

**Universidad de las Ciencias Informáticas  
Facultad 6**



**Título:** Mercado de datos para el área de Tecnologías de la información y la comunicaciones del Sistema de Información de Gobierno.

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

**Autora:**

Lianet Carrillo Barroso

**Tutores:**

Ing. Elio Luis Toledo García

Ing. Yosvany Arrastia Machin

La Habana 2013

”Año 55 de la Revolución”

*Para el logro del triunfo siempre ha sido indispensable pasar por la senda de los sacrificios.*

*Simón Bolívar*

**Declaración de autoría**

Declaro ser autora del presente trabajo “Mercado de datos para el área de Tecnologías de la información y las comunicaciones del Sistema de Información de Gobierno” y reconozco a la Universidad de las Ciencias Informáticas (UCI) los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmo la presente a los \_\_\_\_ días del mes de \_\_\_\_\_ del año 2013.

---

Lianet Carrillo Barroso

Autora

---

Ing. Yosvany Arrastia Machin

Tutor

---

Ing. Elio Luis Toledo García

Tutor

### Datos de contacto

Tutor: Ing. Elio Luis Toledo García.

Especialidad de graduación: Ingeniería en Ciencias Informáticas.

Categoría docente: Instructor.

Categoría científica: Ingeniero.

Años de experiencia en el tema: 1

Años de graduado: 2

Correo electrónico: [eltoledo@uci.cu](mailto:eltoledo@uci.cu)

Tutor: Ing. Yosvany Arrastia Machin

Especialidad de graduación: Ingeniería Informática

Categoría docente: Instructor

Categoría científica: Ingeniero

Años de experiencia en el tema: 2

Años de graduado: 3

Correo electrónico: [yarrastia@uci.cu](mailto:yarrastia@uci.cu)

*Quiero agradecerles a dos personitas más importantes en mi vida, a mis padres. Por darme la felicidad del mundo, por enseñarme y por creer en mí.*

*A la Mimi, por ser la mejor mamá del mundo, por poner toda su confianza en mí y por ser el ejemplo a seguir en cada paso que doy.*

*A mi Papi, por darme todo el apoyo, esfuerzo y dedicación en cada instante de mi vida. Gracias por tus consejos y por todo el amor que me has dado.*

*A mi hermano por ser mi confidente y ser mi mejor amigo.*

*Existe una personita que en tan poco tiempo me hizo la reina de su mundo. Gracias por todo este tiempo, por aguantar mis malas crianzas y por las noches de desvelos que dedicaste para cumplir este gran sueño. Gracias mi Piti Bu.*

*A mi abuela Yiya por ser tan especial, por crear y cuidar a la familia que tengo.*

*A mis tías Magalis y María Elena, a mis primos Yoe, Yenly, Eddy y Eliza y a mi madrina, que me han apoyado y ayudado en todo momento.*

*A mi amiga del Pre Universitario Keilin, por preocuparse por mí y por su energía positiva.*

*A mi cuñada y a mis suegros por darme el apoyo que necesitaba en los momentos difíciles, por su respeto, por estar siempre ahí en el momento y lugar adecuado.*

*Les agradezco a mis tutores Elio y Yosvany por darme su voto de confianza, por su atención en todo momento y por depositar todos sus esfuerzos, en especial a Elio por su eterna paciencia y por todos esos momentos que estuve aprendiendo a ser mejor profesional.*

*A mis compañeros y compañeras de estudio, en especial a Yoa, Anne, Yasma y Luis Ángel. Por brindarme su amistad y por tantos momentos imborrables.*

*A todos los profes que contribuyeron a mi formación como ingeniera, en especial a Wendy, Monchi, Esley, Ismael, Manrresa, Rayko, Mario, Yonelbys por tener siempre una respuesta a mis preguntas.*

*A todas aquellas personas que confiaron en mí.*

*A mis padres, por amarme, por darme la vida. Por guiarme a lo largo de los años con su amor, paciencia y comprensión.*

*A mi hermano por ser mi confidente.*

*A mi novio por ayudarme y apoyarme en todo momento.*

*A toda mi familia que ha esperado lo mejor de mí.*

**Resumen**

La investigación surge como parte de la colaboración que existe entre la Universidad de las Ciencias Informáticas y la Oficina Nacional de Estadísticas e Información. Dicha entidad cuenta con diferentes áreas, entre ellas se encuentra el área de Tecnologías de la información y las comunicaciones, la cual recoge la información mediante el Sistema Integrado de Gestión y Estadísticas. Además almacena la información de la encuesta anual realizada a las entidades de esta área. La metodología y las herramientas utilizadas durante la investigación; en conjunto con el análisis y el diseño, la implementación y las pruebas arrojaron como resultado final la realización del Mercado de datos para el área de Tecnologías de la información y las comunicaciones del Sistema de Información de Gobierno, lo que permite hacer análisis históricos de los principales indicadores de esta área, facilitando la toma de decisiones a los directivos de la Oficina Nacional de Estadísticas e Información.

**Palabras claves:**

Oficina Nacional de Estadísticas e Información, Tecnologías de la información y las comunicaciones, Mercado de Datos, Toma de decisiones.

**Índice de contenido**

|   |    |
|---|----|
| Resumen   |    |
| Introducción.....   | 1  |
| Capítulo 1: Fundamentos teóricos.....   | 5  |
| 1.1 Almacenes de Datos.....   | 5  |
| 1.1.1 Características de los Almacenes de Datos .....   | 5  |
| 1.1.2 Ventajas y desventajas de los Almacenes de Datos .....  | 5  |
| 1.2 Mercado de Datos.....   | 6  |
| 1.3 Modelo Multidimensional .....   | 6  |
| 1.4 Procesamiento analítico en línea (OLAP) .....   | 8  |
| 1.5 Metodologías para el desarrollo de los Almacenes de Datos .....   | 9  |
| 1.5.1 Metodología a utilizar en la solución .....   | 9  |
| 1.6 Herramientas a utilizar en la solución .....  | 11 |
| 1.6.1 Herramienta para el modelado de los datos .....   | 11 |
| 1.6.2 Sistema Gestor de Base de Datos .....   | 12 |
| 1.6.3 Herramientas para la Extracción, Transformación y Carga (ETL) .....   | 13 |
| 1.6.4 Herramientas de Inteligencia de negocio (BI) .....  | 14 |
| Conclusiones del capítulo: .....  | 14 |
| Capítulo 2: Análisis y diseño del Mercado de datos para el área de Tecnologías de la información y las comunicaciones. .... | 16 |
| 2.1 Necesidad de información .....  | 16 |
| 2.2 Requisitos de información .....   | 17 |
| 2.3 Requisitos funcionales .....  | 19 |
| 2.4 Requisitos no funcionales .....   | 20 |
| 2.5 Reglas del negocio .....  | 21 |
| 2.6 Caso de uso del sistema .....   | 21 |
| 2.6.1 Actores del sistema .....   | 22 |
| 2.6.2 Especificación de Casos de uso .....  | 22 |
| 2.7 Arquitectura del sistema .....  | 24 |

|  |    |
|--|----|
| 2.8 Diseño del subsistema de almacenamiento.....   | 25 |
| 2.8.1 Dimensiones .....  | 25 |
| 2.8.2 Hechos y medidas .....   | 26 |
| 2.8.3 Matriz bus.....  | 27 |
| 2.8.4 Modelo de datos .....  | 28 |
| 2.9 Diseño de los procesos de integración de datos .....   | 29 |
| 2.9.1 Perfilado de datos .....   | 29 |
| 2.9.2 Diseño de las transformaciones .....   | 29 |
| 2.10 Diseño del subsistema de visualización.....   | 33 |
| 2.10.1 Diseño de los cubos OLAP.....   | 33 |
| 2.10.2 Diseño de la arquitectura de información.....   | 34 |
| 2.11 Política de respaldo y recuperación .....   | 35 |
| 2.12 Esquemas de seguridad.....  | 35 |
| 2.12.1 Seguridad en el subsistema de almacenamiento .....  | 35 |
| 2.12.2 Seguridad en el subsistema de integración .....   | 36 |
| 2.12.3 Seguridad en el subsistema de visualización .....   | 36 |
| Conclusiones del capítulo: .....   | 36 |
| Capítulo 3: Implementación del Mercado de datos para el área de Tecnologías de la información y las comunicaciones ..... | 38 |
| 3.1 Implementación del subsistema de almacenamiento .....  | 38 |
| 3.1.1 Estandarización de los nombres .....   | 38 |
| 3.1.2 Estructura de los datos .....  | 39 |
| 3.1.3 Estrategias de indexado .....  | 40 |
| 3.2 Implementación del subsistema de integración de datos (ETL) .....  | 40 |
| 3.2.2 Gestión del cambio lento en las dimensiones.....   | 43 |
| 3.2.3 Gestión de los metadatos del proceso de integración .....  | 45 |
| 3.2.4 Gestión de cambios en los datos fuentes .....  | 45 |
| 3.3 Implementación del subsistema de visualización de datos .....  | 45 |
| 3.3.1 Navegación de la capa de visualización .....   | 46 |
| 3.3.2 Implementación de los reportes candidatos.....   | 46 |

|  |    |
|--|----|
| 3.3.3 Seguridad de la aplicación .....   | 47 |
| Conclusiones del capítulo .....  | 47 |
| Capítulo 4: Prueba del Mercado de datos para el área de Tecnologías de la información y las comunicaciones ..... | 49 |
| 4.1 Calidad de software .....  | 49 |
| 4.2 Pruebas de calidad de software .....   | 49 |
| 4.2.1 Modelo V .....   | 49 |
| 4.3 Herramientas de pruebas .....  | 51 |
| 4.4 Resultados de las pruebas .....  | 52 |
| 4.5 Calidad de datos .....   | 53 |
| 4.5.1 Perfilado de datos .....   | 53 |
| 4.5.2 Auditoría de datos .....   | 54 |
| Conclusiones del capítulo: .....   | 54 |
| Conclusiones.....  | 56 |
| Recomendaciones .....  | 57 |
| Referencias bibliográficas.....  | 58 |
| Anexos .....   | 61 |
| Glosario de términos .....   | 62 |

### Índice de Figuras

|   |    |
|---|----|
| Figura 1. Esquema en estrella .....   | 7  |
| Figura 2. Esquema copo de nieve .....   | 7  |
| Figura 3. Constelación de hechos .....  | 8  |
| Figura 4. Fases de la Propuesta de metodología para el desarrollo de Almacenes de Datos en DATEC. . | 11 |
| Figura 5. Diagrama de caso de uso del sistema.....  | 22 |
| Figura 6. Arquitectura del sistema.....   | 25 |
| Figura 7. Modelo de datos .....   | 28 |
| Figura 8. Diseño de la transformación dim_pregunta .....  | 30 |
| Figura 9. Diseño de las transformaciones hech_encuesta_anual_tic .....                              | 31 |
| Figura 10. Diseño de las transformaciones hech_ingresos_por_tic y hech_informacion_tic.....         | 32 |
| Figura 11. Diseño del cubo hech_informacion_tic .....   | 34 |

|  |    |
|--|----|
| Figura 12. Arquitectura de información. ....   | 34 |
| Figura 13. Transformación del hecho hech_informacion_tic. ....   | 42 |
| Figura 14. Implementación del trabajo para los hechos hech_informacion_tic y hech_ingresos_por_tic. .... | 43 |
| Figura 15. Implementación del trabajo para la encuesta. ....   | 43 |
| Figura 16. Implementación general. ....  | 43 |
| Figura 17. Arquitectura del Área de Análisis. ....   | 46 |
| Figura 18. Vista de análisis Presentar información TIC. ....   | 47 |
| Figura 19. Modelo V. ....  | 50 |
| Figura 20. Resultado del perfilado de datos al hech_informacion_tic. ....                                | 54 |

### Índice de Tablas

|   |    |
|---|----|
| Tabla 1. CUI Presentar Información TIC. ....  | 24 |
| Tabla 2. Matriz bus. ....   | 27 |
| Tabla 3. Seguridad en la base de datos. ....  | 35 |
| Tabla 4. Seguridad en la aplicación. ....   | 36 |
| Tabla 5. Estándares de codificación del Mercado de datos para el área de Tecnologías de la información y las comunicaciones. .... | 39 |
| Tabla 6. Estructura del Mercado de datos para el área de Tecnologías de la información y las comunicaciones. ....                 | 40 |
| Tabla 7. Resultado de la aplicación de las listas de chequeo a los artefactos de los procesos de ETL. ....                        | 53 |

## Introducción

A lo largo de la historia el hombre ha necesitado encontrar métodos rápidos y efectivos para resolver sus cálculos, almacenar datos y transmitir información. Desde que se inició la era de la computadora, las organizaciones han usado los datos desde sus sistemas operacionales para atender sus necesidades de información. Para ello ha creado distintos mecanismos y herramientas que le faciliten el trabajo. Con el avance de la ciencia y la tecnología surgen los Almacenes de Datos, proporcionando a las instituciones que realicen un adecuado análisis de la información.

Hoy en día los Almacenes de Datos se aplican en la mayoría de los campos de la actividad humana, desde comercial, científico e industrial, ayudando a la toma de decisiones. Esto ha llevado al desarrollo de nuevas soluciones y planteamientos de las empresas a nivel mundial. Cuba se encuentra inmersa en esta evolución tecnológica, que trae consigo un proceso creciente en su aplicación a los diversos procesos institucionales, para la obtención de información confiable y actualizada en los procesos de dirección y toma de decisiones.

La Universidad de las Ciencias Informáticas (UCI) forma parte del programa que se ha propuesto la revolución cubana para alcanzar un avance en la informatización de la sociedad y desarrollar la industria de software en el país. El Centro de Tecnologías de Gestión de Datos (DATEC) es uno de los centros de producción de la UCI. Este centro tiene la misión de proveer soluciones integrales, productos y servicios relacionados con las tecnologías de gestión de datos. DATEC está compuesto por cuatro líneas de producción de software: PostgreSQL, Bioinformática, Integración de Soluciones y Almacenes de Datos.

Actualmente la línea de Almacenes de Datos cuenta con numerosos proyectos entre ellos se encuentra el Sistema de Información de Gobierno (SIGOB), que recoge en un Almacén de Datos (AD) la información de las diferentes áreas de la Oficina Nacional de Estadísticas e Información (ONEI). Dicha entidad es la encargada de obtener, analizar, organizar, procesar y aprobar la información según las necesidades y los objetivos del gobierno. La ONEI está dividida por áreas, entre ellas se encuentra el área de Tecnologías de la información y las comunicaciones (TIC).

Dentro de esta área se evidencian algunos problemas como: parte de la información es recogida a través de la herramienta Microsoft Access, la cual puede ser consultada por un especialista que posea conocimiento del negocio y la informática. Los ficheros que se generan con dicha herramienta tienen una periodicidad anual provocando un gran cúmulo de información lo que dificulta el análisis estadístico de los

mismos. El resto de la información es recogida a través del Sistema Integrado de Gestión y Estadística (SIGE) el cual permite almacenar, manipular y consultar la información de cada una de las áreas de la ONEI y sin embargo no brinda la posibilidad de registrar los cambios que pueden sufrir los indicadores con el transcurso del tiempo. Dificultando el análisis histórico lo cual influye negativamente en la disponibilidad de la información requerida para la toma de decisiones en el gobierno.

Por la situación anteriormente descrita se identifica el siguiente **problema de la investigación**: ¿Cómo contribuir a la toma de decisiones en el área de Tecnologías de la información y las comunicaciones del Sistema de Información de Gobierno?

La investigación tiene como **objeto de estudio** los Almacenes de Datos enmarcado en el **campo de acción** Mercado de datos para el área de Tecnologías de la información y las comunicaciones del Sistema de Información de Gobierno.

Para dar solución al problema de la investigación se ha definido como **objetivo general** Desarrollar el Mercado de datos para el área de Tecnologías de la información y las comunicaciones del Sistema de Información de Gobierno que contribuya a la toma de decisiones, el cual se desglosa en los siguientes **objetivos específicos**:

- Fundamentar la selección de la metodología y las herramientas a utilizar en el desarrollo del Mercado de datos para el área de Tecnologías de la información y las comunicaciones.
- Realizar el análisis y el diseño del Mercado de datos para el área de Tecnologías de la información y las comunicaciones.
- Implementar el Mercado de datos para el área de Tecnologías de la información y las comunicaciones.
- Probar el Mercado de datos para el área de Tecnologías de la información y las comunicaciones.

