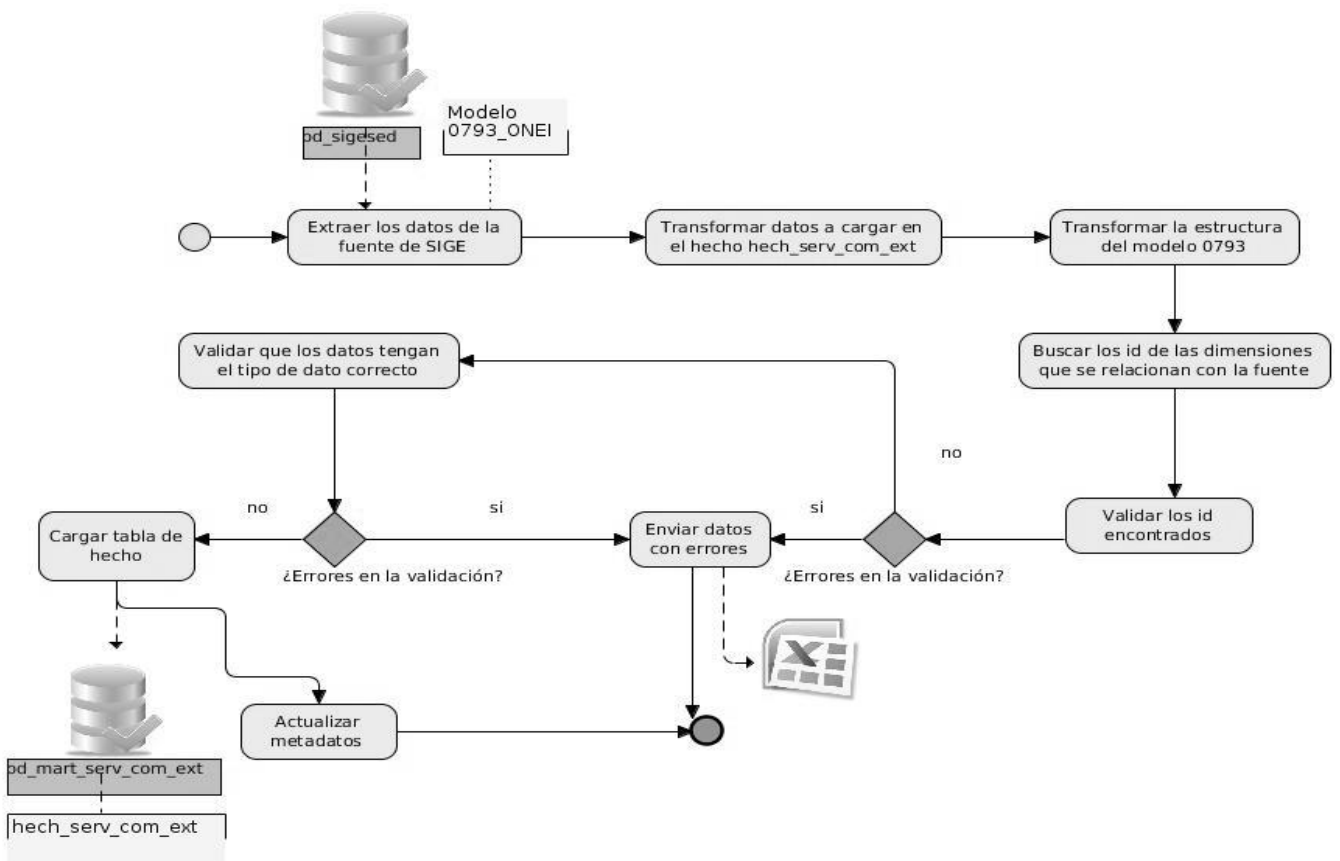
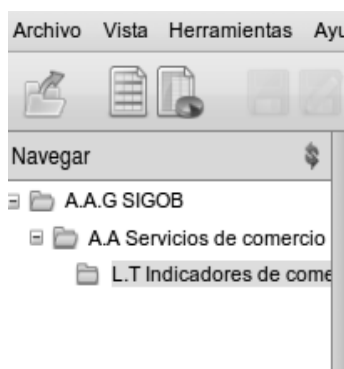


Figura 10: Diseño del subsistema de integración para el hecho.

### 2.3.3 Subsistema de visualización

En correspondencia con los requisitos de información identificados para el área de Servicios de comercio exterior, se definieron 12 reportes agrupados en un libro de trabajo (LT) ubicados dentro del área de análisis (AA) Servicios de comercio exterior, dicha área se corresponde con una sección del almacén de datos del proyecto SIGOB, mientras que el libro de trabajo se definió en correspondencia con la tabla de hecho.

A continuación se detallan los elementos que componen las estructuras de navegación de la información, los mismos estarán organizados de la siguiente manera: un área de análisis (AA) específica para el área de Servicios de comercio exterior y un libro de trabajo (L.T) incluido en esta área de análisis.



*Figura 11: Mapa de navegación.*

#### Descripción del Área General

- ✓ **A.A.G SIGOB:** Agrupa toda la información referente a todos los mercados de datos de la ONEI que forman el almacén SIGOB.

#### Descripción del Área de Análisis

- ✓ **A.A Servicios de comercio exterior:** Agrupa toda la información referente a los indicadores del área de Servicios de comercio exterior y contiene los libros de trabajo donde se encuentran los reportes.

#### Descripción del Libro de Trabajo

- ✓ **L.T. Indicadores de comercio exterior:** Agrupa todos los reportes relacionados con los servicios, como por ejemplo cantidad de ingresos por servicio.

### 2.3.4 Roles y permisos

Un sistema que involucra a todo o gran parte del personal, requiere de un control estricto de a qué información se quiere dar acceso y cuál es el grado de permisos que se puede asignar a cada empleado. Diferentes roles permiten asignar a cada usuario que pueden hacer con la información, sólo consultarla, darla de alta o modificarla, visualizar los datos de carácter personal o que estos permanezcan ocultos. Cada rol está formado por un conjunto de acciones que se pueden personalizar, variando los tipos de acciones que puede realizar cada uno. Para el acceso al mercado de datos Servicios de comercio exterior se define un usuario por cada uno de los roles existentes en el sistema, con el objetivo de garantizar un control de acceso basado en roles y así cada usuario opera en el sistema según los permisos que se le definan al rol.

*Tabla 4: Roles y permisos.*

| Roles                | Permisos |           |
|----------------------|----------|-----------|
|                      | Lectura  | Escritura |
| <b>Base de datos</b> |          |           |
| Administrador        | X        |           |
| Administrador ETL    | X        | X         |
| Especialista         | X        |           |
| <b>Aplicación</b>    |          |           |
| Administrador        | X        | X         |
| Especialista         | X        |           |
| Administrador ETL    |          |           |

### Conclusiones del capítulo

En el presente capítulo se realizó el análisis y diseño del mercado de datos Servicios de comercio exterior donde después de realizadas las entrevistas con los clientes se realizó un estudio del área perteneciente al mercado de datos lo cual permitió identificar los requerimientos necesarios para conocer las necesidades de los usuarios, todo esto permitió diseñar y analizar la solución centrándose en las necesidades del cliente, identificándose 10 requisitos de información, 25 requisitos funcionales y 17 requisitos no funcionales.

Se diseñó el modelo de casos de uso del sistema donde se identificaron los actores y casos de uso, así como los roles y permisos.

Se realizó el diseño del modelo de datos permitiendo determinar las distintas relaciones entre las dimensiones y los hechos identificados.

Se diseñó la arquitectura del sistema con cada uno de sus subsistemas permitiendo seleccionar los elementos estructurales y sus interfaces a través de los cuales se constituye el sistema, composición de los elementos estructurales y de comportamiento en subsistemas.