Para cumplir los objetivos específicos planteados se proponen las siguientes **tareas de investigación**:

1. Fundamentación de la metodología y las herramientas a utilizar en el desarrollo Almacenes de Datos para guiar los procesos de desarrollo del Mercado de datos para el área de Tecnologías de la información y las comunicaciones.
2. Levantamiento de los requisitos de información para identificar las necesidades del cliente.
3. Descripción de los casos de uso del Mercado de datos para el área de Tecnologías de la información y las comunicaciones para determinar cada una de las funcionalidades del sistema.

4. Definición de los hechos, las medidas y las dimensiones del Mercado de datos para el área de Tecnologías de la información y las comunicaciones identificando cada uno de los elementos que van a formar parte del modelo de datos.
5. Definición de la arquitectura del Mercado de datos para el área de Tecnologías de la información y las comunicaciones para definir los subsistemas que forman parte de la solución.
6. Diseño del modelo de datos para identificar los elementos que va a contener el modelo físico.
7. Implementación del modelo de datos para garantizar la disponibilidad de las estructuras de la base de datos.
8. Diseño del subsistema de integración de datos para definir la carga de las dimensiones y hechos.
9. Diseño del subsistema de visualización para realizar el diseño de los cubos OLAP y definir la estructura de la capa de visualización.
10. Diseño de los casos de pruebas para apoyar la realización de las pruebas.
11. Implementación del subsistema de integración de datos para poblar el Mercado de datos para el área de Tecnologías de la información y las comunicaciones.
12. Implementación del subsistema de visualización para mostrar la información al cliente.
13. Aplicación de las listas de chequeo para garantizar la calidad en los procesos de integración.
14. Aplicación de los casos de prueba para garantizar el correcto funcionamiento del sistema.

### **Estructura del trabajo de diploma**

El trabajo de diploma se encuentra estructurado: resumen, introducción, cuatro capítulos, conclusiones, recomendaciones, referencias bibliográficas, bibliografía, anexos y glosario de términos.

#### **Capítulo 1:** Fundamentos teóricos.

En este capítulo se abordan definiciones y conceptos de los AD y los Mercados de Datos (MD). Además se documentan la metodología y las herramientas para el desarrollo de un AD.

**Capítulo 2:** Análisis y diseño del Mercado de datos para el área de Tecnologías de la información y las comunicaciones.

En este capítulo se realiza un estudio y análisis preliminar del negocio. Se especifican las necesidades de información, las reglas del negocio, los requisitos funcionales, los requisitos no funcionales y los requisitos de información, así como los casos de uso del sistema. Se realiza la arquitectura del sistema, el diseño de

los subsistemas de almacenamiento, integración y visualización del Mercado de datos para el área de Tecnologías de la información y las comunicaciones.

**Capítulo 3:** Implementación del Mercado de datos para el área de Tecnologías de la información y las comunicaciones.

En este capítulo se realiza la implementación de cada uno de los subsistemas, el subsistema de almacenamiento, el subsistema de integración y el subsistema de visualización que conforman el Mercado de datos para el área de Tecnologías de la información y las comunicaciones.

**Capítulo 4:** Prueba del Mercado de datos para el área de Tecnologías de la información y las comunicaciones.

Existen diferentes tipos de pruebas que pueden ser aplicadas a un MD con el objetivo de garantizar el cumplimiento de las exigencias del cliente y la calidad del producto. En este capítulo se prueba la solución obteniendo resultados en cada una de ellas.

## Capítulo 1: Fundamentos teóricos.

### Introducción

En este capítulo se abordan definiciones y conceptos de los AD y los MD así como las ventajas, las desventajas y sus características del uso de los mismos. Además se documenta la metodología que guía el desarrollo de un AD, así como las herramientas que permitirán dar solución al problema planteado.

### 1.1 Almacenes de Datos

Un AD es una colección de datos que recoge información de múltiples sistemas fuentes u operacionales dispersos, cuya actividad se centra en el análisis de la información. Guarda los datos reunidos y permite el acceso histórico; proporcionando al usuario una interfaz para los datos y facilitando la realización de consultas para la toma de decisiones. (1)

Bill Inmon define un AD de la siguiente manera: *"Un Almacén de Datos es una recopilación de datos temáticos, integrados, no volátiles y con historial para la toma de decisiones"*. (2)

Ralph Kimball, otro conocido autor en el tema de los AD, define a estos como: *"Una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis"*. (3)

La investigación se ajustó a la definición planteada por Bill Inmon, pues es el concepto más completo definiendo las características de los AD.

#### 1.1.1 Características de los Almacenes de Datos

Los AD se caracterizan por ser:

- **Orientado a tema:** Los datos están organizados por materias o por temas.
- **Integrado:** Los datos deben ser consistentes.
- **No volátil:** La información una vez incorporada al AD debe mantenerse, en general, invariable. Cargar nuevos datos y consultar los almacenados, no permite borrar, ni modificar los mismos.
- **Variantes en el tiempo:** Los informes que se generan, reflejan las variaciones que se han producido en los datos a lo largo del tiempo. (2)

#### 1.1.2 Ventajas y desventajas de los Almacenes de Datos

Existen diversas ventajas por la que es recomendable usar el AD. Algunas de ellas son:

**Ventajas:**

1. Facilitan la información de manera correcta, en el tiempo y formato necesario para la toma de decisiones empresariales.
2. Se obtiene una base de datos histórica y clasificada por temas.

Utilizar AD también tiene sus inconvenientes.

**Desventajas:**

1. Tienen un alto coste, en cuanto a la utilización de recursos humanos, medios básicos y material gastable. (4)

**1.2 Mercado de Datos**

Un MD *“es una base de datos departamental, especializada en el almacenamiento de datos de un área de negocio específica. Se caracteriza por disponer de una estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicha área de negocio.”*

(4)

Características de los MD:

- Contienen poco volumen de datos.
- Mayor rapidez de consulta.
- Se realizan consultas SQL y MDX. (5)

**1.3 Modelo Multidimensional**

En el diseño de los AD se recomienda utilizar el modelo multidimensional (MMD), que tiene como ventaja con respecto al modelo entidad-relación (MER), ser flexible y estar desnormalizado; el cual es representado a través de tablas de hechos con sus respectivas dimensiones.

**Tablas de dimensiones:** Las tablas de dimensiones almacenan los valores que se utilizan en las tablas de hechos. Contienen una llave simple y atributos que la describen. (6)

**Tablas de hechos:** Contienen los datos de la organización y generalmente son de tipo numéricos. Cada tabla de hechos incluye las llaves primarias de las tablas de dimensiones relacionadas, así como los atributos de los registros de hechos. (7)

Existen esquemas para representar los modelos multidimensionales: el esquema en estrella, el copo de nieve y constelación de hechos:

**Esquema en estrella:** Consta de una tabla de hecho y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves. En la Figura 1 se puede apreciar un esquema en estrella:

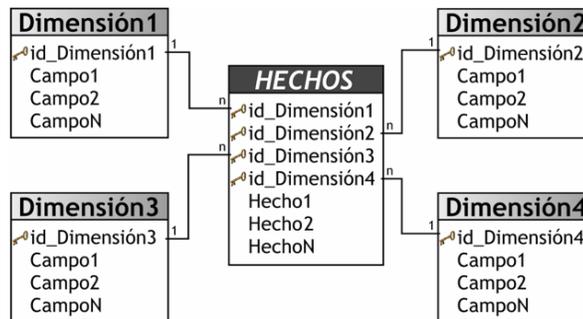


Figura 1. Esquema en estrella.

**Esquema copo de nieve:** Este esquema representa una extensión del modelo en estrella, cuando las tablas de dimensiones se organizan en jerarquías de dimensiones (ver Figura 2).

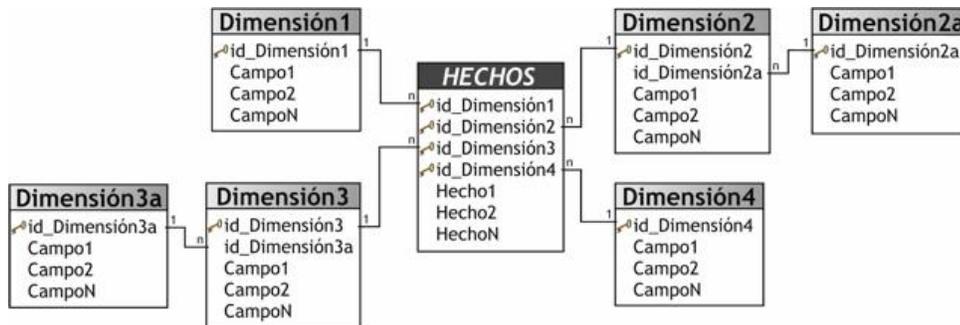


Figura 2. Esquema copo de nieve.

**Constelación de hechos:** Un esquemas de constelación de hechos es una combinación de una serie de esquemas en estrella, es decir es un conjunto de tablas de hechos que comparten algunas tablas de dimensiones. Tiene como ventaja la reutilización de las tablas de dimensiones, pues una tabla de dimensión puede ser utilizada para varias tablas de hechos. En la Figura 3 se puede apreciar un esquema constelación de hechos. (8)

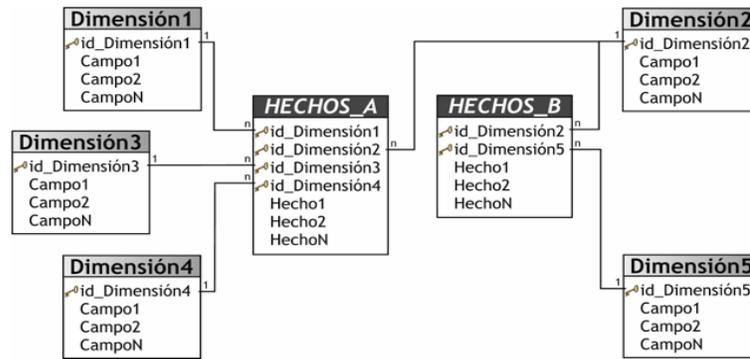


Figura 3. Constelación de hechos.

### 1.4 Procesamiento analítico en línea (OLAP)

Procesamiento analítico en línea (OLAP): Es una solución utilizada en el campo de la Inteligencia de Negocios (9). Son sistemas que permiten recolectar, organizar la información y disponer de ella mediante gráficos y reportes. Se basa en el análisis multidimensional, lo que permite al usuario ver los datos desde diferentes perspectivas (dimensiones) del negocio. (10)

A continuación se presentan los tipos de almacenamientos OLAP:

- **Proceso analítico en línea relacional (ROLAP):** Trabaja directamente con las bases de datos relacionales, es decir los datos son almacenados en filas y columnas de forma relacional. Una desventaja de los sistemas ROLAP, es que los datos de los cubos se deben calcular cada vez que se ejecuta una consulta sobre ellos. Esto provoca que ROLAP no sea muy eficiente en cuanto a la rapidez de respuesta ante las consultas de los usuarios. (11)
- **Proceso analítico en línea multidimensional (MOLAP):** El objetivo de los sistemas MOLAP es almacenar físicamente los datos en estructuras multidimensionales, es decir la información se almacena y visualiza multidimensionalmente. El principal motivo de pre calcular los datos de los cubos, es que posibilita que las consultas sean respondidas con mucha rapidez, pues los mismos no deben ser calculados en tiempo de ejecución. (9)

**Proceso analítico en línea híbrido (HOLAP):** Constituye un sistema híbrido entre MOLAP y ROLAP, combina estas dos implementaciones para almacenar algunos datos en un motor relacional y otros en una base de datos multidimensional. Se utilizará ROLAP para navegar y explorar los datos y se empleará MOLAP para la realización de tableros. (9)

En la investigación se utilizó la arquitectura ROLAP pues las capacidades OLAP se soportan mejor contra las bases de datos relacionales. La principal razón por la que se selecciona esta arquitectura es porque se utilizará el Sistema Gestor de Bases de Datos (SGBD) PostgreSQL, el cual permite utilizar un modelo relacional.

### 1.5 Metodologías para el desarrollo de los Almacenes de Datos

Las metodologías de desarrollo de software son un conjunto de procedimientos, técnicas y ayudas a la documentación para el desarrollo de productos software. En estos se van indicando paso a paso todas las actividades a realizar para lograr el producto informático deseado. Además indican qué personas deben participar en el desarrollo de las actividades y qué papel juegan. También se detalla la información que se debe producir como resultado de una actividad y la información necesaria para comenzarla. (11)

Existen dos enfoques de desarrollo para la construcción de los AD, el enfoque ascendente y el enfoque descendente. La visión de Bill Inmon se basa en un enfoque descendente (del inglés *top-down*) pues se debe ir desde una visión más general de las distintas partes que componen el almacén y posteriormente ir concretando y refinando cada una de las partes por separado. Por ello, una vez desarrollado el AD, se empieza el desarrollo de los MD. Sin embargo, esta metodología se contrapone con el enfoque ascendente (del inglés *bottom-up*) que defiende Kimball. La idea es construir MD independientes para evaluar las ventajas del nuevo sistema a medida que se avanza. Las partes individuales se diseñan con detalle y luego se enlazan para formar componentes más grandes, que a su vez se relacionan hasta formar el sistema completo. (12)

#### 1.5.1 Metodología a utilizar en la solución

Para guiar el proceso de desarrollo del Mercado de datos para el área de Tecnologías de la información y las comunicaciones se seleccionó como metodología: Propuesta de metodología para el desarrollo de Almacenes de Datos en DATEC, esta toma como base la metodología Ciclo de vida de Kimball por las siguientes razones:

- Identifica las tablas de hechos y las tablas de dimensiones, lo que proporciona mayor agilidad en el proceso de desarrollo y con ello la toma de decisiones.
- Propone la construcción de MD departamentales, para después integrarlo como un todo en el AD. Esto permite ir presentando resultados parciales a los clientes en cortos plazos.
- Existe amplia documentación sobre la misma y puede ser consultada ante cualquier duda.

Añade la fase de prueba, lo que permite fortalecer la salida del producto. Se basa en la tesis de doctorado Metodología para el Diseño Conceptual de Almacenes de Datos de Leopoldo Zenaido Zepeda Sánchez, del cual incluye los casos de uso para guiar los procesos de desarrollo.

En la Figura 4 se muestran las ocho fases de la Propuesta de metodología para el desarrollo de Almacenes de Datos en DATEC. En la investigación se hace uso de cinco fases pues el alcance de la investigación termina en la fase de prueba y las restantes fases la desarrollan los especialistas del centro DATEC:

- **Estudio preliminar y planeación:** Se realiza un estudio en la entidad cliente, con el fin de determinar lo que se desea construir y las condiciones que existen para el desarrollo y montaje de la misma. Se definen los objetivos y el alcance preliminar.
- **Requisitos:** Se realiza el proceso de entrevistas al cliente para determinar los requisitos de información. Se hace un levantamiento detallado de las fuentes de datos para validar la disponibilidad de la información. Además se definen los requisitos funcionales y no funcionales de la solución y se hace el análisis de los requisitos que dan paso al diseño e implementación.
- **Arquitectura:** Se define la arquitectura de la solución, la comunicación entre los subsistemas y las herramientas a utilizar.
- **Diseño e implementación:** Se define el diseño de las estructuras de almacenamiento de datos, se diseñan los procesos de integración de datos como el mapa lógico de datos, los cubos OLAP para la presentación de la información, así como el diseño gráfico de la aplicación definido por el cliente. Después se implementan cada uno de los subsistemas (repositorio de datos, integración de datos, presentación de datos).
- **Prueba:** Se realizan las pruebas que validan la calidad del producto, comenzando por las pruebas de unidad, las pruebas de integración y sistema, hasta llegar a las pruebas de aceptación con el cliente final.
- **Despliegue:** Se realiza un despliegue piloto, donde se configuran los servidores, se instalan las herramientas según la arquitectura definida, se carga una muestra de los datos en un ambiente controlado, con el fin de demostrarle al cliente final que la solución funciona. Una vez aceptada la solución por el cliente, se realiza la carga histórica de los datos, la capacitación y la transferencia tecnológica.

- **Soporte y mantenimiento:** Después de implantada la solución se brindan los servicios, que pueden ser soporte en línea, vía telefónica, web y correo, hasta el acompañamiento junto al cliente según el contrato firmado.
- **Gestión y administración del proyecto:** Esta fase se ejecuta a lo largo de todo el ciclo de vida del proyecto. Es aquí donde se controla, gestiona y chequea todo el desarrollo, los gastos, las utilidades, los recursos, las adquisiciones, los planes y cronogramas. (12)

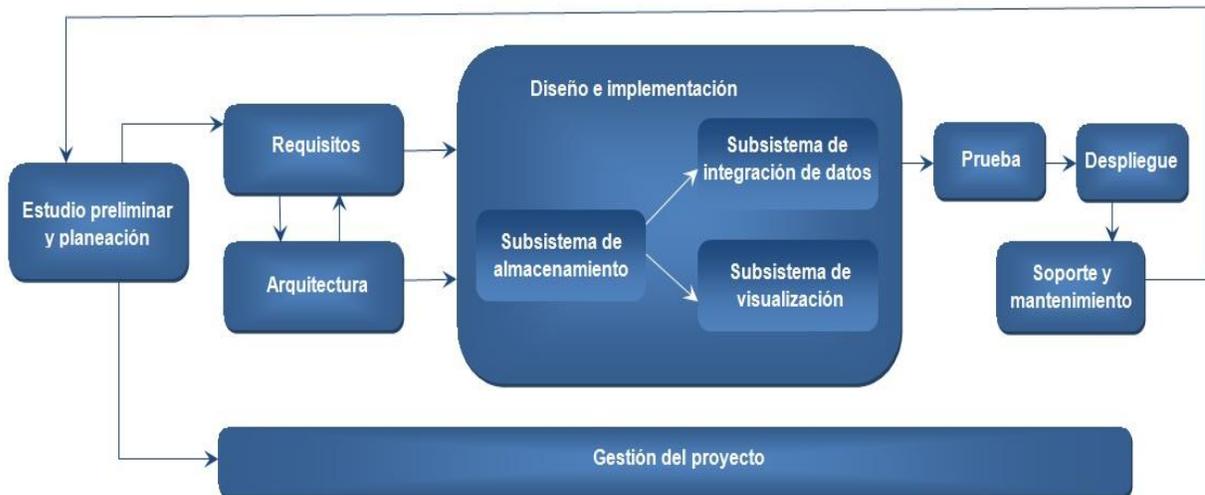


Figura 4. Fases de la Propuesta de metodología para el desarrollo de Almacenes de Datos en DATEC.

## 1.6 Herramientas a utilizar en la solución

En la actualidad se han desarrollado diversas herramientas para el desarrollo de los AD. A continuación se exponen las herramientas a utilizar en la solución del MD.

### 1.6.1 Herramienta para el modelado de los datos

Se definen a las Herramientas CASE como un conjunto de programas y ayudas que dan asistencia a los analistas, ingenieros de software y desarrolladores, durante todos los pasos del ciclo de vida de desarrollo de un software. (13)

En la investigación se utilizó para el diseño el Visual Paradigm 8.0. Es multiplataforma y está disponible en varios idiomas, además es fácil de instalar y fácil de actualizar.

### ✓ Visual Paradigm 8.0

Visual Paradigm es una herramienta UML profesional que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. Permite dibujar todos los tipos de diagramas de clases, generar código desde diagramas. También proporciona abundantes tutoriales de UML. Presenta licencia comercial.

Características principales:

- Soporte de UML versión 2.1.
- Diagramas de Procesos de Negocio.
- Ingeniería inversa - Código a diagrama.
- Generación de bases de datos - Transformación de diagramas de Entidad-Relación en tablas de base de datos.
- Exportación de ficheros XML. (13)

## 1.6.2 Sistema Gestor de Base de Datos

### ✓ PostgreSQL 9.1

PostgreSQL es un sistema gestor de bases de datos objeto-relacional, distribuido bajo licencia BSD (por sus siglas en inglés *Berkeley Software Distribution*, permite el uso del código fuente en software no libre) y con su código fuente disponible libremente. Es el sistema de gestión de bases de datos de código abierto más potente del mercado. (14)

**Características:**

- Se ejecuta en los sistemas operativos: Linux, Unix, Windows.
- Soporte de todas las características de una base de datos profesional (triggers, funciones, relaciones, reglas, vistas). (14)

PostgreSQL ofrece las siguientes ventajas:

- Estabilidad y confiabilidad.
- Herramientas gráficas de diseño y administración de bases de datos. (15)

## Herramienta de administración

### ✓ PgAdmin III 1.14.1

PgAdmin III 1.14.1 es una herramienta de código abierto para la administración de bases de datos PostgreSQL. En ella se incluyen:

- Interfaz administrativa gráfica.
- Herramienta de consulta SQL.

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 interfaz gráfica soporta todas las características de PostgreSQL y hace simple la administración. Está disponible en más de una docena de lenguajes y para los sistemas operativos como Windows y Linux. (16)

## 1.6.3 Herramientas para la Extracción, Transformación y Carga (ETL)

### ✓ Data Cleaner 1.5.4

El perfilado de datos es una de las primeras tareas a realizar en el proceso de calidad de los datos y consiste en realizar un análisis inicial sobre los datos fuentes, con el propósito de empezar a conocer su estructura, formato y nivel de calidad. Data Cleaner es una herramienta cuyo objetivo principal es detectar errores e inconsistencias en los datos. Por ejemplo, permite detectar entradas duplicadas, incompletas y a partir del análisis establecer reglas para corregirlas. Es una herramienta de código abierto y multiplataforma. Genera informes y gráficos que permiten a los usuarios identificar y analizar la estructura del origen de los datos y determinar el nivel de calidad de los mismos. (17)

### ✓ Pentaho Data Integration 4.2.1

Pentaho Data Integration es un motor junto con un conjunto de herramientas encargadas de los procesos de extracción, transformación y carga más conocido como los procesos ETL.

Pentaho Data Integration abre, limpia e integra la información y la pone en manos del usuario. Permite evitar grandes cargas de trabajo manual. Esta herramienta es de código abierto y sin costes de licencia. Las características son:

- Entorno gráfico orientado al desarrollo rápido basado en dos áreas: la de trabajo y la de diseño/vista.

- Uso de tecnologías estándar: Java y JavaScript.
- Fácil de instalar y configurar.
- Multiplataforma: Windows, Macintosh y Linux. (18)

#### 1.6.4 Herramientas de Inteligencia de negocio (BI)

##### ✓ Pentaho Schema Workbench 3.2

Es una herramienta de desarrollo que permite crear, modificar y publicar esquemas del cubo OLAP. Es un programa java multiplataforma. Permite crear todos los objetos que soporta Mondrian como esquemas, cubos, dimensiones y métricas. Esta herramienta tiene dos áreas: la zona de edición de las propiedades de cada elemento y la zona en la que se muestra la estructura jerárquica del esquema OLAP. (19)

El motor Mondrian procesa las expresiones multidimensionales (MDX) con los esquemas ROLAP. Permite modelar un XML (por sus siglas en inglés *eXtensible Markup Language*, lenguaje de marcas extensibles) que se crea en una estructura específica que utiliza el motor de Mondrian.

##### ✓ Pentaho BI Server 3.10

La Plataforma Pentaho BI provee el soporte y la infraestructura necesarios para crear soluciones de inteligencia empresarial a problemas de negocios. El marco proporciona los servicios básicos como autenticación y servicios web. La plataforma también incluye un motor de solución que integra reportes, análisis y componentes de minería de datos.

Esta herramienta permite a los usuarios de negocio analizar grandes cantidades de datos en tiempo real. Está diseñada para integrarse fácilmente en cualquier proceso de negocio. Es una herramienta de código abierto (20).

#### Conclusiones del capítulo:

Con esta investigación se arribó a las siguientes conclusiones:

- Se estudiaron los aspectos importantes de los AD y de los MD, sus componentes, la metodología para su desarrollo y las herramientas a utilizar, lo que permitió tener una mejor comprensión del proceso de desarrollo del MD.
- La Propuesta de metodología para el desarrollo de Almacenes de Datos en DATEC sirvió como guía de desarrollo y permitió tener una adecuada organización de las fases de desarrollo,

adquiriendo documentos en cada una de ellas, el cual contendrá información clave del proceso de desarrollo.

- El conjunto de herramientas seleccionadas permitió dar cumplimiento al desarrollo de la solución.

## **Capítulo 2: Análisis y diseño del Mercado de datos para el área de Tecnologías de la información y las comunicaciones.**

### **Introducción**

En este capítulo se realiza un estudio y análisis preliminar del negocio. Se especifican las necesidades de información, las reglas del negocio (RN) identificadas mediante entrevistas con el cliente, los requisitos funcionales (RF), los requisitos no funcionales (RNF) y los requisitos de información (RI), así como los casos de uso del sistema (CUS). Se realiza además, el diseño de los subsistemas de almacenamiento, integración y visualización del Mercado de datos para el área de Tecnologías de la información y las comunicaciones. De manera general se aborda el resultado del análisis y el diseño del MD.

### **2.1 Necesidad de información**

Se realizaron reuniones con la especialista del área de Tecnologías de la información y las comunicaciones de la ONEI para analizar los modelos: “Ingresos por tecnologías de la informática” (ver Anexo 1), “Indicadores de las Tecnologías de la información y las comunicaciones” (ver Anexo 2) y la “Encuesta anual de las Tecnologías de la información y las comunicaciones”. A continuación se detallan las necesidades de los usuarios identificadas para cada uno de los modelos mencionados:

El usuario necesita analizar la información del modelo 0009 “Indicadores de las Tecnologías de la información y las comunicaciones” por División Político Administrativa (DPA), por Nomenclador de la Actividad Económica (NAE), por entidad, por organismos, por tiempo y por indicador. Los datos que se solicitan en el modelo: plan real, real del mes, real acumulado, real del año anterior.

El usuario necesita analizar la información del modelo 0764 “Ingresos por tecnologías de la informática”, por DPA, por NAE, por organismos, por entidad, por indicador, por tiempo y por variante. Los datos que se solicitan en el modelo: cantidad de producción nacional en mp (miles de pesos), cantidad de importados en mp, cantidad de producción nacional en cuc (pesos convertibles), cantidad de importados en cuc, cantidad de exportación en cuc, cantidad de donaciones en cuc y cantidad de computadoras.

El usuario necesita analizar la información de la “Encuesta anual de las Tecnologías de la información y las comunicaciones”, por entidad, por tiempo, por preguntas y por organismo. Los datos que se solicitan en la encuesta: valor de respuesta.

## 2.2 Requisitos de información

Los RI describen la información que debe almacenar el sistema para satisfacer las necesidades de los usuarios (21). Para el desarrollo del Mercado de datos para el área de Tecnologías de la información y las comunicaciones se identificaron 12 RI, estos representan toda la información que debe estar disponible en la aplicación. De ellos se identificaron variables de entrada y de salida. Estos requisitos se encuentran en el artefacto **0113\_Especificación de Requisitos de Software**, del expediente de proyecto. A continuación se desglosan los RI:

### Requisitos de información:

- ✓ **RI\_1.** Obtener la cantidad de producción nacional en mp, por indicador, por tiempo, por organismos, por NAE, por DPA, por entidad.
- ✓ **RI\_2.** Obtener la cantidad de importados en mp, por indicador, por tiempo, por organismos, por NAE, por DPA, por entidad.
- ✓ **RI\_3.** Obtener la cantidad de producción nacional en cuc, por indicador, por tiempo, por organismos, por NAE, por DPA, por entidad.
- ✓ **RI\_4.** Obtener la cantidad de importados en cuc, por indicador, por tiempo, por organismos, por NAE, por DPA, por entidad.
- ✓ **RI\_5.** Obtener la cantidad de exportación en cuc, por indicador, por tiempo, por organismos, por NAE, por DPA, por entidad.
- ✓ **RI\_6.** Obtener la cantidad de donaciones en cuc, por indicador, por tiempo, por organismos, por NAE, por DPA, por entidad.
- ✓ **RI\_7.** Obtener la cantidad de computadoras, por indicador, por tiempo, por organismos, por NAE, por DPA, por entidad.
- ✓ **RI\_8.** Obtener el plan real por entidad, por tiempo, por organismos, por NAE, por DPA, por indicador, por variante.
- ✓ **RI\_9.** Obtener el real del mes por entidad, por tiempo, por organismos, por NAE, por DPA, por indicador, por variante.
- ✓ **RI\_10.** Obtener el real acumulado por entidad, por tiempo, por organismos, por NAE, por DPA, por indicador, por variante.
- ✓ **RI\_11.** Obtener el real del año anterior por entidad, por tiempo, por organismos, por NAE, por DPA, por indicador, por variante.

- ✓ **RI\_12.** Obtener el valor de respuesta por entidad, por tiempo, por preguntas, por organismos.

Una vez identificados los RI se concluye que estos presentan variables de entrada y variables de salida.

**Variables de entrada:** Son las aristas de análisis por la que va a ser filtrada y analizada la información.

**Variables de salida:** Son los indicadores a medir que representen las posibles medidas del AD.

Variables de entrada y de salida correspondiente a los RI del modelo “Indicadores de las TIC”:

Variables de entrada:

- ✓ tiempo
- ✓ DPA
- ✓ NAE
- ✓ organismo
- ✓ entidad
- ✓ indicador
- ✓ variante

Variables de salida:

- ✓ plan real
- ✓ real del mes
- ✓ real acumulado
- ✓ real del año anterior

Variables de entrada y de salida correspondiente a los RI del modelo “Ingresos por tecnologías de la informática”:

Variables de entrada:

- ✓ tiempo
- ✓ DPA
- ✓ NAE
- ✓ organismo
- ✓ entidad
- ✓ indicador

Variables de salida:

- ✓ cantidad de producción nacional en mp
- ✓ cantidad de importados en mp
- ✓ cantidad de producción nacional en cuc
- ✓ cantidad de importados en cuc
- ✓ cantidad de exportación en cuc
- ✓ cantidad de donaciones en cuc
- ✓ cantidad de computadoras

Variables de entrada y de salida correspondiente a los RI de la “Encuesta Anual de las TIC”:

Variables de entrada:

- ✓ entidad
- ✓ temporal
- ✓ pregunta
- ✓ organismo

Variables de salida:

- ✓ valor de respuesta

### 2.3 Requisitos funcionales

Los requisitos funcionales describen las actividades que el equipo de desarrollo debe construir. Estos incluyen las funcionalidades que deben implementarse en los tres subsistemas que se desarrollan en las soluciones de los AD, teniendo en cuenta las características del negocio (12). Se identificaron 15 RF, estos se encuentran descritos en su totalidad en el artefacto **0113\_Especificación de Requisitos de Software**, del expediente de proyecto. A continuación se enumeran algunos RF:

- RF 1 Crear metadatos asociados a los procesos de ETL como fecha de actualización de la última carga realizada al MD, nombre de la fuente y nombre del mercado.
- RF 2 Organizar la información mediante libros de trabajo.
- RF 3 Exportar la información a excel y a pdf.
- RF 4 Visualizar la información mediante gráficos y tablas.
- RF 5 Autenticar usuario.

- RF 6 Adicionar usuario.
- RF 7 Eliminar usuario.
- RF 8 Adicionar rol.
- RF 9 Eliminar rol.
- RF 10 Adicionar reporte.
- RF 11 Eliminar reporte.

#### **2.4 Requisitos no funcionales**

Los RNF son propiedades o cualidades que el producto debe tener (12). Luego de un análisis detallado se identificaron cinco RNF del Mercado de datos para el área de Tecnologías de la información y las comunicaciones, los cuales se encuentran descritos en su totalidad en el artefacto **0113\_Especificación de Requisitos de Software** dentro del expediente de proyecto. A continuación se enumeran algunos de ellos:

##### **Usabilidad:**

**RNF\_1.** Agilizar el acceso a los reportes del MD mediante la distribución de la información por áreas de análisis.

El usuario podrá acceder de manera rápida a la información que solicita en el área correspondiente de acuerdo al objetivo de su solicitud.