## CAPÍTULO 3: IMPLEMENTACIÓN DEL MERCADO DE DATOS

### Introducción

En este capítulo se realizará la implementación de la solución propuesta, se implementará el modelo de datos anteriormente diseñado, se realizarán los procesos de ETL y se desarrollará el subsistema de visualización para gestionar los reportes candidatos necesarios dándole solución a los requisitos de información planteados por el cliente.

### 3.1 Implementación del subsistema de almacenamiento

Durante la implementación del modelo de datos se tuvo en cuenta todo lo especificado anteriormente en el modelo físico, los tipos de datos de las variables, los esquemas que agrupan las dimensiones y el hecho del mercado de datos.

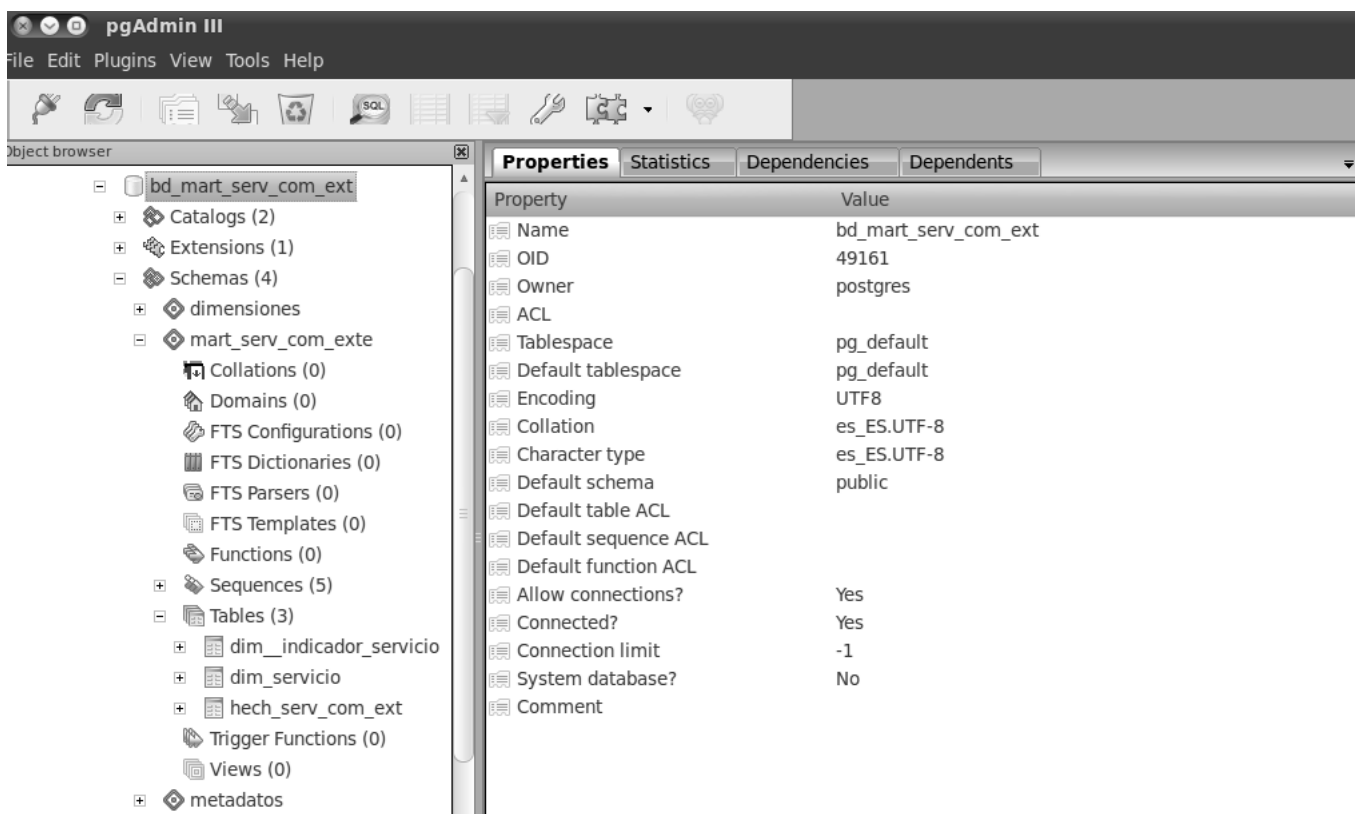


Figura 12: Estructura física de la base de datos.

### 3.2 Implementación del subsistema de integración de los datos

Antes de comenzar la integración de datos, es necesario analizar y estudiar previamente la fuente de datos de los cuales se extraerán los datos necesarios, este estudio permitirá realizar las transformaciones para facilitar el proceso de integración de los datos, es aquí donde desempeñan un papel fundamental el uso de las tecnologías disponibles para la integración de los datos.

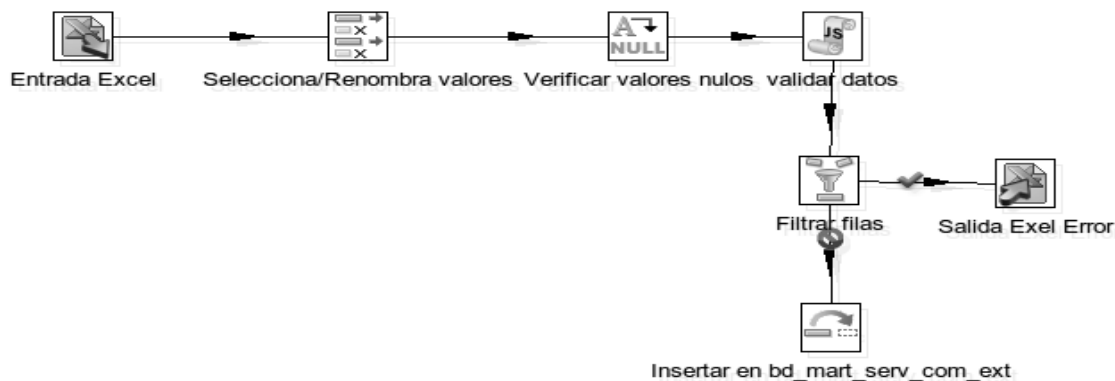
### 3.2.1 Implementación de los procesos de ETL

#### Extracción de los datos

Es en esta fase donde se extraen los datos necesarios de las fuentes existentes basándose en las necesidades y requisitos del usuario. Los datos correspondientes al área de Servicios de comercio exterior se encuentran en diferentes formatos. Los datos fueron estandarizados para iniciar el proceso de transformación, en el caso de las dimensiones se tomó como fuente los clasificadores establecidos por dicha entidad y para el hecho se utilizaron los datos referentes al tema de análisis consultados por los especialistas y el cliente.

#### Transformación y carga de los datos

Este proceso de ETL tiene como objetivo extraer los datos, transformarlo en función de reglas de negocio, a través de un proceso de homogeneización de los tipos y formatos de datos. Y finalmente cargar los datos ya convertidos para el mercado de datos.