##### **Requisitos para la documentación de usuarios en línea y ayuda del sistema:**

**RNF\_2.** 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.

##### **Requisitos Legales, de Derecho de Autor y otros:**

**RNF\_3.** Entregar el sistema a la ONEI.

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

## 2.5 Reglas del negocio

Las RN describen restricciones, controlan la estructura y funcionamiento de una organización mediante políticas, medidas y operaciones, las cuales se deben seguir para cumplir las necesidades de los clientes (22). Las reglas del negocio del MD se identificaron a partir de las entrevistas con el cliente y del perfilado de los datos. A continuación se muestran:

### Reglas de almacenamiento:

- ✓ **RN1.** No pueden existir campos con valores negativos.
- ✓ **RN2.** Los campos de la columna empresa van a contener un máximo de 5 caracteres.

### Reglas de transformación:

- ✓ **RN3.** En el MD no pueden existir campos vacíos.
- ✓ **RN4.** Los id de las dimensiones no pueden tomar valores nulos.

### Reglas de visualización:

- ✓ **RN5.** Los valores numéricos se mostrarán a lo sumo con un lugar decimal.

## 2.6 Caso de uso del sistema

En el diagrama CUS se representa la relación que existe entre los casos de uso (CU) y los actores del sistema. Los CUS se dividen en dos categorías:

- Los casos de uso de información (CUI) que agrupan los RI teniendo en cuenta el criterio de tipo de información de negocio que manejan.
- Los casos de usos funcionales (CUF) que agrupan los RF en correspondencia con las actividades a realizar.

Posterior al análisis de los requisitos se diseña el diagrama de CUS (ver Figura 5).

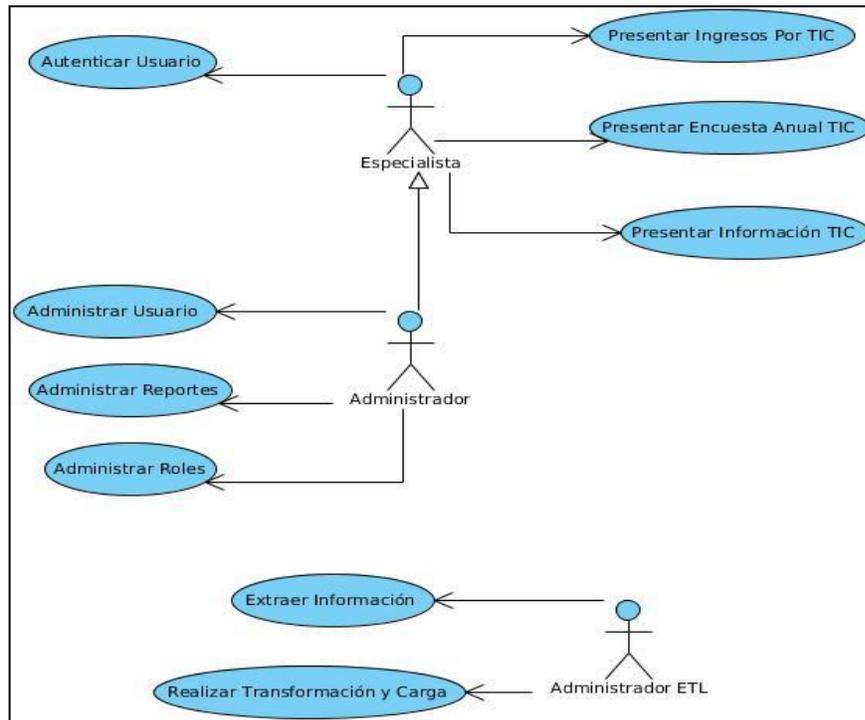


Figura 5. Diagrama de caso de uso del sistema.

### 2.6.1 Actores del sistema

**Administrador:** Es el responsable de la administración de los usuarios, roles y reportes en el AD.

**Administrador\_ETL:** Realiza la extracción de los datos de los ficheros fuentes, así como la transformación y carga de los datos hacia el MD.

**Especialista:** Consulta la información de los reportes disponibles en el sistema.

### 2.6.2 Especificación de Casos de uso

➤ **Casos de uso de información:**

- ✓ Presentar Ingresos Por TIC.
- ✓ Presentar Encuesta Anual TIC.
- ✓ Presentar Información TIC.

➤ **Casos de uso funcionales:**

- ✓ Extraer Información.
- ✓ Realizar Transformación y Carga.

- ✓ Autenticar Usuario.
- ✓ Administrar Usuario.
- ✓ Administrar Reportes.
- ✓ Administrar Roles.

En la Tabla 1 se presenta la descripción textual del CUI Presentar Información TIC. El resto de las descripciones textuales se encuentran en el artefacto **0114\_Especificación de casos de uso** del expediente de proyecto.

|   |   |  |
|---|---|--|
| <b>Objetivo</b>   | Presentar Información TIC.  |  |
| <b>Actores</b>  | Especialista.   |  |
| <b>Resumen</b>  | El CU inicia cuando el actor desea consultar información referente a los Indicadores de las TIC. El mismo selecciona los campos por los que quiere que se le muestre la información. Luego se muestra la información correspondiente a la selección. El CU finaliza cuando el actor analice la información. |  |
| <b>Complejidad</b>  | Alta.   |  |
| <b>Prioridad</b>  | Media.  |  |
| <b>Precondiciones</b>                                     | Los datos correspondientes fueron cargados al MD. El usuario tiene que estar autenticado y los reportes referentes a esa información tienen que haber sido creados.   |  |
| <b>Postcondiciones</b>                                    | Los reportes correspondientes fueron analizados por el actor.   |  |
| <b>Flujo de eventos</b>                                   |   |  |
| <b>Flujo básico: Presentar indicadores TIC</b>            |   |  |
|   | <b>Actor</b>  | <b>Sistema</b>   |
| 1.  | Selecciona SIGOB.   |  |
| 2.  |   | Muestra las áreas de análisis contenidas en SIGOB.   |
| 3.  | Selecciona el A.A TIC.  |  |
| 4.  |   | Muestra los libros de trabajo correspondientes al AA TIC.  |
| 5.  | Selecciona el L.T Indicadores de las TIC.   |  |
| 6.  |   | Muestra los reportes correspondientes a la selección.  |
| 7.  | Selecciona el reporte deseado.  |  |
| 8.  |   | Muestra la información contenida en el reporte seleccionado. Brinda las opciones que aparecen en la barra de herramientas. Finaliza el CU. |
| <b>Opciones de reportes de Presentar información TIC.</b> |   |  |
|   | <b>Perspectivas de análisis</b>   | <b>Posibles resultados</b>   |

|   | Medidas  | Periodicidad  |
|---|--|---|
| <ul style="list-style-type: none"> <li>• Tiempo</li> <li>• DPA</li> <li>• NAE</li> <li>• Organismo</li> <li>• Entidad</li> <li>• Indicador general</li> <li>• Variante</li> </ul> | <ul style="list-style-type: none"> <li>• plan real</li> <li>• real del mes</li> <li>• real acumulado</li> <li>• real del año anterior</li> </ul> | <ul style="list-style-type: none"> <li>• Mensual</li> </ul> |
| <b>Prototipo de interfaz de usuario</b>   |  |   |
| <b>Relaciones</b>   | <b>CU Incluidos</b>  | No aplica.  |
|   | <b>CU Extendidos</b>   | No aplica.  |
| <b>Requisitos no funcionales</b>  | Sección: “3.2 Requisitos no funcionales” del documento: 0114_Especificación de requisitos de software.   |   |
| <b>Asuntos pendientes</b>   | No aplica.   |   |

Tabla 1. CUI Presentar Información TIC.

## 2.7 Arquitectura del sistema

Para representar la arquitectura de un MD se debe tener en cuenta el origen de los datos, la comunicación y los subsistemas. La arquitectura general del Mercado de datos para el área de Tecnologías de la información y las comunicaciones (ver Figura 6) está compuesta por las fuentes de datos (fichero “M9015\_22.mdb” y los modelos 0009 y 0764 de la base de datos del SIGE) y por los subsistemas: el subsistema de integración, el subsistema de almacenamiento y el subsistema de visualización:

- **Fuente de datos:** Se refiere al origen de los datos.
- **Subsistema de integración:** Incluye los procesos que permiten que los datos sean extraídos de las fuentes, transformados e integrados.
- **Subsistema de almacenamiento:** Almacena toda la información correspondiente del Mercado de datos para el área de Tecnologías de la información y las comunicaciones.

- **Subsistema de visualización:** Se encarga de consultar los datos existentes en el Mercado de datos para el área de Tecnologías de la información y las comunicaciones, para luego representarlos mediante gráficos y tablas contribuyendo a la toma de decisiones.



Figura 6. Arquitectura del sistema.

El subsistema de integración obtiene los datos de diferentes fuentes (archivo "M9015\_22.mdb" y de la base de datos del SIGE) y se encarga de los procesos que integran y transforman la información para su carga.

El subsistema de almacenamiento contiene los datos cargados en los procesos de ETL. La información es almacenada en el SGBD PostgreSQL.

El subsistema de visualización permite mostrar la información en forma de reportes. Los usuarios que acceden a este subsistema son los autorizados a la visualización de los reportes.

## 2.8 Diseño del subsistema de almacenamiento

Para el correcto funcionamiento del MD se realiza el modelo dimensional de los datos, el cual contiene las tablas de hechos identificadas en el negocio, las dimensiones y las relaciones que existen entre ellas.

### 2.8.1 Dimensiones

Las tablas de dimensiones representan las características de un hecho, las cuales están descritas por un conjunto de atributos organizados en jerarquías. Cada dimensión identificada en el estudio de la investigación se muestra a continuación junto con la jerarquía:

1. **Dimensión pregunta** (dim\_pregunta): Describe el universo de valores bajo los cuales puede clasificarse la información atendiendo a la respuesta de la encuesta.
2. **Dimensión dpa** (dim\_dpa): Describe el universo de valores bajo los cuales puede clasificarse la información atendiendo a la situación geográfica. Presenta un nivel provincia y dentro de esta se encuentran los municipios.  
Provincia → Municipio.
3. **Dimensión temporal mes** (dim\_temporal\_mes): Describe el universo de valores bajo los cuales puede clasificarse la información atendiendo los valores anuales, semestrales, trimestrales y mensuales. Presenta los niveles año, semestre, trimestre y mes.  
Año → Semestre → Trimestre → Mes.
4. **Dimensión nae** (dim\_nae): Describe el universo de valores bajo los cuales puede clasificarse la información atendiendo Nomencladores de Actividad Económica. Presenta los niveles sección, división y clase.  
Sección → División → Clase.
5. **Dimensión organismo** (dim\_organismo): Describe el universo de valores bajo los cuales puede clasificarse la información atendiendo al organismo que la emite.
6. **Dimensión entidad** (dim\_entidad): Describe el universo de valores bajo los cuales puede clasificarse la información de acuerdo con el centro que emite la misma.
7. **Dimensión indicador general** (dim\_indicador\_general): Describe el universo de valores bajo los cuales puede clasificarse la información atendiendo a los indicadores que se recogen del modelo 0009 y el modelo 0764. Presenta el nivel temático, subtemática e indicador.  
Temática → Subtemática → Indicador.
8. **Dimensión variante** (dim\_variante): Describe el universo de valores bajo los cuales puede clasificarse la variante según la empresa que la emite.

### 2.8.2 Hechos y medidas

Las tablas de hechos representan un objeto o evento del negocio para analizar el comportamiento de los datos de una empresa. Estas tablas contienen las medidas que son valores numéricos que serán analizadas por los usuarios y las llaves asociadas a cada una de las dimensiones con que se relacionan. En la investigación se identificaron los siguientes hechos y medidas:

1. **Hecho información tic** (hech\_informacion\_tic): Este hecho contiene la información de los Ingresos por tecnologías de la informática generados en el modelo 0764. Las medidas relacionadas con este hecho son las siguientes: plan real, real del mes, real acumulado y real del año anterior.
2. **Hecho ingresos tic** (hech\_ingresos\_por\_tic): Este hecho contiene la información de los Indicadores de las TIC generados en el modelo 009. Las medidas relacionadas con este hecho son las siguientes: cantidad de producción nacional en mp, cantidad de importados en mp, cantidad de producción nacional en cuc, cantidad de importados en cuc, cantidad de exportación en cuc, cantidad de donaciones en cuc y cantidad de computadoras.
3. **Hecho encuesta anual** (hech\_encuesta\_anual\_tic): Este hecho contiene la información de la Encuesta Anual TIC. La medida relacionada con este hecho es la siguiente: valor de respuesta.

### 2.8.3 Matriz bus

La matriz bus describe la relación que existe entre las dimensiones y los hechos. En ella se verifica que no exista solapamiento entre los hechos y se identifican las dimensiones comunes de los hechos. Las columnas de la tabla representan los hechos identificados en el MD y las filas las dimensiones. Las celdas marcadas con una X indican que la fila correspondiente a esa dimensión está relacionada con la columna de los hechos. En la Tabla 2 se muestra la matriz bus.

| Dimensiones / Hechos  | hech_informacion_tic | hech_ingresos_por_tic | hech_encuesta_anual_tic |
|-----------------------|----------------------|-----------------------|-------------------------|
| dim_pregunta          |                      |                       | X                       |
| dim_dpa               | X                    | X                     |                         |
| dim_temporal_mes      | X                    | X                     | X                       |
| dim_nae               | X                    | X                     |                         |
| dim_organismo         | X                    | X                     | X                       |
| dim_entidad           | X                    | X                     | X                       |
| dim_variante          | X                    |                       |                         |
| dim_indicador_general | X                    | X                     |                         |

Tabla 2. Matriz bus

La matriz bus permitió conocer las dimensiones compartidas para los tres hechos identificados en el negocio, en este caso se encuentran las dimensiones temporal mes, organismo y entidad. También arrojó

como resultado que no existen dos hechos que compartan exactamente las mismas dimensiones del MD por lo que se demostró que no existe solapamiento entre los mismos.

### 2.8.4 Modelo de datos

Una vez definida la matriz bus se procede a diseñar el modelo de datos. El principal objetivo está dirigido a que los datos del negocio queden representados en una estructura lógica, evidenciando las relaciones que existen entre las tablas de hechos y las dimensiones. A continuación se muestra el modelo de datos diseñado para el MD, donde se evidencia el uso de la topología constelación de hechos, atendiendo que existen tablas de hechos que comparten algunas dimensiones. En él se crearon tres tablas de hechos y ocho tablas de dimensiones relacionados entre sí. Para observar el modelo de datos en su totalidad puede ser consultado el artefacto **Plantilla Especificación del modelo de datos** del expediente de proyecto. En la Figura 7 se muestra un fragmento del modelo de datos.