*Figura 13: Transformación para cargar la dimensión dim\_servicio.*

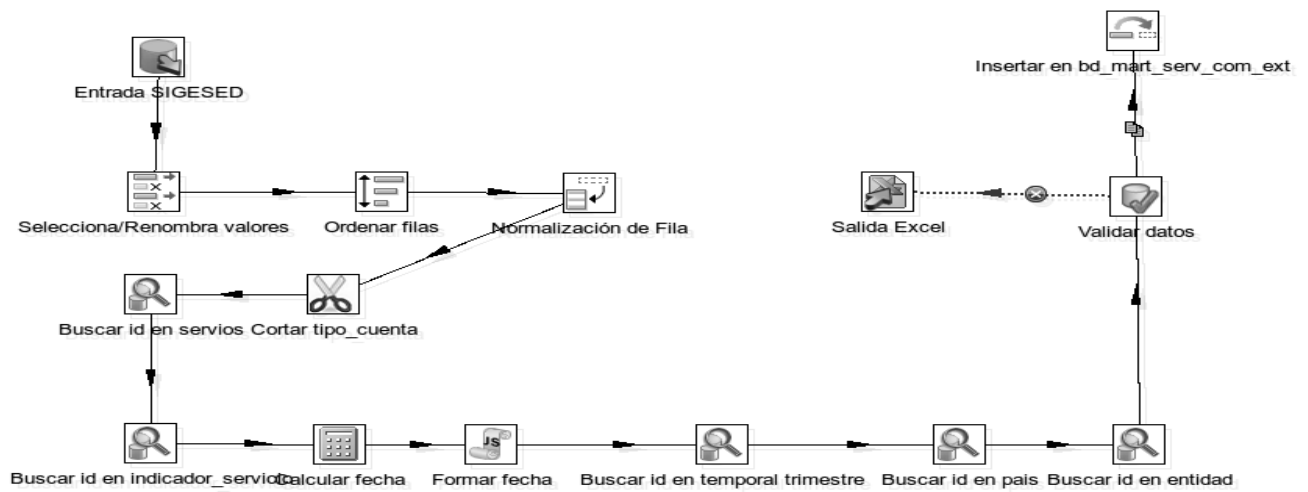


Figura 14: Transformación para cargar el hecho hech\_serv\_com\_ext.

### 3.3 Implementación del subsistema de visualización

#### 3.3.1 Cubos OLAP

El diseño y la implementación del cubo OLAP se realizó en la herramienta Pentaho Schema Workbench, esta herramienta permite generar un fichero xml con las dimensiones con sus niveles de jerarquía y la medida definida anteriormente.

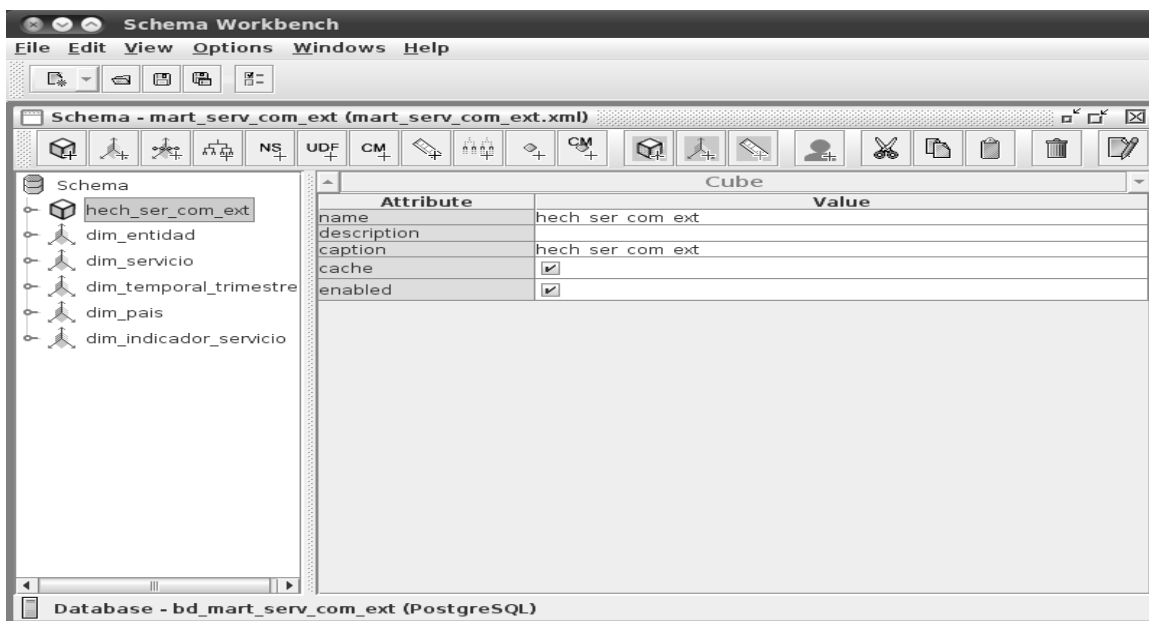


Figura 15: Diseño general del esquema.

### 3.3.2 Creación y administración de los reportes

Los reportes representan la información que el cliente desea que se muestre como finalidad del producto según sus necesidades e intereses. Fueron seleccionados luego de realizar un análisis de los modelos antes descritos, donde se recoge toda la información referente al área de Servicios de comercio exterior. La visualización de los reportes se realizará a través de consultas MDX que serán administradas por la herramienta Pentaho BI Server. Se crearon 12 reportes en total. En la siguiente figura se muestra un reporte del hecho hech\_serv\_com\_ext desarrollado en la herramienta antes mencionada.

| Cantidad de egreso..                |                                     |                      |                  |                            |   |
|-------------------------------------|-------------------------------------|----------------------|------------------|----------------------------|---|
| Fecha                               |                                     |                      |                  |                            |   |
| + Todas                             |                                     |                      |                  |                            |   |
| Indicadores                         |                                     |                      |                  |                            |   |
| Entidad                             | Cuentas por pagar al inicio del año | Servicios facturados | Pagos realizados | Cuentas por pagar al final | Servicios donados y compensados recibidos |
| - Todas                             | \$23.344                            | \$38.948             | \$27.409         | \$35.611                   | \$35.12                                   |
| E.GRABACIONES Y EDICIONES MUSICALES | \$858                               | \$200                | \$4.259          | \$375                      | \$35                                      |
| EMP.COMER.RTV                       | \$24                                | \$91                 | \$1.844          | \$520                      | \$15                                      |
| E.CONTRATA.ASISTEN.TECNIC A         | \$2.476                             | \$2.927              | \$1.829          | \$9.326                    | \$21.71                                   |
| E.EXPORT.CIENCIA Y TEC.AGROP        | \$3.955                             | \$350                | \$331            | \$129                      | \$30                                      |
| REPRESENTACIONES PLATINO S.A.       | \$3.311                             | \$307                | \$807            | \$5.499                    | \$1.98                                    |
| UNION EMP. CONSTRUCTORA CARIBE S.A. | \$95                                | \$182                | \$33             | \$39                       | \$22                                      |
| INVERSIONES GANMA S.A.              | \$1.877                             | \$4.021              | \$4.241          | \$5.529                    | \$28                                      |
| BANCO INTERNACIONAL DE COMERCIO S.A | \$5.946                             | \$10.966             | \$10.724         | \$869                      | \$2.20                                    |
| LEX S.A AGE. PROPIEDAD INDUSTRIAL   | \$17                                | \$123                | \$97             | \$3                        | \$88                                      |
| BUFETE INTERNACIONAL S.A.           | \$4.570                             | \$8.971              | \$2.436          | \$12.485                   | \$6.12                                    |
| CORATUR S.A.                        | \$2                                 | \$131                | \$1              | \$1                        | \$  |
| CONSULTORIA ECONOMICA CANEC S.A.    | \$213                               | \$10.679             | \$807            | \$836                      | \$87                                      |

Figura 16: Imagen del reporte cantidad de egresos por entidad.

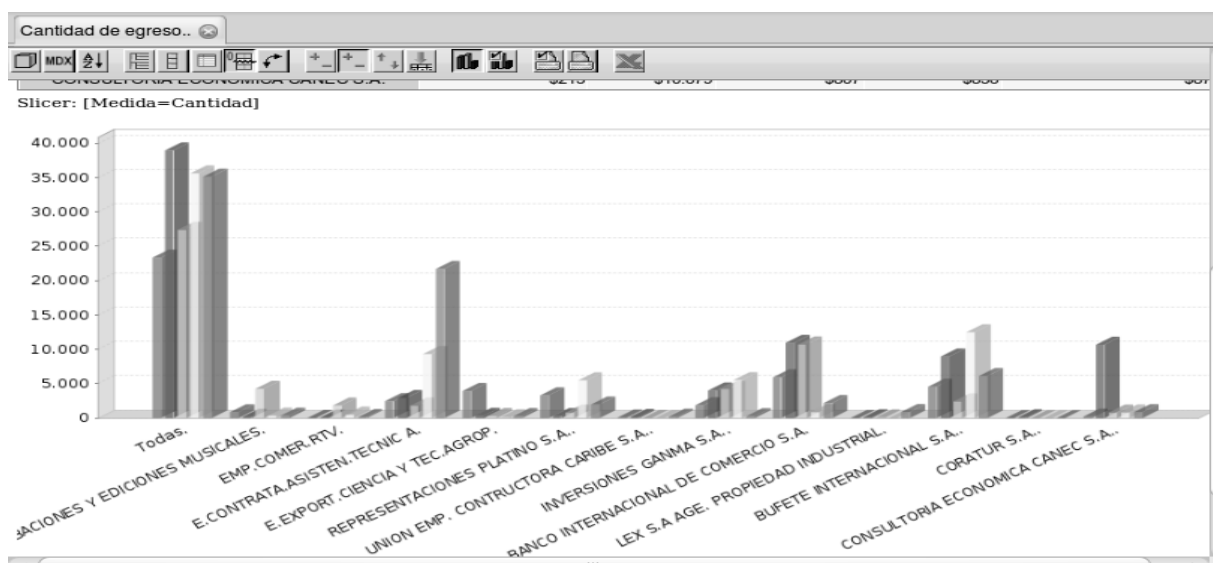


Figura 17: Reporte cantidad de egresos por entidad en forma de gráfica.

## **Conclusiones**

En este capítulo se describieron los elementos de implementación para el desarrollo del mercado de datos Servicios de comercio exterior y se implementaron los tres subsistemas. Luego de finalizada la implementación del MD, el subsistema de almacenamiento quedó estructurado en tres esquemas: dimensiones, mart\_serv\_com\_ext y metadatos, logrando una mejor organización física y lógica de la estructura de la base de datos, facilitando la posterior implementación del proceso de integración de datos. Las transformaciones y trabajos realizados en el subsistema de integración, fueron implementados teniendo en cuenta las reglas de transformación definidas en la fase de análisis, permitiendo una correcta carga de los datos hacia el mercado. Finalmente, el subsistema de visualización implementado, permite estructurar la información que será mostrada al cliente por temas, facilitando el análisis de los datos a través de los reportes que han sido creados para ser consultados por los especialistas.

## CAPÍTULO 4: PRUEBA

### Introducción

En el presente capítulo se realizarán las pruebas al mercado de datos Servicios de comercio exterior, mediante las listas de chequeo aplicadas a los artefactos de ETL, los casos de prueba basados en los casos de uso, las pruebas unitarias y de integración, las pruebas del sistema y las pruebas de aceptación. Posteriormente se evaluarán los resultados para comprobar si el sistema cumple con los requerimientos y necesidades planteadas por el cliente.

### 4.1 Pruebas

El único instrumento adecuado para determinar el estatus de la calidad de un software es el proceso de pruebas. En este proceso se ejecutan pruebas dirigidas a componentes del software o al sistema de software en su totalidad, con el objetivo de medir el grado en que el software cumple con los requerimientos. A continuación se describen las pruebas y herramientas aplicadas al MD:

Prueba unitaria y de integración: Una prueba unitaria es una forma de probar el correcto funcionamiento de un módulo de código. Esto sirve para asegurar que cada uno de los módulos funcione correctamente por separado. Luego, con la Prueba de Integración, se podrá asegurar el correcto funcionamiento del sistema o subsistema en cuestión.

Una prueba unitaria debe cumplir los siguientes requisitos:

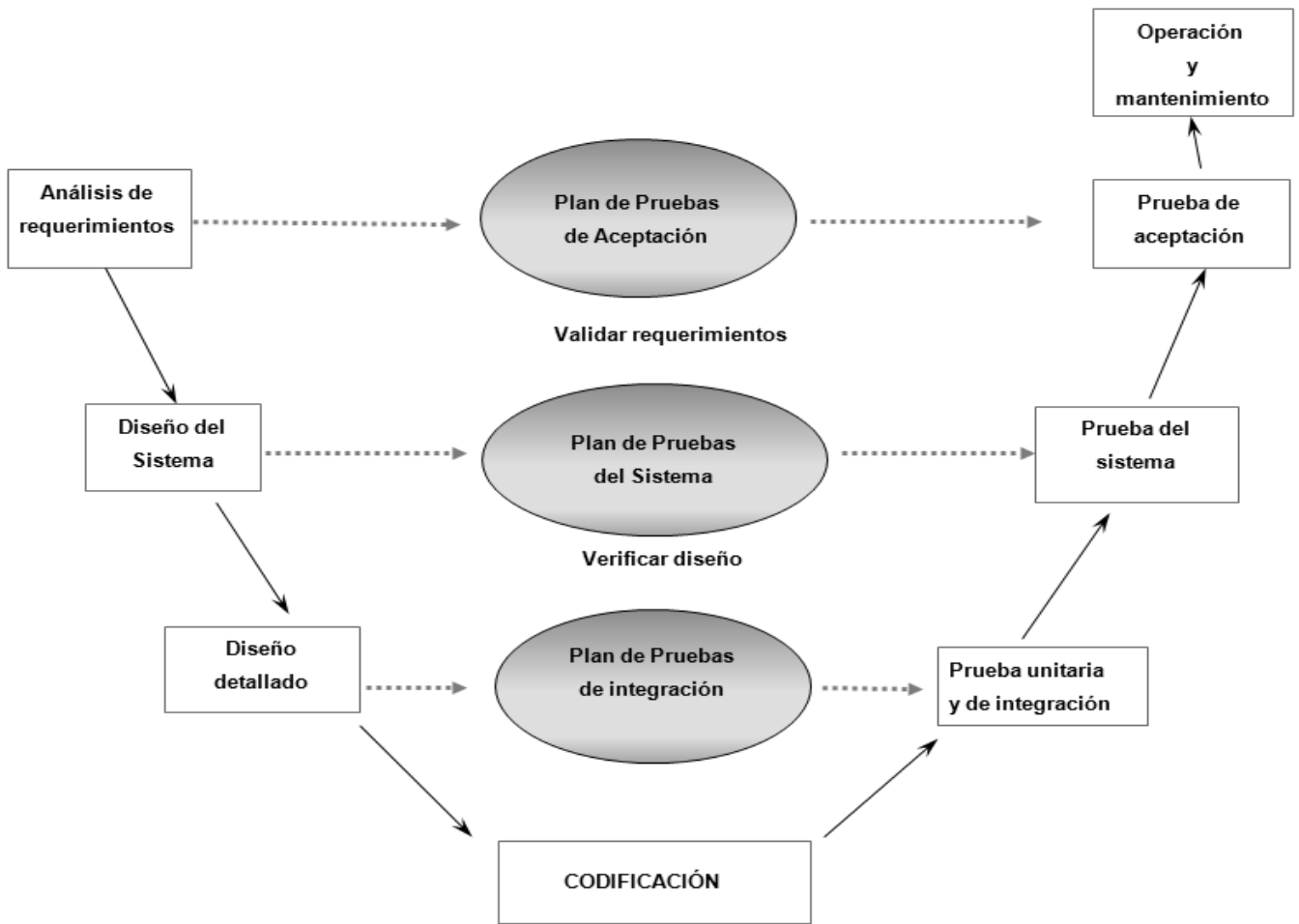
- Automatizable: no debería requerirse una intervención manual.
- Completas: deben cubrir la mayor cantidad de código.
- Repetibles o Reutilizables: no se deben crear pruebas que solo puedan ser ejecutadas una sola vez.
- Independientes: la ejecución de una prueba no debe afectar a la ejecución de otra.
- Profesionales: las pruebas deben ser consideradas igual que el código, con la misma profesionalidad y documentación.