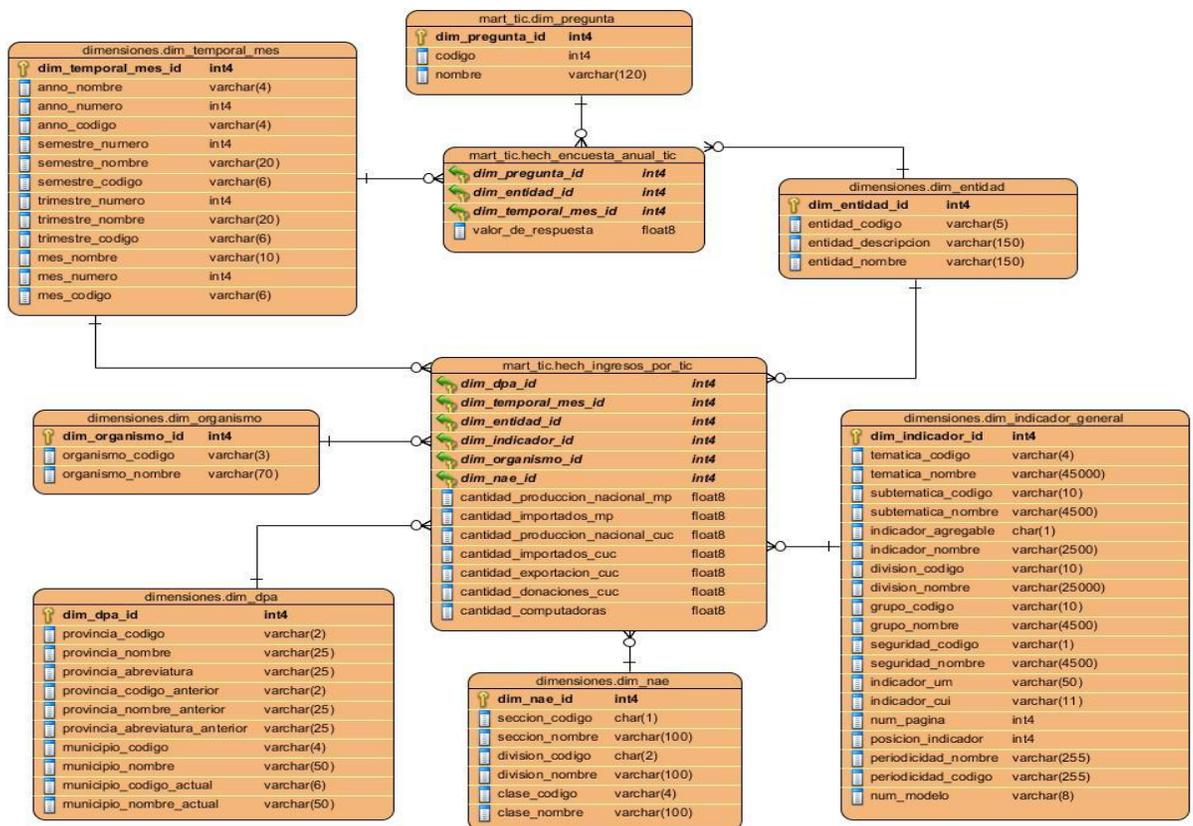


Figura 7. Modelo de datos.

## 2.9 Diseño de los procesos de integración de datos

El diseño del subsistema de integración lleva implícito realizar un análisis de los datos provenientes de la fuente, esto se conoce como perfilado de datos y el diseño de las transformaciones que al ser ejecutado permiten que los datos se encuentren disponibles en una tabla de salida.

### 2.9.1 Perfilado de datos

El perfilado de datos es el proceso que se encarga de analizar las fuentes para conocer el estado que se encuentran los datos, así como la calidad y la estructura de los mismos. (23) La herramienta que se utilizó fue el Data Cleaner para verificar la existencia de valores nulos, duplicados, distintos y para identificar la cantidad de caracteres que van poseer los campos de las tablas. A continuación se describen los resultados arrojados por la herramienta Data Cleaner:

- ✓ Una vez realizado el perfilado, se pudo observar que la fuente de datos de la encuesta tiene formato de fecha MM/dd/yyyy y permitió conocer que la columna “emp” contiene un máximo de 5 caracteres. También arrojó como resultado que la fuente no contenía valores nulos.
- ✓ Los modelos que se encuentran en el SIGE no contienen valores nulos.

### 2.9.2 Diseño de las transformaciones

Una vez que se conoce la estructura de los datos, se procede a realizar el diseño de las transformaciones. A continuación se muestra el diseño de las transformaciones para la dimensión dim\_pregunta (ver Figura 8) y los hechos hech\_encuesta\_anual\_tic (ver Figura 9), hech\_ingresos\_por\_tic y hech\_informacion\_tic (ver Figura 10).

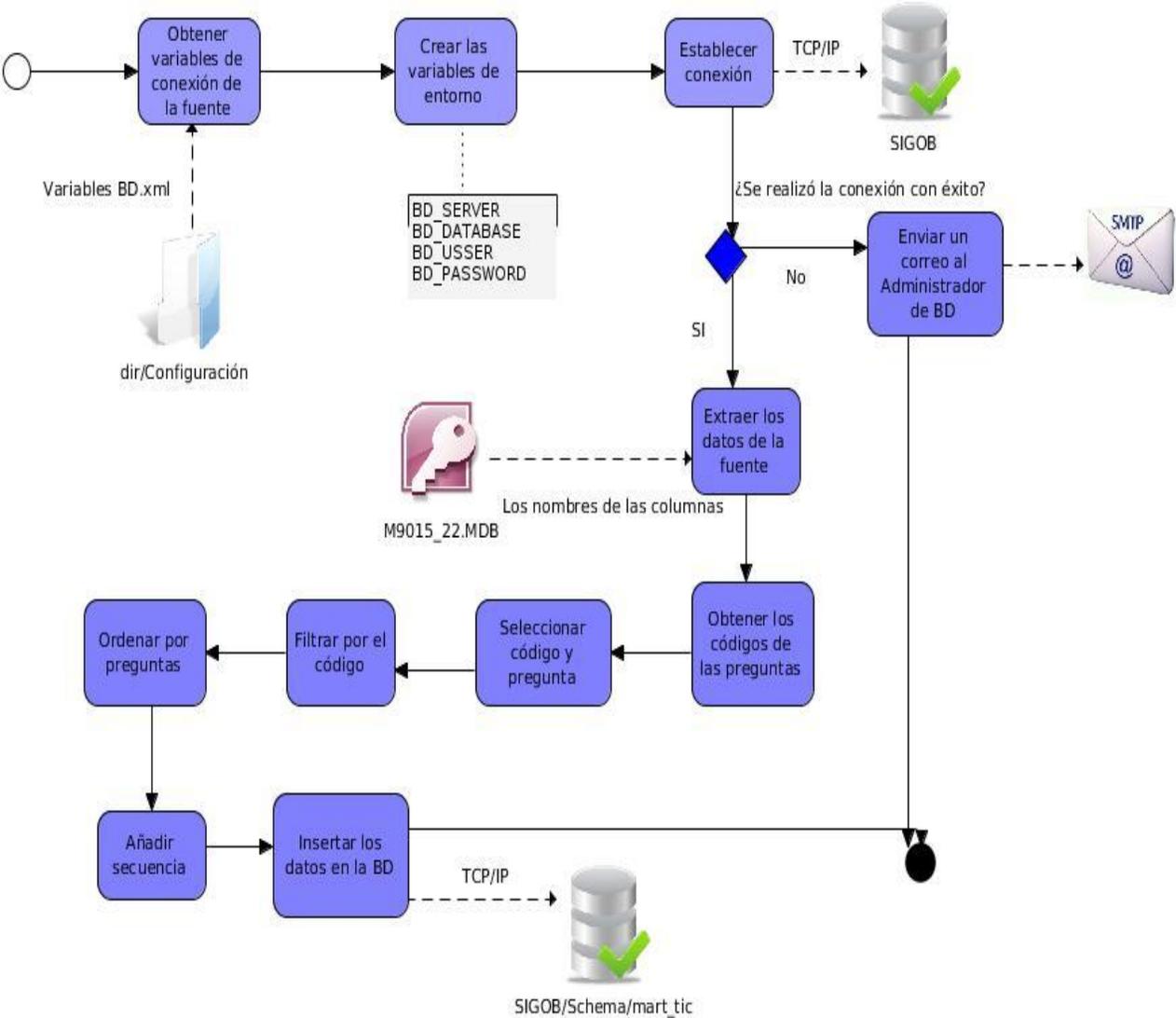


Figura 8. Diseño de la transformación dim\_pregunta.

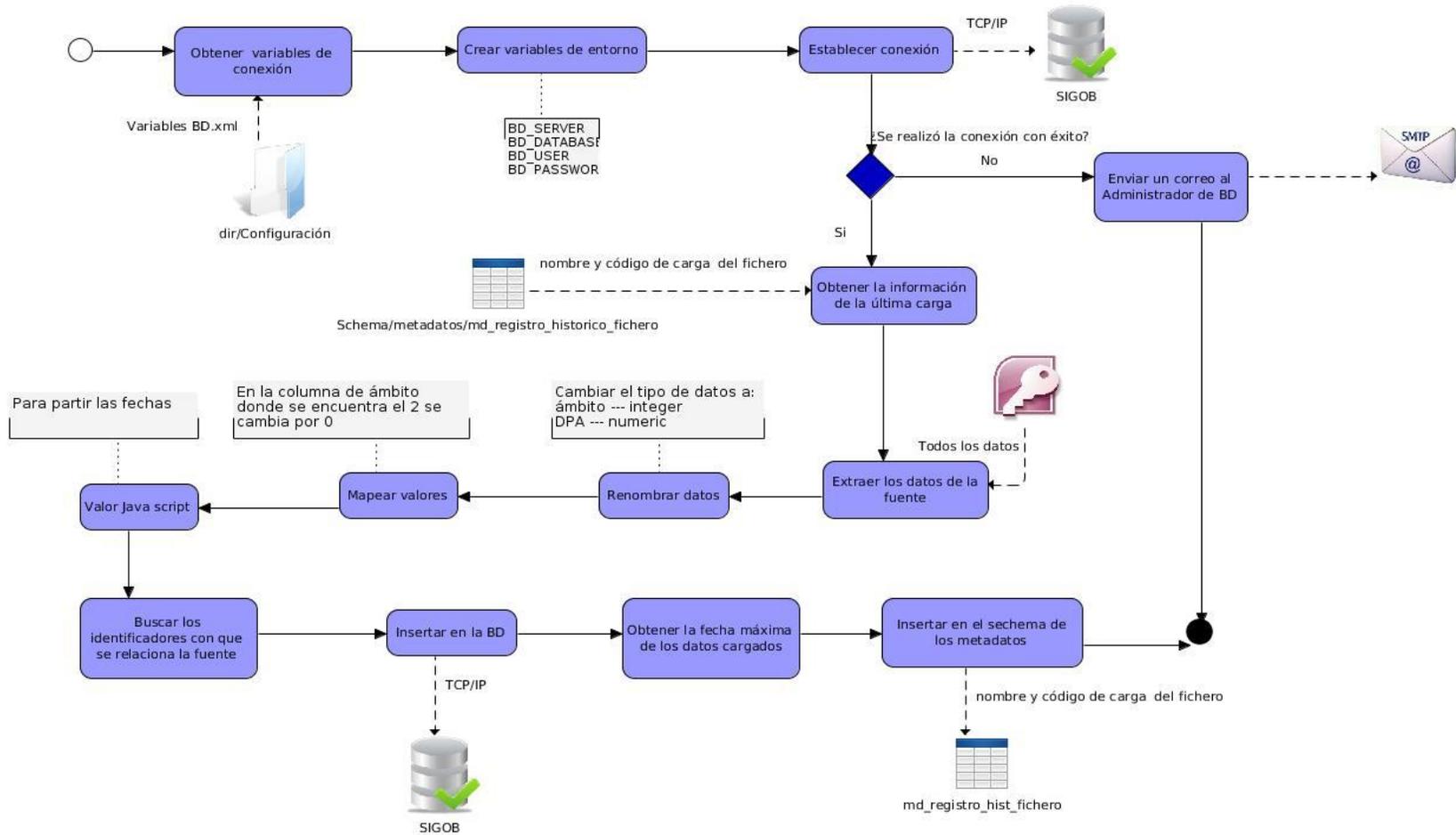


Figura 9. Diseño de las transformaciones hech\_encuesta\_anual\_tic.

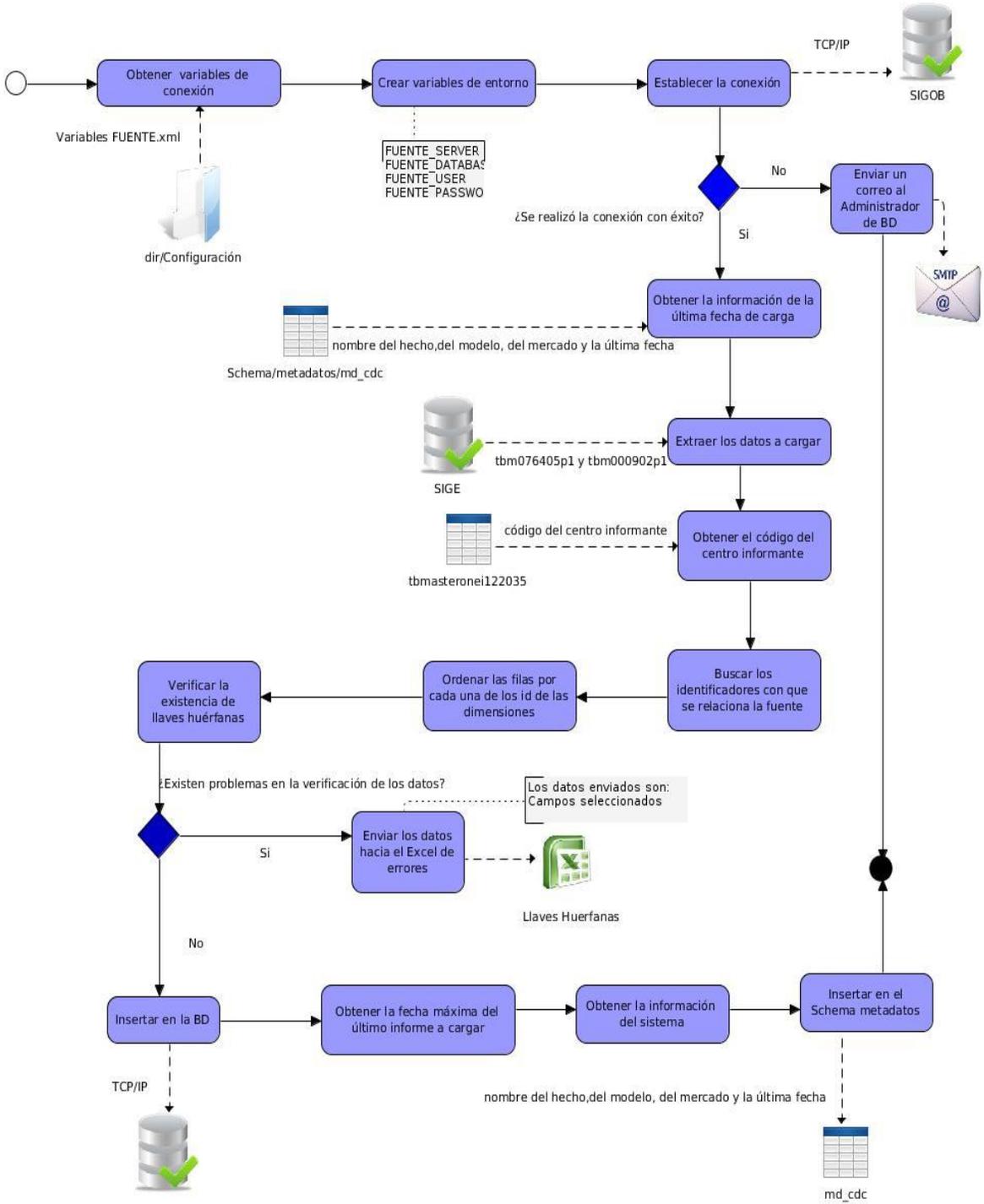


Figura 10. Diseño de las transformaciones hech\_ingresos\_por\_tic y hech\_informacion\_tic.

A continuación se explica el diseño de la transformación del hecho información tic. El diseño comienza obteniendo las variables de conexión que se encuentra en la siguiente dirección *Configuración/Variables/FUENTE.xml*, creando para cada una de ellas las variables de entorno. Seguido de crear estas variables se establece la conexión a la base de datos de SIGOB obteniendo la fuente de datos del mismo. Luego se obtiene la información de la última fecha de carga en la tabla md\_cdc del esquema metadatos. Se obtienen los datos a cargar de la tabla tbm076405p1 del esquema mod\_modelos. Posterior se busca los identificadores de cada una de las tablas de las dimensiones. Una vez capturados los id se verifica la existencia de llaves huérfanas. En caso de que existan se envían hacia un Excel de error para su posterior análisis y los demás datos que no presenten errores serán cargados al MD en la tabla de hecho hech\_informacion\_tic.

## 2.10 Diseño del subsistema de visualización

El diseño del subsistema de visualización comprende la creación de los cubos OLAP para el análisis multidimensional de la información. Se definen los reportes que serán implementados como parte de la solución.

### 2.10.1 Diseño de los cubos OLAP

En la investigación se definen los cubos multidimensionales: hech\_encuesta\_anual\_tic, hech\_informacion\_tic, hech\_ingresos\_por\_tic. Los cubos fueron especificados para cada uno de los hechos existentes, pues constituyen la fuente de información fundamental de la organización de los datos. El MD contiene ocho dimensiones, doce medidas físicas y ocho reportes, distribuido de la siguiente manera:

El cubo **Ingresos por TIC** contiene seis dimensiones utilizables, siete medidas físicas y dos reportes.