Pruebas del sistema: Son similares a las pruebas de caja negra, solo que estas buscan probar al sistema en conjunto. Están basadas en los requerimientos generales y abarca todas las partes

combinadas del sistema, como aspectos específicos de su comportamiento, tales como seguridad o rendimiento.

Pruebas de aceptación: El objetivo de las pruebas de aceptación es validar que un sistema cumple con el funcionamiento esperado y permitir al usuario de dicho sistema que determine su aceptación, desde el punto de vista de su funcionalidad y rendimiento.

Para el desarrollo de cualquier producto de software se realizan diferentes actividades desde que surge la idea inicial hasta la obtención del producto final. Para establecer un orden de ejecución de estas actividades se utilizó el modelo V, el cual es empleado por el centro DATEC para garantizar la calidad del producto final. En la siguiente figura, se aprecia una representación gráfica del ciclo de vida del software propuesta en el modelo V, donde a la izquierda se muestran las diferentes etapas de desarrollo, mientras que a la derecha se muestran las pruebas correspondientes a cada una de ellas.



**Figura 18:** Modelo V.

#### 4.1.1 Elaboración, evaluación y aplicación de las Listas de chequeo

Esta herramienta utiliza preguntas orientadas a identificar problemas por áreas y sirven para motivar posibles soluciones o la detección de oportunidades de mejora. Las Listas de chequeo se realizan con el objetivo de evaluar la calidad de los artefactos que se generan en el análisis de la solución.

La lista de chequeo definida para la evaluación del mercado de datos Servicios de comercio exterior 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.



- Elementos definidos por el modelo de desarrollo: abarca todos los indicadores a evaluar durante la etapa de desarrollo del mercado según el modelo de desarrollo.
- 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, Elementos definidos por el modelo de desarrollo y Semántica del documento.
- **Evaluación (Eval):** es la forma de evaluar el indicador. El mismo se evalúa de 1 en caso de que exista alguna no conformidad sobre el indicador y 0 en caso de que el indicador revisado no presente inconformidades.
- **No Procede (N.P):** 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.

*Tabla 5: Lista de chequeo.*

| <b>Estructura del documento</b> |  |             |             |  |                    |
|---------------------------------|--|-------------|-------------|--|--------------------|
| <b>Peso</b>                     | <b>Indicadores a evaluar</b>   | <b>Eval</b> | <b>(NP)</b> | <b>Cantidad de elementos afectados</b> | <b>Comentarios</b> |
| crítico                         | 1. ¿El alcance del proyecto describe correctamente los datos de las dimensiones y hechos del mercado de datos? | 0           |             |  |                    |

|         |   |   |  |  |  |
|---------|---|---|--|--|--|
| crítico | 2. ¿El objetivo expresa correctamente el propósito del documento?                                 | 0 |  |  |  |
| crítico | 3. ¿Existe una adecuada correspondencia entre el origen de los datos y los atributos del mercado? | 0 |  |  |  |
|         | 4. ¿Se hace un uso adecuado del control del documento?  | 0 |  |  |  |
|         | 5. ¿En la sección de acrónimos se definen todos los acrónimos utilizados en el documento?         | 0 |  |  |  |

#### 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    |      |                                 |             |

#### Semántica del documento

| Peso    | Indicadores a evaluar  | Eval | (NP) | Cantidad de elementos afectados | Comentarios |
|---------|--|------|------|---------------------------------|-------------|
| crítico | • ¿Se han identificado errores ortográficos en el documento?                     | 0    |      |                                 |             |
| crítico | • ¿Se entiende claramente lo que se ha especificado en el documento?             | 0    |      |                                 |             |
|         | • ¿El número de página que aparece en el índice coincide con el contenido que se |      | NP   |                                 |             |

|  |                                    |  |  |  |  |
|--|------------------------------------|--|--|--|--|
|  | refleja realmente en dicha página? |  |  |  |  |
|--|------------------------------------|--|--|--|--|

#### 4.1.2 Diseño de Casos de pruebas

Se realizó un caso de prueba para el caso de uso de información “Presentar indicadores de servicios de comercio exterior”, a continuación se muestra un ejemplo explicando todo el desarrollo del proceso.

*Tabla 6: Caso de prueba Presentar indicadores de servicios de comercio exterior.*

| Escenario  | Descripción   | Variables de entrada                                | Variables de salida | Respuesta del sistema  | Flujo central   |
|--|---|---|---------------------|--|---|
| EC 1.<br>Cantidad de ingresos y egresos por entidad, servicio, y país. | Muestra los valores de los indicadores de ingreso y egreso por entidad, servicio y país del año 2009. | Indicadores<br>Fecha<br>País<br>Entidad<br>Servicio | Cantidad            | El sistema muestra todas las variables disponibles para el análisis, ubicados en las filas y las columnas que pueden ser visualizadas para cada reporte. | En la parte superior izquierda se selecciona el A.A SIGOB. Se selecciona el área de análisis A.A Servicios de comercio exterior. Se selecciona el libro de trabajo L.T Prueba. En la parte inferior izquierda se selecciona el reporte que corresponde al escenario seleccionado. En el área de trabajo se visualiza la información del reporte seleccionado. |

### **4.1.3 Resultados y discusión de las pruebas**

Aplicadas las pruebas descritas anteriormente al MD Servicios de comercio exterior, se arrojaron los siguientes resultados.

#### **Pruebas unitarias y de integración**

Las pruebas unitarias y de integración se aplicaron a los subsistemas que componen el MD identificándose 4 No Conformidades (NC) que son detalladas a continuación.

En el subsistema de almacenamiento no se identificaron NC.

En el subsistema de integración se identificaron como NC:

**NC1:** No tienen la implementación de la transformación para la carga de las variables de entorno definidas para las conexiones.

**NC2:** Se señalaron algunos detalles en el diseño de integración de datos.

**NC3:** Ponerles nombres sugerentes a los pasos utilizados.

Las NC detectadas en este subsistema presentan complejidad Media.

En el subsistema de visualización se identificaron como NC:

**NC4:** Arreglar datos fijos en el reporte BI.

La NC detectada en este subsistema presenta complejidad Media.

Posteriormente se aplicaron cuatro listas de chequeo a los artefactos generados en el proceso de ETL, donde se identificaron 2 NC, de ellas una de complejidad Alta y otra de complejidad Baja.

#### **Pruebas de sistema**

Las pruebas se aplicaron por parte del grupo de calidad interno del departamento. El caso de prueba basado en caso de uso, permitieron identificar un total de 2 NC, una de complejidad Baja y otra de complejidad Alta.

**NC1:** El flujo central descrito en el caso de prueba no se corresponde con la aplicación

**NC2:** Error ortográfico en las descripciones del LT Indicadores de comercio exterior.

#### **Pruebas de aceptación**

El MD fue presentado al especialista del área, que pudo comprobar que las funcionalidades implementadas satisfacen sus necesidades, emitiendo la carta de aceptación la cual valida que el sistema está listo para ser desplegado.

### **Conclusiones**

En este capítulo se describieron las distintas pruebas realizadas y se detallaron las herramientas aplicadas en el proceso de pruebas, detectándose 8 no conformidades, las cuales permitieron comprobar si el sistema cumple o no con las necesidades del cliente. Estas no conformidades fueron resueltas satisfactoriamente quedando el sistema listo para su total funcionamiento.

## CONCLUSIONES

Luego de haber culminado el proceso de desarrollo de la solución se puede plantear que fueron cumplidos los objetivos propuestos, logrando satisfacer las necesidades del cliente:

- La selección de la metodología, las herramientas y las tecnologías permitieron el desarrollo de todo el proceso de construcción de la solución.
- Se realizó el proceso de análisis y diseño de la solución, en el que se identificaron los requisitos necesarios para solucionar las necesidades del cliente.
- Se realizó el modelo de datos, quedando lista la base de datos relacionada con los servicios de comercio exterior.
- Se implementaron los subsistemas de integración y de visualización de datos, dando como resultado el mercado de datos poblado y funcional.
- La realización de las distintas pruebas, permitieron detectar y solucionar las no conformidades detectadas.

## RECOMENDACIONES

Se recomienda para el mejoramiento del mercado de datos:

- Realizar la integración de otras fuentes de datos que complementen la información del mercado de datos.
- Realizar otra versión del mercado de datos para cumplir con todos los requerimientos y necesidades nuevas planteadas por el cliente.

## REFERENCIAS BIBLIOGRÁFICAS

1. Torrillas, Francisco José Lucas-Torres. Modelos Avanzados de Bases de Datos. ALMACENES DE DATOS Y BASES DE DATOS XML. .
2. BASES DE DATOS VIRTUALES. *BASES DE DATOS VIRTUALES*. [En línea] [Citado el: 14 de Abril de 2012.] <http://basesdedatos.wetpaint.com/page/MODELO+DE+DATOS>.
3. Hernández, Yanisbel González. PROPUESTA DE METODOLGIA PARA EL DESARROLLO DE ALMACENES DE DATOS EN DATEC. La Habana : s.n., 2012.
4. Galeon. *Galeon*. [En línea] [Citado el: 29 de 5 de 2012.] <http://www.galeon.com/zuloaga/Doc/AnalisisRequer.pdf>.
5. Visual Pardigm. Visual Paradigm. [Online] 11 7, 2011. [Citado: 3 14, 2012.] <http://visual-paradigm.com>.
6. PostgreSQL. PostgreSQL. [Online] Copyright 2009-2011. [Citado: 3 15, 2012.] [http://www.postgresql.org.es/sobre\\_postgresql](http://www.postgresql.org.es/sobre_postgresql).
7. Scribd. *Scribd*. [En línea] 2012. [Citado el: 13 de Abril de 2012.] <http://es.scribd.com/doc/36570462/postgreSQL-investigacion>.
8. PostgreSQL : la base de datos mas grande del mundo. [Online] [Citado: 3 15, 2012.] <http://www.arpug.com/ar/trac/wiki/PgAdmin>.
9. La revista de tecnología y estrategia de negocio en Internet. La empresa multidimensional : OLAP. Borja, Ronda : s.n.
10. Data Cleaner. [Online] [Citado: 3 de mayo de 2012] <http://datacleaner.eobjects.org>.
11. PEÑALOZA, LUCIA VICTORIA HERNANDEZ. Diseño y construcción de un DATA MART para la manteción de indicadores de sostenibilidad de la industria del salmon. Santiago de Chile : s.n., 2008.



12. Portada sobre la plataforma Pentaho Open Source Business Intelligence. Portada sobre la plataforma Pentaho Open Source Business Intelligence. [En línea] 2011. [Citado el: 15 de Mayo de 2012.] <http://pentaho.almacen-datos.com>.
14. L., Joaquin E. Oramas. La inteligencia de negocios, un concepto informático. La inteligencia de negocios, un concepto informático.
15. Ibarra, María de los Angeles. IBM DB2 OLAP SERVER. IBM DB2 OLAP SERVER. 2005.
16. Pentaho Kettle Project. Pentaho Kettle Project. [En línea] 2012. [Citado el: 3 de Mayo de 2012.] <http://kettle.pentaho.com/>.
17. Pentaho Data Integration. Pentaho Data Integration. [En línea] 2012. [Citado el: 3 de Mayo de 2012.] <http://www.pentaho.com>.
18. Pentaho BI Platform / Server. Pentaho BI Platform / Server. [En línea] 2012. [Citado el: 15 de Mayo de 2012.] <http://community.pentaho.com>.

## BIBLIOGRAFÍA

1. BERRÍOS, Br. ERICKA GRACIELA SEVILLA. Scribd. *Scribd*. [En línea] 2011. [Citado el: 15 de Marzo de 2012.] <http://es.scribd.com/doc/39978465/GUIA-METODOLOGICA-PARA-LA-DEFINICION-Y-DESARROLLO-DE-UN-DATAWAREHOUSE>.
2. DATA WAREHOUSING: Investigación y sistematización de conceptos. Hefesto: Metodología propia para la construcción de un Datawarehouse. Ricardo Dario, Ing. Bernabeu. 2009.
3. W. H Inmon. Building the Data Warehouse. QED Press/John Wiley, 1992.
4. Nazir, Amer Bin. The feasibility of an effective data warehousing. *The feasibility of an effective data warehousing*. South Africa : s.n., 2008.
5. Almacenando Datos. *Almacenando Datos*. [En línea] 4 de Mayo de 2010. [Citado el: 18 de Marzo de 2012.] [proyectopentahodw.wordpress.com/2010/05/04/modelo-dimENSIONAL](http://proyectopentahodw.wordpress.com/2010/05/04/modelo-dimENSIONAL).
6. Dataprix. 2010. Dataprix. OLAP, MOLAP y ROLAP. [En línea] 2010. [Citado el: 30 de Marzo de 2012.].<http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/i-data-warehousing-investigacion-y-sistematizacion-concepto-17>
7. Diaz, Josep Curto. 2007. Information Management. Reflexiones sobre las tecnologías de la información. [En línea] 7 de Octubre de 2007. [Citado el: 21 de Abril de 2012.]
8. Sinnexus. *Sinnexus*. [En línea] 2008. [Citado el: 1 de Abril de 2012.] [http://www.sinnexus.com/business\\_intelligence/index.aspx](http://www.sinnexus.com/business_intelligence/index.aspx).
9. bprbenchmark. *bprbenchmark*. [En línea] 2010. [Citado el: 2 de Abril de 2012.] [http://www.bprbenchmark.com/articles\\_dir/Inteligenciadenegocios](http://www.bprbenchmark.com/articles_dir/Inteligenciadenegocios).