El cubo **Información TIC** contiene cinco dimensiones utilizables, cuatro medidas físicas y dos reportes.

El cubo **Encuesta Anual TIC** contiene dos dimensiones utilizables, una medida física y cuatro reportes.

En la Figura 11 se muestra el cubo multidimensional del hecho hech\_informacion\_por\_tic.



Figura 11. Diseño del cubo hec\_h\_informacion\_tic.

### 2.10.2 Diseño de la arquitectura de información

La arquitectura de información permite tener una visualización respecto a los elementos que estructuran el sistema y serán mostrados en la capa de visualización del MD. En el Mercado de datos para el área de Tecnologías de la información y las comunicaciones se identificó un Área de Análisis General, un Área de Análisis (AA), que se divide en tres Libros de Trabajos (LT), los cuales contienen un grupo de reportes. A continuación se describen cada uno de los elementos que corresponden a la estructura de navegación.

Dentro del Área de Análisis General SIGOB se encuentra el A.A TIC, el cual presenta los siguientes L.T: L.T Ingresos por TIC, L.T Indicadores de las TIC y L.T Encuesta anual TIC, con un total de 8 reportes (ver Figura 12).



Figura 12 Arquitectura de información.

### 2.11 Política de respaldo y recuperación

Para establecer las políticas de respaldo y recuperación se tuvo en cuenta los siguientes aspectos fundamentales:

- **Periodicidad de las salvallas:** Las salvallas de toda la información contenida en la base de datos se realizan mensuales y trimestrales para los modelos 0764 y 0009. Para las salvallas de la información de la encuesta anual de las TIC se realizan anualmente.
- **Tablas involucradas:** Las tablas involucradas en el proceso son tres tablas de hechos y ocho tablas de dimensiones identificadas en la fase de análisis.
- **Backup o salvallas existentes:** En la ONEI se realizan backup mensuales a la base de datos y se realizan salvallas diariamente a la información.

### 2.12 Esquemas de seguridad

La seguridad del Mercado de datos para el área de Tecnologías de la información y las comunicaciones, está determinada por los niveles de acceso al sistema. Dicha seguridad se rige por los roles y permisos que los usuarios poseen en su interacción con la base de datos y la aplicación.

#### 2.12.1 Seguridad en el subsistema de almacenamiento

Para la seguridad de la base de datos se definió los siguientes roles con sus permisos:

| Roles             | Permisos  |
|-------------------|---|
| Administrador     | Realiza la administración de la base de datos relacional que contiene todos los esquemas del AD. Tiene todos los permisos sobre la base de datos.   |
| Administrador ETL | Es el responsable de realizar los procesos de Extracción, Transformación y Carga sobre los datos. Tiene permisos de lectura y escritura sobre los esquemas pertenecientes al Mercado de datos para el área de Tecnologías de la información y las comunicaciones. |

Tabla 3. Seguridad en la base de datos.

### 2.12.2 Seguridad en el subsistema de integración

En los procesos de ETL se garantiza la seguridad a nivel de sistema operativo, en este caso utiliza GNU/Linux. Este sistema operativo permite asignar permisos a los archivos para diferentes usuarios, mediante la opción de solo lectura en las propiedades de la carpeta donde se almacena todos los datos fuentes, las transformaciones y los trabajos.

### 2.12.3 Seguridad en el subsistema de visualización

Para la seguridad en la aplicación se definió los siguientes roles con sus permisos:

| Roles         | Permisos  |
|---------------|---|
| Administrador | Tiene acceso total en el AD y además es el responsable de administrar los permisos y los roles a cada usuario.                      |
| Especialista  | Tiene acceso de solo lectura al A.A TIC. Puede visualizar los reportes contenidos en cada uno de los LT pertenecientes a esta área. |

Tabla 4. Seguridad en la aplicación.

### Conclusiones del capítulo:

En este capítulo se realizó el análisis y diseño del Mercado de datos para el área de Tecnologías de la información y las comunicaciones, el cual arrojó las siguientes conclusiones:

- Se identificaron 12 RI y 15 RF agrupados en 3 CUI y 6 CUF, sirviendo de base para elaborar el diagrama de caso de uso del sistema; además se identificaron 5 RN lo que permitió definir las reglas de transformación que serán utilizadas en los procesos de integración. Así se obtiene de manera más clara la necesidad de información del cliente.
- La arquitectura diseñada contribuyó a definir los tres componentes fundamentales que formarán parte del MD lo que proporcionó una vista general de la solución.
- A través del diseño del subsistema de almacenamiento se determinaron los elementos que formarán parte del modelo físico de datos, se identificaron tres hechos, ocho dimensiones y doce medidas.
- El diseño del subsistema de integración realizado sirvió de apoyo para la implementación de los procesos de ETL.

- Luego de diseñado el subsistema de visualización quedó definida la estructura de la capa de presentación. De igual manera se diseñaron tres cubos OLAP que constituyen la fuente de información fundamental de la organización de los datos.

**Capítulo 3: Implementación del Mercado de datos para el área de Tecnologías de la información y las comunicaciones.**

**Introducción**

En este capítulo se realiza la implementación de cada uno de los subsistemas, el subsistema de almacenamiento, el subsistema de integración y el subsistema de visualización que conforman el Mercado de datos para el área de Tecnologías de la información y las comunicaciones. Se realiza la estandarización de los nombres para lograr un patrón en los términos utilizados y así entender las estructuras del MD.

**3.1 Implementación del subsistema de almacenamiento**

La implementación del subsistema de almacenamiento incluye la estandarización de los nombres, la estructura física de almacenamiento y el indexado.

**3.1.1 Estandarización de los nombres**

La estandarización de los nombres tiene como objetivo lograr un patrón que tribute a la correcta normalización de los términos utilizados (ver Tabla 5). Esto permite a los desarrolladores entender las estructuras del MD.

| Tipo de Objeto | Función                 | Nomenclatura    | Descripción  |
|----------------|-------------------------|-----------------|--|
| Esquema        | dimensiones compartidas | dimensiones     | Esquema donde se organizan las dimensiones compartidas por varios MD.  |
|                | esquemas                | mart_[temática] | Esquema donde se almacenan las tablas de los hechos y vistas materializadas definidas para gestionar los datos asociados a cada área temática. |
| Tablas         | dimensiones             | dim_[nombre]    | Tablas dimensionales utilizadas como perspectivas de análisis.   |
|                | hechos                  | hech_[nombre]   | Tabla de hechos que definen las principales medidas requeridas   |

|  |           |             |  |
|--|-----------|-------------|--|
|  |           |             | para calcular indicadores.   |
|  | metadatos | md_[nombre] | Tabla de metadatos donde se guarda la información correspondiente a los datos del mercado. |

Tabla 5. Estándares de codificación del Mercado de datos para el área de Tecnologías de la información y las comunicaciones.

### 3.1.2 Estructura de los datos

La estructura de los datos constituye una forma de organizar los datos con el objetivo de facilitar el trabajo y la correcta manipulación de los mismos. Los esquemas representan de forma organizada toda la información contenida en la base de datos. En las tablas se guardan los datos obtenidos de las fuentes, las cuales están compuestas por campos que corresponden al nombre de la columna y el tipo de dato. La solución cuenta con los siguientes esquemas:

- ✓ **dimensiones:** Está compuesta por las tablas de dimensiones que son comunes para todos los MD de SIGOB.
- ✓ **mart\_tic:** Está compuesta por las tablas de dimensiones y las tablas de hechos propias del mercado.
- ✓ **metadatos:** Está compuesta por las tablas encargadas de almacenar los procesos de metadatos en el MD.
- ✓ **correlacionadores:** Este esquema está compuesto por la tabla correlacionadores, la cual contiene el código correspondiente a cada uno de los modelos del AD.

En la Tabla 6 se muestra la relación entre las tablas y los esquemas del MD.

| Esquemas           | Tablas                |
|--------------------|-----------------------|
| <b>dimensiones</b> | dim_indicador_general |
|                    | dim_dpa               |
|                    | dim_temporal_mes      |
|                    | dim_nae               |
|                    | dim_organismo         |
|                    | dim_variante          |

|                          |                          |
|--------------------------|--------------------------|
|                          | dim_entidad              |
| <b>mart_tic</b>          | dim_pregunta             |
|                          | hech_ingresos_por_tic    |
|                          | hech_informacion_tic     |
|                          | hech_encuesta_anual_tic  |
| <b>metadatos</b>         | md_carga_hist            |
|                          | md_registro_hist_fichero |
|                          | md_temporal              |
|                          | md_transformacion        |
|                          | md_mercado               |
|                          | md_cdc                   |
| <b>correlacionadores</b> | correlacionadores        |

Tabla 6. Estructura del Mercado de datos para el área de Tecnologías de la información y las comunicaciones.

### 3.1.3 Estrategias de indexado

Sobre un MD se realizan consultas y en muchos casos son complejas, pues solicitan información cumpliendo determinados criterios de búsqueda. Estas incluyen operaciones de *join* entre tablas, lo que demora en tiempo de respuesta. Para realizar estas operaciones se han investigado y creado técnicas especializadas que hoy ofrecen varios gestores, como los índices.

Los índices se utilizan para reducir los tiempos de ejecución de las consultas. En la investigación para realizar la búsqueda de los datos y las uniones entre las tablas, se utilizarán las llaves primarias y foráneas. Todas ellas tendrán un índice *B-Tree* (Árbol B) que posee el gestor PostgreSQL, de esta manera cualquier búsqueda que se realice permite disminuir el tiempo de respuesta de las consultas mediante el uso de las llaves anteriormente mencionadas. (24)

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

El proceso de integración consta de tres etapas fundamentales relacionadas entre sí: extracción, transformación y carga. No es recomendable iniciar estos procesos sin haber realizado el perfilado de datos a la fuente con la que se va a trabajar, para identificar los problemas como errores en los datos, valores nulos y duplicados. A continuación se describen los subsistemas que se utilizaron durante la implementación de la solución identificados por Ralph Kimball:

- ✓ **Perfilado de datos:** Este subsistema permite explorar los datos para verificar su calidad.
- ✓ **Captura del cambio de los datos:** Detecta los cambios correspondientes a las fechas que permitirán realizar las cargas de la información: mensuales, trimestrales y anuales.
- ✓ **Extracción:** Este subsistema permite la extracción de los datos desde las fuentes de origen. Para ello se tuvo presente la información relacionada con cada uno de los hechos y dimensiones.
- ✓ **Transformación:** Este subsistema permite realizar transformaciones como el mapeo de valores, la normalización de las tablas, las búsquedas de datos en tablas, filtrado de valores.
- ✓ **Carga:** Este subsistema permite realizar la carga de los datos a las tablas de dimensiones y hechos del MD.
- ✓ **Cambio lento de las dimensiones (SCD):** Implementa la lógica para crear atributos de variabilidad lenta a lo largo del tiempo. Fue utilizado el tipo 2 de SCD para dar tratamiento al cambio de la información asociada a la dimensión pregunta.
- ✓ **Llave subrogada:** Permite crear claves subrogadas independientes para cada tabla.
- ✓ **Programador de trabajos:** Permite gestionar los procesos de integración mediante los trabajos.
- ✓ **Repositorio de metadatos:** Captura los metadatos de los procesos de ETL y de los aspectos técnicos.
- ✓ **Rastreo de eventos de errores:** Captura todos los errores que proporcionan información valiosa sobre los datos. (25)

Los datos que serán cargados al mercado serán proporcionados por SIGE, los cuales serán gestionados por el subsistema de extracción a través de las tablas tbm000902p1 y tmb076405 que se corresponde con el modelo 0009 Indicadores de las TIC y el modelo 0764 Ingresos por tecnologías de la informática respectivamente. Estas tablas contienen la información referente a las Tecnologías de la información y las comunicaciones, que le proveen diferentes entidades de todos los municipios del país a la ONEI. También fueron extraídos de la base de datos de Access el fichero "M9015\_22.mdb", el cual contiene los datos de la Encuesta Anual de las TIC.

Luego de extraídos los datos se empiezan a realizar las transformaciones como mapeo de valores, normalización, seleccionar y renombrar, según las reglas de negocio identificadas; quedando listos para ser almacenados en las tablas de hechos y dimensiones del MD. Una vez cargados en la base de datos son mostrados al cliente para su posterior análisis.

### Implementación de las transformaciones

Las transformaciones se realizaron teniendo en cuenta las reglas de negocio identificadas. Se confeccionaron cuatro transformaciones, una para la carga de la dimensión y tres para los hechos propios del MD. Todas estas transformaciones tienen como objetivo poblar el MD.

En la Figura 13 se muestra el flujo de la transformación que posee los pasos para la carga del hecho hech\_informacion\_tic. En este flujo se hace uso del esquema metadatos, el cual posibilita la gestión de la captura del cambio de los datos, detectando los cambios correspondientes a las fechas que permitirán realizar las cargas. Además se hace uso del subsistema de extracción para obtener los datos fuentes y del subsistema de transformación quedando listos los datos para ser almacenados en la base de datos. El rastreo de errores es otro de los subsistemas utilizados, el cual permite capturar todos los datos que presentan errores una vez realizadas las transformaciones, enviándolos a un Excel de errores correspondientes al hecho hech\_informacion\_tic, que serán tratados posteriormente por el cliente. Mediante el subsistema de carga se procede a insertar los datos ya transformados e integrados al MD en las tablas de hechos y dimensiones correspondientes. Finalmente se hace uso del repositorio de metadatos para obtener los metadatos de los procesos de ETL.

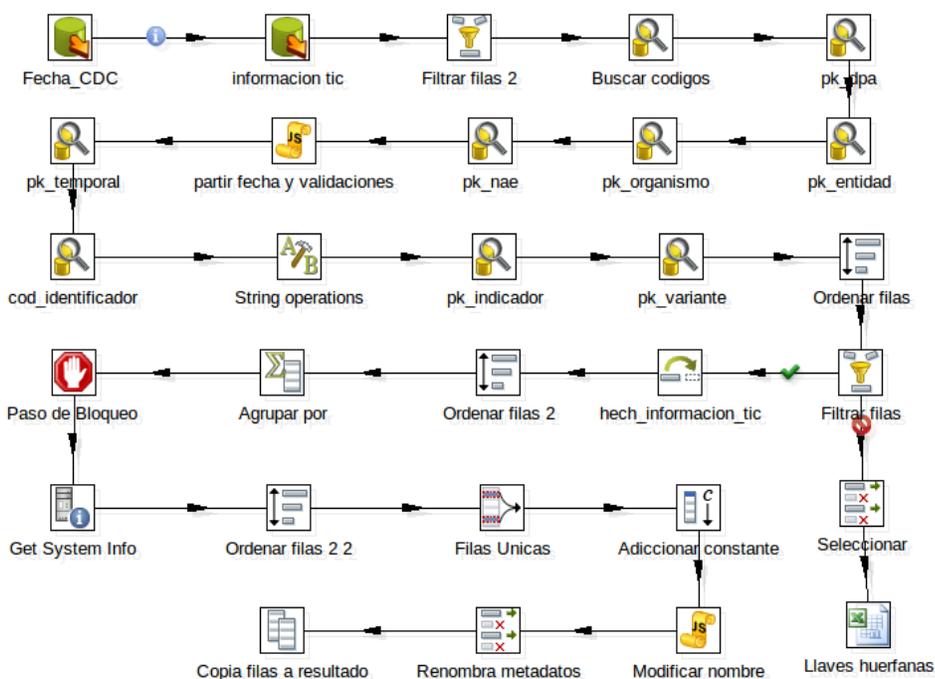


Figura 13. Transformación del hecho hech\_informacion\_tic.

### Implementación del trabajo

El trabajo es utilizado para realizar la carga de los datos de forma organizada, permitiendo ejecutar las transformaciones que han sido diseñadas. Se realizaron en la investigación tres trabajos, uno para la carga de los hechos: hech\_informacion\_tic y hech\_ingresos\_por\_tic (ver Figura 14), el siguiente se realizó para la carga de la dimensión dim\_pregunta y el hecho hech\_encuesta\_anual\_tic (ver Figura 15) y por último un trabajo general (ver Figura 16) que ejecuta los anteriores trabajos mencionados.