10. Wiklund, Andreas. El rincón del BI. *El rincón del BI*. [En línea] [Citado el: Abril de 4 de 2012.] <http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimENSIONAL/>.
11. León, Eduardo. Scribd. *Scribd*. [En línea] 2012. [Citado el: 3 de Abril de 2012.] <http://es.scribd.com/doc/36636137/Tutorial-Visual-Paradigm>.
12. Arpug Grupo de usuarios PostgreSQL de Argentina. *Arpug Grupo de usuarios PostgreSQL de Argentina*. [En línea] [Citado el: 4 de Abril de 2012.] <http://www.arpug.com.ar/trac/wiki/PgAdmin>.
13. Ilustrados. *Ilustrados*. [En línea] 2011. [Citado el: 5 de Abril de 2012.] <http://www.ilustrados.com/tema/8205/Almacenes-Datos-Importancia-Estandarizacion-direcciones-para.html>.
14. PostgreSQL. *PostgreSQL*. [En línea] 2012. [Citado el: 10 de Abril de 2012.] <http://www.postgresql.org>.
15. Visual Paradigm. *Visual Paradigm*. [En línea] 2012. [Citado el: 6 de Abril de 2012.] <http://www.visual-paradigm.com/>.
16. Mundo en la web. *Mundo en la web*. [En línea] 2011. [Citado el: 4 de Abril de 2012.] <http://compuroo.blogspot.com/2011/11/data-mart-mercado-de-datos.html>.
17. Pulso social. *Pulso social*. [En línea] 2011. [Citado el: 9 de Abril de 2012.] <http://pulsosocial.com/2011/11/21/mercado-de-datos-crecera-por-diez-segun-proyeccion-de-ericsson/>.
18. LABORSTA Internet. *LABORSTA Internet*. [En línea] 2010. [Citado el: 6 de Abril de 2012.] [http://laborsta.ilo.org/sti/sti\\_S.html](http://laborsta.ilo.org/sti/sti_S.html).
19. Scribd. *Scribd*. [En línea] 2012. [Citado el: 13 de Abril de 2012.] <http://es.scribd.com/doc/36570462/postgreSQL-investigacion>.

20. Pulso social. Pulso social. [En línea] Socialatom Ventures , 2011. [Citado el: 15 de marzo de 2012.] <http://pulsosocial.com>.

21. Data Integration: Using ETL. EAI and EII Tools to Create an Integrate Enterprise. [En línea] [Citado el: 20 de febrero de 2012.] <http://www.bi-bestpractices.com/view-articles/4737>.

## ANEXOS

### Requerimientos no funcionales

#### Usabilidad

**RNF1.** Cumplir con las pautas de diseño de las interfaces.

El sistema debe tener una interfaz gráfica uniforme que incluya pantallas, menús y opciones. Las pautas de diseño se realizarán siguiendo la arquitectura de información definida.

**RNF2.** Mostrar los mensajes, títulos y demás textos que aparezcan en la interfaz del sistema en idioma español e inglés.

Los títulos de los componentes de la interfaz y los mensajes para informar a los usuarios, deben ser en idioma español e inglés y tener una apariencia uniforme en todo el sistema.

**RNF3.** Establecer tiempo de entrenamiento requerido para que usuarios normales sean productivos operando el sistema.

El tiempo de entrenamiento requerido para que usuarios normales sean productivos operando el sistema deberá ser entre 5 y 10 días. Para aquellos usuarios con un nivel avanzado se define como valor máximo 5 días. Para lograr el cumplimiento de los tiempos establecidos por parte de los usuarios es necesario un dominio del funcionamiento del negocio en correspondencia con el rol que ocupen.

**RNF4.** Asegurar la disponibilidad del sistema y la recuperación ante un fallo.

El sistema debe estar disponible durante el horario de trabajo. En caso de fallo, la recuperación del servicio debe fluctuar entre 8 minutos y 72 horas. Este tiempo comprende la solución al problema, así como su validación y prueba.

**RNF5.** Garantizar la conexión de múltiples usuarios al mismo tiempo.

El mercado de datos debe permitir que existan como máximo 20 usuarios conectados de forma simultánea.

#### Confiabilidad

**RNF6.** Garantizar la persistencia de la información.

Se debe realizar un respaldo total de los datos del mercado de datos con una frecuencia trimestral. 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 ONEI.

#### Soporte

**RNF 7.** Lograr la homogeneidad de la estructura de los elementos definidos en el almacén.

Las estructuras del almacén de datos deben tener un nombre estándar teniendo en cuenta el tipo de estructura que sea. En la siguiente tabla se definen convenciones de nombrado con el objetivo de manejar un vocabulario común en todo el almacén de datos, permitiendo un entendimiento claro y conciso por parte de los desarrolladores.

| Estructura            | Descripción  | Ejemplo                          |
|-----------------------|--|----------------------------------|
| Tablas de hechos      | Todas las tablas de hechos tendrán una cadena que demuestra que son hechos y el concepto que describen.  | hech_<concepto>                  |
| Tablas de dimensiones | Todas las tablas de hechos tendrán una cadena que demuestra que son hechos y el concepto que describen.  | dim_<concepto>                   |
| Llaves primarias      | Todas las llaves primarias tendrán una cadena que demuestra que son llaves primarias y el nombre de la tabla a la que pertenecen.              | hech_<concepto>                  |
| Atributos compuestos  | En los atributos donde el nombre es compuesto se debe especificar el primer componente del atributo separado del segundo por un carácter de _. | <Primer nombre>_<Segundo nombre> |

### Restricciones de diseño

**RNF8.** Utilizar los lenguajes de programación y las herramientas definidas.

- El Visual Paradigm 8.0 como herramienta de modelado.

- El gestor de base de datos que se utilizará es PostgreSQL 9.1.2 y como interfaz de administración de dicho gestor PgAdmin 1.14.1. Como lenguaje dentro del sistema gestor de base de datos para la programación en el mercado de datos se utilizará PL/pgSQL.
- DataCleaner 1.5.3 como herramienta para realizar el perfilado de los datos.
- Para el proceso de integración de datos se usará la herramienta Pentaho Data Integrator 4.2.1. En la implementación de los procesos de integración de datos se utilizará el lenguaje JavaScript. También se hará uso del lenguaje MDX para realizar las consultas.
- Schema Workbench 3.2.0 como herramienta gráfica que se utiliza para construir el esquema multidimensional que soportará la creación de los reportes multidimensionales.
- Pentaho BI Server 3.8 como servidor 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. Esta herramienta usa como servidor OLAP el Mondrian OLAP Server 3.2.1.
- Apache Tomcat 6.0.20 como herramienta que se utiliza como servidor web de aplicaciones.

Para el uso de las herramientas anteriores se requiere la instalación de la máquina virtual de java (Java Virtual Machine 6.0 o superior).

### **Requerimientos para la documentación de usuarios en línea y ayuda del sistema.**

**RNF9.** Confección de un manual de usuario.

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

### **Interfaz**

**RNF10.** Acceso al sistema.

El usuario deberá acceder a la aplicación mediante el protocolo HTTP, usando como navegador web Firefox 2.0 en adelante.

### **Interfaces de usuario**

**RNF11.** Garantizar una interfaz amigable al usuario.

El sistema debe tener una interfaz amigable y sencilla de utilizar, teniendo en cuenta que los usuarios finales no son personas adiestradas en el campo de la informática.

### **Interfaces de hardware**

**RNF12.** Definir las interfaces de hardware que soportará el sistema.

El sistema podrá interactuar solamente con una interfaz de hardware: la impresora. Esta interacción se ocasionará cuando se necesite imprimir un reporte en formato físico. El acceso a la impresora será mediante el protocolo TCP/IP a través de la interfaz que ofrece el hardware.

**RNF13.** Proporcionar características mínimas de hardware a las estaciones de trabajo.

- Sistema operativo Linux (distribución Debian, Ubuntu, Nova o cualquiera de las existentes).
- 1 GB RAM.
- 1 Microprocesador Core2Duo.