Figura 14. Implementación del trabajo para los hechos hech\_informacion\_tic y hech\_ingresos\_por\_tic.



Figura 15. Implementación del trabajo para la encuesta.



Figura 16. Implementación general.

### 3.2.2 Gestión del cambio lento en las dimensiones

Las dimensiones de cambio lento o SCD (*por sus siglas en inglés Slowly Changing Dimensions*) determinan cómo se manejan los cambios históricos en las tablas de dimensiones. Estos cambios pueden ocurrir de forma ocasional o constante:

Inicialmente Ralph Kimball plantea en su libro “*The Data Warehouse ETL Toolkit*” tres estrategias a seguir para las SCD, estas son las SCD tipo 1, 2 y 3. Con el transcurso del tiempo se ha profundizado en el

estudio de las mismas y se han creado nuevas SCD como los tipos 0, 4 y 6. A continuación se describen las mismas (23):

**SCD Tipo 0:** No se tiene en cuenta la gestión histórica y no se realiza esfuerzo alguno para lidiar con los problemas del cambio de la dimensión.

**SCD Tipo 1 (sobrescribir):** Es utilizado cuando la información histórica no es importante. Este tipo sobrescribe los datos antiguos con nuevos. Es utilizado mayormente para corregir errores de datos en las dimensiones. A pesar de ser fácil de implementar presenta como desventaja principal que no permanece ningún registro histórico en la dimensión.

**SCD Tipo 2 (añadir fila):** Cuando hay un cambio se crea una nueva entrada en la tabla. Al nuevo registro se le asigna una nueva llave subrogada y a partir de este momento será el valor usado para futuras entradas, las antiguas usarán el valor anterior. En este modo se gestiona un versionado que puede incluir fechas para indicar los períodos de validez, así como numeradores de registros o indicadores de registros activos o no. Este tipo permite guardar toda la información histórica en el AD.

**SCD Tipo 3 (añadir columna):** Esta estrategia requiere que se agregue una nueva columna a la tabla por cada columna cuyos valores se desea mantener un historial de cambios. De este modo en la nueva columna se coloca el valor antiguo antes de sobrescribir el valor actual con el nuevo. Este tipo presenta como principal desventaja que solo permite guardar un historial limitado de los datos, dependiendo del número de columnas que se añadan.

**SCD Tipo 4 (tabla de historia separada):** Su función es almacenar en una tabla adicional los detalles de cambios históricos realizados a la tabla de dimensión. La tabla con la información histórica indicará el tipo de operación que se ha realizado, sobre qué campo se realizó el cambio y la fecha del mismo. Esta tabla tiene como objetivo mantener un detalle de los cambios realizados.

**SCD Tipo 6 (híbrido):** Este método combina los tipos anteriores 1, 2 y 3 y se le denomina tipo 6 debido a la suma de los tres tipos que integra ( $1+2+3=6$ ). Esta estrategia utiliza el Tipo 1 (sobrescribir) junto con el Tipo 2 (añadir filas) y el Tipo 3 (añadir columnas), añadiendo además una pareja adicional de columnas para indicar el rango de fechas al cual aplica cada fila en particular.

En la investigación se utilizó las dimensiones de cambio lento de tipo 2 para la dimensión pregunta, pues anualmente las preguntas de la encuesta se modifican, lo que significa que todos los años se va a tener una encuesta nueva. De esta manera existen cambios en la tabla `dim_pregunta` por lo que se agregará tres columnas: fecha inicio, fecha fin y versión.

### 3.2.3 Gestión de los metadatos del proceso de integración

Los metadatos son información descriptiva sobre los datos, es decir que ayudan a identificar y controlar la información. Pueden describir un recurso en particular o un solo elemento. Los metadatos pueden ser agrupados en tres categorías:

- **Metadatos técnicos:** Enfocado a los diseñadores, desarrolladores y administradores durante el desarrollo y mantenimiento. Este es el punto técnico que agrupa las herramientas, aplicaciones y sistemas, para que juntos constituyan la solución.
- **Metadatos del negocio:** Posibilita obtener los datos y la información referente a los aspectos del negocio, como son los datos provenientes de la fuente.
- **Metadatos de proceso:** Es la presentación de las estadísticas sobre los resultados de la ejecución del propio proceso de ETL, incluyendo medidas tales como filas cargado con éxito, las filas rechazadas y la cantidad de tiempo de carga, particularmente es importante en el proceso de limpieza de metadatos. (26)

En la investigación se hizo uso de los metadatos de procesos para obtener la información correspondiente a los procesos de ETL como líneas leídas y el nombre de la transformación y de los metadatos técnicos para la gestión de cambios en los datos fuentes.

### 3.2.4 Gestión de cambios en los datos fuentes

Para definir la estrategia de gestión de los cambios en el Mercado de datos para el área de Tecnologías de la información y las comunicaciones se tuvo en cuenta la periodicidad de la información proveniente de la fuente. Mensualmente y trimestralmente las entidades de todos los municipios del país informan a la ONEI mediante un balance, los resultados obtenidos por los distintos indicadores que maneja el área de las TIC. Teniendo en cuenta la frecuencia de carga de la información, la captura de los cambios en los datos se gestionará mensualmente y trimestralmente a través de las tablas de metadatos md\_cdc. Una vez realizada la carga se actualiza la fecha, nombre de la transformación y el nombre del mercado.

## 3.3 Implementación del subsistema de visualización de datos

Una vez realizada la carga de los datos, se procede a realizar los cubos OLAP y los reportes candidatos, así como la configuración del control de acceso al sistema.

### 3.3.1 Navegación de la capa de visualización

El mapa de navegación es una representación gráfica donde se encuentra organizada toda la información. El Mercado de datos para el área de Tecnologías de la información y las comunicaciones está formado por un Área de Análisis General, un A.A y tres L.T, dentro del cual se encuentran los reportes. En la Figura 17 se detalla la estructura de la capa de visualización que posee el MD.

✓ **Descripción del Área de Análisis General:**

**SIGOB:** Agrupa la información referente a los MD de las distintas áreas de la ONEI, los cuales conforman en su conjunto el AD SIGOB.

✓ **Descripción del Área de Análisis:**

**A.A TIC:** Contiene la información referente a las Tecnologías de la información y las comunicaciones.

✓ **Descripción de los Libros de Trabajo:**

**L.T Encuesta Anual TIC:** Presenta los reportes relacionados con la encuesta que se realiza anualmente en el área de las TIC.

**L.T Indicadores TIC:** Presenta los reportes referentes a los indicadores de las TIC del modelo 0009.

**L.T Ingresos por TIC:** Presenta los reportes referentes a los ingresos de las TIC del modelo 0764.

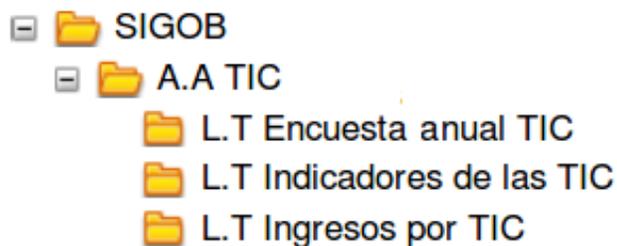


Figura 17. Arquitectura del Área de Análisis.

### 3.3.2 Implementación de los reportes candidatos

Cada L.T contiene un conjunto de reportes, que son realizados mediante consultas MDX. MDX constituye un lenguaje que es utilizado para hacer consultas en los cubos OLAP. Es una sintaxis que sostiene la elaboración y manipulación de objetos multidimensionales y los datos (27). Esto permite la generación de reportes hacia la toma de decisiones basada en el análisis de datos históricos.

Para la elaboración de cada uno de los reportes se especifica una consulta MDX, por lo que se realizaron ocho consultas. En la Figura 18 se muestra la vista de análisis “Presentar información TIC”, que se encuentra en L.T Indicadores de las TIC.

| Indicadores                                     | ● Plan real                       | ● Real del mes | ● Real acumulado | ● Real del año anterior |
|---|-----------------------------------|----------------|------------------|-------------------------|
|   | Potencia radiada de radiodifusión | 16598,6        | 2691,2           | 16457,8                 |
| Potencia radiada de ondas medias                | 11078,5                           | 1805,8         | 11023,5          | 10016,5                 |
| Potencia radiada de frecuencia modulada         | 1534,5                            | 246,5          | 1530,7           | 1508,1                  |
| Potencia radiada de radiodifusión internacional | 3985,6                            | 638,9          | 3903,6           | 4242,6                  |
| Potencia radiada de televisión                  | 2588,6                            | 426,1          | 2604,2           | 2524,0                  |
| Confiabilidad del servicio de ondas medias      | 97,0                              | 98,8           | 98,8             | 98,6                    |

Figura 18. Vista de análisis Presentar información TIC.

### 3.3.3 Seguridad de la aplicación

Con el objetivo de restringir el acceso al sistema y tener el control de su seguridad en el mismo, se definen dos roles y dos usuarios con diferentes permisos.

**Rol administrador:** Tiene el control total sobre la aplicación y es asignado al usuario admin.

**Rol analista:** Posee permisos de ejecución sobre la aplicación, eso permite navegar por los reportes y visualizarlos pero no le permite realizar ningún cambio sobre ellos, ni crear nuevas vistas de análisis. Es asignado al usuario analista.

### Conclusiones del capítulo

En este capítulo se realizó la implementación del Mercado de datos para el área de Tecnologías de la información y las comunicaciones y arrojó las siguientes conclusiones:

- Fueron implementados los tres subsistemas que componen el Mercado de datos para el área de Tecnologías de la información y las comunicaciones, se tuvo como resultado la disponibilidad de la información de manera organizada.
- La estructura física del MD está compuesta por tres esquemas: dimensiones, mart\_tic y metadatos, lo que posibilitó la integración de los datos al mercado.

- Los metadatos de procesos implementados almacenan la información acerca de la ejecución de las transformaciones, lo cual permitió la gestión del sistema y el correcto funcionamiento del mismo.

## Capítulo 4: Prueba del Mercado de datos para el área de Tecnologías de la información y las comunicaciones.

### Introducción

En este capítulo se aplican las pruebas unitarias, de integración, de sistema y de aceptación. Estas pruebas se realizan para garantizar el cumplimiento de las exigencias del cliente y la calidad del producto. Se validan las entradas de los datos verificando que los datos obtenidos se correspondan con los resultados esperados mediante los casos de pruebas. También se realizan las listas de chequeo a los artefactos que se generan en los procesos de ETL.

### 4.1 Calidad de software

Todo proceso de creación de software está sujeto a fallos. La calidad es un aspecto importante a tener en cuenta, debido a que cuando un producto tiene calidad cumple con los requisitos establecidos y el cliente se encuentra satisfecho. (28)

### 4.2 Pruebas de calidad de software

Para determinar la calidad del producto se realizan distintas pruebas de software. Estas son aplicadas desde el inicio del software hasta que la aplicación llega a manos del cliente, cumpliendo con todas las funcionalidades.

#### 4.2.1 Modelo V

Para realizar las pruebas necesarias en el desarrollo de la solución se decidió tomar el Modelo V con el objetivo de definir un software de mayor calidad. El modelo V relaciona las etapas de desarrollo del software y las pruebas.

En la Figura 19 se muestra una representación del ciclo de vida del Modelo V. En la parte izquierda de la figura se detallan las etapas de desarrollo del software y en la parte derecha, las pruebas correspondientes a cada una de las etapas. La codificación forma el vértice de la V, con el análisis y el diseño a la izquierda y las pruebas y el mantenimiento a la derecha. La unión mediante las líneas discontinuas entre las fases y las pruebas indica en qué fase de desarrollo se deben definir las pruebas correspondientes o para saber a qué fase de desarrollo hay que volver si se encuentran fallos en las pruebas efectuadas. (29)

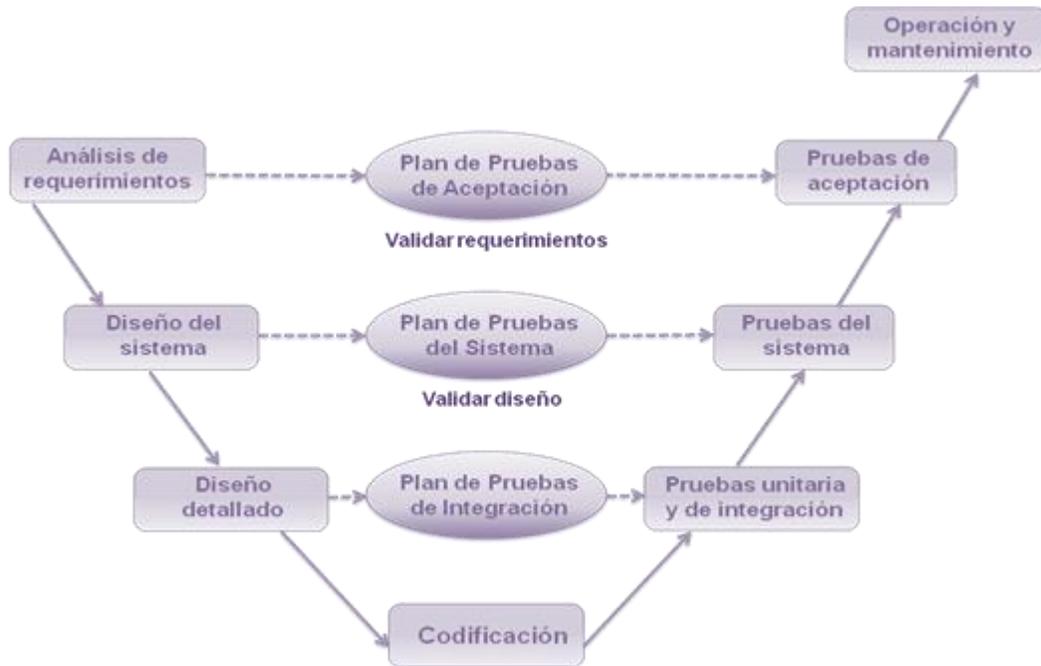


Figura 19. Modelo V.

A continuación se detallan las pruebas a utilizar en el proceso de calidad del software, guiadas a partir del Modelo V:

- Las **pruebas unitarias** se realizan en la fase de diseño, las cuales permiten probar el correcto funcionamiento de un subsistema específico. (12)
- Las **pruebas de integración** son realizadas en la fase de diseño. El objetivo de estas pruebas es verificar el correcto ensamblaje entre los distintos subsistemas una vez que han sido probados unitariamente, con el fin de comprobar que interactúan correctamente a través de sus interfaces. (30)
- Las **pruebas de sistema** son realizadas en la fase de diseño. Permiten validar el cumplimiento de los requisitos de información definidos por el cliente y el software. Son las pruebas más cercanas a la realidad del cliente debido a que los probadores utilizan el sistema de la misma manera que será usado por los clientes. (12)
- Las **pruebas de aceptación** son las pruebas realizadas en la fase de requisitos. Estas pruebas son realizadas por el cliente para verificar que se cumple con los requisitos planteados por el mismo y validar su conformidad con el producto. (12)

### 4.3 Herramientas de pruebas

Para aplicar las pruebas fueron utilizadas las siguientes herramientas:

#### Casos de prueba

Mediante los casos de prueba, el probador podrá determinar si el requisito de una aplicación es parcial o completamente satisfactorio. Para probar la aplicación se utilizaron los casos de pruebas basados en CU. A continuación se mencionan los casos de prueba realizados:

- Caso de prueba Encuesta anual TIC.
- Caso de prueba Ingresos por TIC.
- Caso de prueba Indicadores de las TIC.

En el expediente de proyecto se pueden encontrar los casos de prueba diseñados.