**RNF14.** Proporcionar características mínimas de hardware a los servidores.

Para lograr una explotación aceptable del sistema los servidores deben contar con los siguientes requerimientos de hardware:

- Sistema operativo Linux (distribución Debian, Ubuntu, Nova o cualquiera de las existentes) o Windows Server 2003 o superior.
- 1 GB RAM.
- 1 Microprocesador Core2Duo

### **Interfaces de software**

**RNF15.** Instalar en las estaciones de trabajo el software necesario para el correcto funcionamiento del sistema.

Las configuraciones de software de las máquinas clientes deben contar al menos con:

- Firefox 2.0 o superior.
- Java Virtual Machine 6.0 o superior.
- Schema Workbench 3.2.1 en caso de que un usuario capacitado requiera la construcción de esquemas multidimensionales para el diseño de nuevos reportes.

### **Requerimientos legales y de derecho de autor.**

**RNF16.** Entregar el sistema a la ONEI.

El sistema debe ser transferido a la ONEI mediante un proceso de transferencia una vez que esté en explotación, incluyendo el código fuente y la documentación correspondiente.

**RNF17.** Requerimientos legales y de derecho de autor.

Se firmará un acuerdo de colaboración entre la UCI y la ONEI donde quedarán establecidas todas las pautas que se deben cumplir durante el desarrollo de la solución. Se reconoce la UCI con derecho de autor y se firmará un proyecto técnico por ambas partes que recoge entre otros elementos, el cronograma de ejecución del producto a desarrollar.



## Acta de aceptación

### ACTA DE ACEPTACIÓN

En La Habana, a los 13 días del mes de junio del 2012

De una parte, la Oficina Nacional de Estadísticas e Información (ONEI) de Cuba, representado en este acto por Elena Lonila Fernández Goría, quien a los fines y efectos derivados del presente documento se denominará como "El cliente", y de otra Parte, el centro de Tecnologías de Gestión de Datos, conocido de forma abreviada como DATEC de la Universidad de las Ciencias Informáticas (UCI), representada en este acto por José Salvador Bermúdez Rodríguez, que a los fines y efectos derivados del presente documento se denominará DATEC.

**Primero:** Que en cumplimiento de los acuerdos, han sido efectuadas las actividades que se describen, **Las partes DECLARAN:**

**CONSIDERANDO:** Que se han efectuado las actividades siguientes:

1. – Diseño del mercado de datos: Servicios de comercio exterior.
2. – Proceso de extracción, transformación y carga de los datos.
3. – Implementación de las vistas de análisis OLAP.

**CONSIDERANDO:** Que las actividades realizadas han sido desarrolladas con la calidad requerida y bajo las condiciones pactadas y aprobadas por **Las Partes**.

**CONSIDERANDO:** Que las actividades que se han ejecutado cumplen con los requerimientos de **El Cliente**.

**CONSIDERANDO:** Que DATEC ha entregado la documentación que avala la ejecución de este acto al **El Cliente**.

**POR TANTO:** **Las Partes** acuerdan formalizar mediante la presente Acta, Aceptadas las actividades que han sido ejecutadas en esta fecha.

Y para que así conste, se extiende la presente Acta en dos (2) ejemplares, rubricados por **Las Partes**.

Por El Cliente

Elena L. Fernández

Nombre y Apellidos

Por DATEC

José Salvador Bermúdez Rodríguez

Nombre y Apellidos

## GLOSARIO

1. **Técnicas de indexación:** son técnicas que se utilizan para acelerar el procesamiento de los datos en bases de datos orientadas a objeto.
2. **JDBC:** Java DataBase Connectivity. Protocolo de conexión de Java a base de datos.
3. **MDX:** Multi-Dimensional Expressions. Es un lenguaje de consulta para bases de datos multidimensionales sobre cubos OLAP, se utiliza en Inteligencia de negocios para generar reportes que apoyen la toma de decisiones basados en datos históricos.
4. **Lista de chequeo:** Instrumento de medición y evaluación que consiste básicamente en un formulario de preguntas referentes al atributo de calidad que se está probando.
5. **IIS:** es un servidor web y un conjunto de servicios para el sistema operativo Microsoft Windows. Originalmente era parte del Option Pack para Windows NT. Luego fue integrado en otros sistemas operativos de Microsoft destinados a ofrecer servicios, como Windows 2000 o Windows Server 2003. Windows XP Profesional incluye una versión limitada de IIS. Los servicios que ofrece son: FTP, SMTP, NNTP y HTTP/HTTPS.
6. **BI:** Se denomina inteligencia empresarial, inteligencia de negocios o BI (del inglés business intelligence) al conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización o empresa.
7. **Datos federados:** Los sistemas de bases de datos federadas son colecciones de componentes cooperativos pero autónomos de sistemas de bases de datos convencionales. En un sistema federado los usuarios tienen acceso a los datos, de los distintos sistemas, a través de una interfaz común sin embargo, no existe un esquema global que describa a todos los datos de las distintas bases de datos, en su lugar hay varios esquemas unificados, cada uno describiendo porciones de bases de datos y archivos para el uso de cierta clase de usuarios.
8. **Sponsors :** Personas con autoridad en una empresa o entidad.
9. **Deduplicación:** Efecto de duplicar o duplicarse.
10. **Bases de datos MongoDB:** MongoDB es un sistema de base de datos multiplataforma orientado a documentos, de esquema libre. Esto significa que cada entrada o registro puede tener un esquema de datos diferentes, con atributos o "columnas" que no tienen por qué repetirse de un registro a otro. Está escrito en C++, lo que le confiere cierta cercanía al bare metal, o recursos de hardware de la máquina, de modo que es bastante rápido a la hora de ejecutar sus tareas. Además, está licenciado como GNU AGPL 3.0, de modo que se trata de un software de licencia libre. Funciona en sistemas operativos Windows, Linux, OS X y Solaris.
11. **API:** es el conjunto de funciones y procedimientos (o métodos, en la programación orientada a objetos) que ofrece cierta biblioteca para ser utilizado por otro software como una capa de

abstracción. Son usadas generalmente en las bibliotecas (también denominadas vulgarmente "librerías").

- 12. Sysml:** por sus siglas en inglés (Systems Modeling Language) es un lenguaje de especificación de sistemas. Es un lenguaje desarrollado como perfil de UML 2.0 desde la Meta-Object Facility (MOF).
- 13. BPMN:** Business Process Modeling Notation (en español Notación para el Modelado de Procesos de Negocio) es una notación gráfica estandarizada que permite el modelado de procesos de negocio, en un formato de flujo de trabajo (workflow).
- 14. XMI o XML:** Metadata Interchange (XML de Intercambio de Metadatos) es una especificación para el Intercambio de Diagramas. La especificación para el intercambio de diagramas fue escrita para proveer de una manera de compartir modelos UML entre diferentes herramientas de modelado. En versiones anteriores de UML se utilizaba un Schema XML para capturar los elementos utilizados en el diagrama; pero este Schema no decía nada acerca de cómo debía representarse el modelo. Para solucionar este problema la nueva Especificación para el Intercambio de Diagramas fue desarrollada mediante un nuevo Schema XML que permite construir una representación SVG (Scalable Vector Graphics). Típicamente esta especificación es utilizada solamente por quienes desarrollan herramientas de modelado UML.
- 15. MAC:** (siglas en inglés de media access control; en español "control de acceso al medio") es un identificador de 48 bits (6 bloques hexadecimales) que corresponde de forma única a una tarjeta o dispositivo de red. Se conoce también como dirección física, y es única para cada dispositivo. Está determinada y configurada por el IEEE (los últimos 24 bits) y el fabricante (los primeros 24 bits) utilizando el organizationally unique identifier. La mayoría de los protocolos que trabajan en la capa 2 del modelo OSI usan una de las tres numeraciones manejadas por el IEEE: MAC-48, EUI-48, y EUI-64, las cuales han sido diseñadas para ser identificadores globalmente únicos. No todos los protocolos de comunicación usan direcciones MAC, y no todos los protocolos requieren identificadores globalmente únicos.