#### Lista de chequeo

La lista de chequeo es un artefacto que contiene una serie de preguntas en forma de cuestionario, la cual verifica el cumplimiento de determinadas reglas establecidas para los procesos de desarrollo del sistema. Las listas de chequeo están estructuradas por los siguientes elementos:

- **Peso:** Define si el indicador a evaluar es crítico o no.
- **Indicadores a evaluar:** Son los indicadores que servirán para evaluar las tres secciones fundamentales que componen las listas de chequeo.
- **Evaluación:** Es el modo de evaluar el indicador, este obtiene evaluación de 1 en caso de que exista alguna dificultad sobre el indicador y de 0 en caso contrario.
- **No procede (N.P):** Especifica que el indicador no presenta evaluación.
- **Cantidad de elementos afectados:** especifica la cantidad de errores que se identificaron en el indicador.
- **Comentario:** Señalamientos o sugerencias que desee incluir la persona encargada de aplicar la lista de chequeo.

Luego de establecida la estructura de la lista de chequeo (consultar expediente de proyecto), es aplicada a los artefactos de los procesos de ETL:

- Lista de chequeo del Mapa Lógico de Datos.

- Lista de chequeo de Registro de Sistema Fuente.
- Lista de chequeo del Perfilado de Datos.
- Lista de chequeo del Diccionario de Datos.

#### 4.4 Resultados de las pruebas

Aplicadas las pruebas descritas anteriormente al Mercado de datos para el área de Tecnologías de la información y las comunicaciones se obtienen los siguientes resultados:

##### Evaluación de las pruebas unitarias

Las pruebas unitarias se aplicaron a cada uno de los subsistemas que componen el MD identificando la siguiente no conformidad (NC):

NC1 Definir que la dimensión DPA presenta dos jerarquías: DPA antigua y DPA actual.

##### Evaluación de las pruebas de integración

Las pruebas de integración se aplicaron a los subsistemas que componen el MD con el fin de comprobar que interactúan correctamente, en las cuales se identificaron las siguientes NC:

- NC1 Agregar los archivos “.properties” para los nombres de los reportes.
- NC2 Crear el A.A TIC.

Posteriormente fueron aplicadas las listas de chequeo a cada uno de los artefactos de los procesos de ETL. La Tabla 7 muestra los principales elementos que componen cada una de estas listas, así como el total de indicadores (TI) por cada una de ellas, la cantidad de indicadores críticos (IC), así no conformidades (NC) detectadas y las no conformidades correspondientes a los indicadores críticos (NC-IC).

| Artefactos           | Estructura | Indicadores | Semántica | TI | IC | NC | NC-IC |
|----------------------|------------|-------------|-----------|----|----|----|-------|
| Diccionario de datos | 9          | 1           | 3         | 13 | 5  | 1  | 0     |
| Perfilado de datos   | 8          | 1           | 3         | 12 | 5  | 0  | 0     |
| Mapa lógico de datos | 5          | 1           | 3         | 9  | 5  | 0  | 0     |

|                            |   |   |   |    |   |   |   |
|----------------------------|---|---|---|----|---|---|---|
| Registro de sistema fuente | 9 | 1 | 3 | 13 | 5 | 1 | 0 |
|----------------------------|---|---|---|----|---|---|---|

Tabla 7. Resultado de la aplicación de las listas de chequeo a los artefactos de los procesos de ETL.

### Evaluación de las pruebas del sistema

Las pruebas del sistema se realizaron por parte de los especialistas de calidad del centro DATEC identificando 4 NC:

- NC1 Errores ortográficos en la fuente de datos.
- NC2 Arreglar el nombre del reporte.
- NC3 Uso excesivo de la mayúscula en el nombre del reporte Encuesta anual TIC del L.T Encuesta anual TIC.
- NC4 Error de interfaz. Los botones *OK* y *Cancel* de la opción Configurar impresión aparecen en inglés, debe ser Aceptar y Cancelar.

Estas NC presentadas durante la realización de las pruebas fueron resultas.

### Evaluación de las pruebas de aceptación

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

## 4.5 Calidad de datos

El concepto calidad de datos se asocia a los sistemas de información, teniendo en cuenta que al procesar grandes volúmenes de datos, pueden encontrarse con información incompleta y valores nulos. Dentro de un MD estos problemas se acentúan cuando se realizan la integración de las distintas fuentes de datos (26). A continuación se describen varios procesos que permiten comprobar la calidad de la información almacenada en el MD.

### 4.5.1 Perfilado de datos

El proceso de perfilado de datos permite obtener información sobre los datos, lo que posibilita corregir los problemas como: valores nulos o valores escritos incorrectamente. Una vez culminado el proceso de integración y la carga de los datos, se realizó el perfilado a los datos que fueron cargados al mercado, con

el objetivo de que posean la calidad requerida. Con el uso de la herramienta Data Cleaner se verificó que en las tablas de hechos y dimensiones no fueran almacenados valores nulos y las medidas fueran valores numéricos. En la Figura 20 se muestra el resultado obtenido del perfilado de datos realizado a la tabla del hecho información tic.

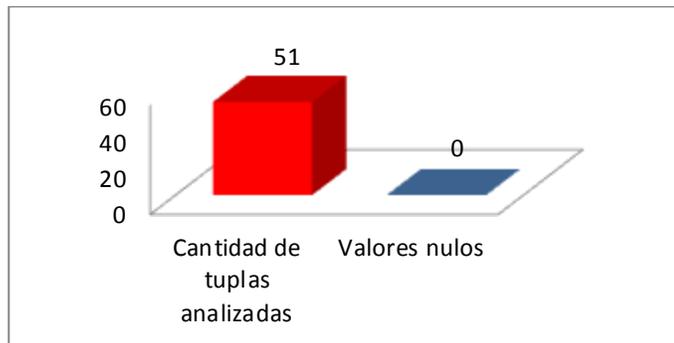


Figura 20. Resultado del perfilado de datos al hech\_informacion\_tic.

#### 4.5.2 Auditoría de datos

La auditoría de datos es el proceso de gestionar cómo los datos se ajustan a los propósitos definidos por la organización. Se establecen políticas para gestionar los criterios de datos para la organización. No es suficiente con actuar sino que se debe vigilar (26).

Aplicando auditoría a los datos se obtiene conocimiento relacionado con la confiabilidad de los mismos. Auditando los datos se puede conocer el nombre de la transformación, la fecha y la hora en que fue ejecutada, el número total de elementos de entrada y salida, así como el número de errores. Para auditar los datos almacenados se basa en el uso del esquema metadatos, el cual contiene la información de los datos cargados hacia el mercado.

#### Conclusiones del capítulo:

En el capítulo se le realizó las pruebas al Mercado de datos para el área de Tecnologías de la información y las comunicaciones, lo que arrojó las siguientes conclusiones:

- Se realizaron los casos de pruebas lo cual validó las entradas de los datos y se verificó que los datos obtenidos se correspondan con los resultados esperados.
- Las listas de chequeo se realizaron a los artefactos que se generan en los procesos de ETL lo que comprobó que la estructura y semántica del documento sea la correcta.

- La aplicación de las pruebas unitarias, las de integración y las de sistema al Mercado de datos para el área de Tecnologías de la información y las comunicaciones responden al cumplimiento de las exigencias del cliente y la calidad del producto.

## Conclusiones

El estudio de los principales conceptos relacionados con el desarrollo de un MD, proporcionó la elaboración de la investigación cuyo resultado fue el Mercado de datos para el área de Tecnologías de la información y las comunicaciones, el cual arrojó los siguientes resultados:

- En el marco teórico de la investigación se seleccionó la Propuesta de metodología para el desarrollo de Almacenes de Datos en DATEC, la cual guió el proceso de desarrollo. El conjunto de herramientas utilizadas en el MD permitió dar cumplimiento al desarrollo de la solución.
- Se realizó el análisis y el diseño del Mercado de datos para el área de Tecnologías de la información y las comunicaciones, identificando 12 RI, 15 RF, 5 RNF y 5 RN, el cual satisfacía las expectativas del cliente y lo requerido en el negocio.
- Se implementaron los subsistemas de integración y visualización lo que permitió obtener como resultado un MD poblado y funcional. Se logró la disponibilidad de la información de manera organizada y facilitó el proceso de toma de decisiones.
- Las pruebas unitarias, de integración, de sistema y de aceptación permitió validar el Mercado de datos para el área de Tecnologías de la información y las comunicaciones lo cual garantizó cumplir con las necesidades del cliente.

## Recomendaciones

Con el propósito de mejorar el trabajo de diploma se sugiere:

- Adicionar nuevas formas de visualización de la información aplicando la técnica de inteligencia de negocio, cuadro de mando integral.

## Referencias bibliográficas

1. **Roberto Hernando Velasco.** Almacenes de datos (Datawarehouse) . [En línea]  
<http://www2.rhernando.net/modules/tutorials/doc/bd/dw.html>.
2. **Inmon, Bill.** *Building the Data Warehouse.*
3. **Kimball, Ralph.** *The Data Warehouse.* New York : s.n., 1998.
4. **Buyto.** [En línea] <http://www.buyto.es/general-business-intelligence/almacenamiento-de-datos-datawarehouse-datamart-en-business-intelligence>.
5. **NEVASOFT.** [En línea]  
[http://www.nebasoft.cl/v2/index.php?option=com\\_content&view=article&id=48:datamart&catid=17:nebasoft&Itemid=63](http://www.nebasoft.cl/v2/index.php?option=com_content&view=article&id=48:datamart&catid=17:nebasoft&Itemid=63).
6. **Ing.Lilliam Vega Torres, Ing.Luis Rojas Díaz.** *La inteligencia de negocio.Su implementación mediante la plataforma Pentaho.* La Habana : s.n., 2008.
7. **Msdn.** [En línea] <http://msdn.microsoft.com/es-es/library/ms244679%28v=vs.80%29.aspx>.
8. **DataPRIX.** [En línea] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse/34-datawarehouse-manager>.
9. **Computer Audio VideoSystem Integrator.** [En línea] <http://www.cavsi.com/preguntasrespuestas/ques-procesamiento-analitico-en-liea-olap/>.
10. **Mastermagazine.** [En línea] <http://www.mastermagazine.info/termino/6841.php#ixzz2LMZB1g7m>.
11. **Universidad de Murcia.** [En línea] <http://www.um.es/docencia/barzana/IAGP/IAGP2-Metodologias-de-desarrollo.html>.
12. **González Hernández, Yanisbel.** *La Metodología para el desarrollo de Almacenes de Datos en DATEC.* noviembre del 2011.
13. **Plusformación.** [En línea] <http://www.plusformacion.com/Recursos/r/Herramientas-CASE-para-proceso-desarrollo-Software>.
14. **PostgreSql Cuba.** [En línea] <http://postgresql.uci.cu/node/63>.
15. **PostgreSQL.** [En línea] <http://archives.postgresql.org/pgsql-es-fomento/2009-07/msg00000.php>.
16. **ArPUG Grupo de Usuarios PostgreSQL Argentina.** [En línea]  
<http://www.arpug.com.ar/trac/wiki/PgAdmin>.

17. **Daniel E. Busso, Matias Torcivia, Valeria Ortiz Quiroz.** Maestría en Ingeniería de Sistemas de Información, Universidad Tecnológica Nacional. [En línea] <http://es.scribd.com/doc/102747055/Calidad-de-Datos>.
18. **Gravitar Información sin Límites.** [En línea] <http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.
19. **Google Books.** [En línea]  
<http://books.google.com.cu/books?id=iU3RAXYQXMkC&pg=PA106&lpg=PA106&dq=pentaho+schema+workbench+%2B+caracteristicas&source=bl&ots=NmccIxbck&sig=bgGz411T6caYAvPYjVIFgp6npMo&hl=es&sa=X&ei=44KdUleQPMWr0AGHwIDYCg&ved=0CCkQ6AEwAg#v=on>.
20. **Summan Manejo Documental e Infraestructura Informática.** [En línea]  
<http://www.summan.com/pentaho/pentaho-bi-platform-server>.
21. **Pressman, Roger s.** *Ingeniería del software: un enfoque práctico*. 2005.
22. **MSDN.** [En línea] <http://msdn.microsoft.com/es-es/library/aa577691%28v=bts.10%29.aspx>.
23. **Kimball, Ralph.** *The Data Warehouse Toolkit*. s.l. : Second Edition. s.l Wiley Publishing.
24. **Douglas Godoy.** sabd15N1. [En línea] <http://es.tldp.org/Postgresql-es/web/navegable/programmer/x6608.html>.
25. **Books Google.** [En línea]  
<http://books.google.com.cu/books?id=iU3RAXYQXMkC&pg=PA90&lpg=PA90&dq=creador+de+cubos+Olap+subsistemas&source=bl&ots=Nmdijxb9at&sig=2HYkS20v5Na2baKHEje3Jys9OhA&hl=es&sa=X&ei=0JiOUbTLOaq40QHglYAY&ved=0CCCoQ6AEwAA#v=onepage&q=creador%20de%20cubos%20Olap%20su>.
26. **Mustelier Medina, Doris.** *Técnicas de Extracción, Transformación y Carga de Datos del Sistema de Información Nacional de Seguridad Ciudadana en la República Bolivariana de Venezuela*. Marzo 2009.
27. **Introducción a MDX.** [En línea] Marzo 21, 2012.  
<http://dragonsql.wordpress.com/2010/04/18/introduccion-a-mdx/>.
28. **ECURED.** [En línea] [http://www.ecured.cu/index.php/Calidad\\_de\\_Software](http://www.ecured.cu/index.php/Calidad_de_Software).
29. **Ciclo de vida- Modelo v.** [En línea] Mayo 2012. <http://spanishpmo.com/index.php/ciclos-de-vida-modelo-en-v>.
30. **Pruebas de integración.** [En línea] Mayo 2012.  
<http://gemini.udistrital.edu.co/comunidad/grupos/arquisoft/fileadmin/Estudiantes/Pruebas/HTML%20-%20Pruebas%20de%20software/node31.html>.

## Bibliografía

1. **Apache** . [En línea] <http://apachefoundation.wikispaces.com/Apache+Tomcat>.
2. **Mustelier Medina, Doris**. *Técnicas de Extracción, Transformación y Carga de Datos del Sistema de Información Nacional de Seguridad Ciudadana en la República Bolivariana de Venezuela*. Marzo 2009.
3. **Pentaho**. [En línea] <http://mondrian.pentaho.com/documentation/olap.php>.
4. **Pruebas unitarias**. [En línea] Mayo 2012. [http://www.dosideas.com/wiki/Prueba\\_Unitaria](http://www.dosideas.com/wiki/Prueba_Unitaria).
5. **Pruebas de sistema**. [En línea] Mayo 2012. <http://gemini.udistrital.edu.co/comunidad/grupos/arquisoft/fileadmin/Estudiantes/Pruebas/HTML%20-%20Pruebas%20de%20software/node37.html>.
6. **Portal ONEI**. [En línea] <http://www.onei.cu>.
7. **UCI. Portal UCI**. [En línea] <http://www.uci.cu>.
8. **34 Subsistemas ETL**. [En línea] <http://bi.social.uoc.edu/smc/blog/34-subsistemas-etl-de-kimball>.

Anexos

|   |  |  |  |                           |   |
|---|--|--|--|---------------------------|---|
| <br>OFICINA NACIONAL DE ESTADÍSTICA E INFORMACIÓN<br>República de Cuba | Sistema de Información Estadístico Nacional (SIEN) | INGRESOS POR TECNOLOGÍAS DE LA INFORMÁTICA | Informe acumulado hasta:<br>___Marzo 31 ___ Septiembre 30<br>___Junio 30 ___ Diciembre 31<br>Año: [ ][ ] |                           | Modelo No. 0764-05<br>Página 1 de 1<br>TRIMESTRAL |
|   |  |  | Centro informante:   | Código centro informante: |   |

| INDICADOR                  | FILA | U.M.: Uno Cantidad | U.M.: MP            |            | U.M.: MCUC          |            |             |            |
|----------------------------|------|--------------------|---------------------|------------|---------------------|------------|-------------|------------|
|                            |      |                    | Producción Nacional | Importados | Producción Nacional | Importados | Exportación | Donaciones |
| A                          | B    | C                  | 1                   | 2          | 3                   | 4          | 5           | 6          |
| Total de ingresos          | 01   |                    |                     |            |                     |            |             |            |
| Equipos de computación     | 02   |                    |                     |            |                     |            |             |            |
| Computadoras               | 03   |                    |                     |            |                     |            |             |            |
| Otros equipos y accesorios | 04   |                    |                     |            |                     |            |             |            |
| Insumos informáticos       | 05   |                    |                     |            |                     |            |             |            |
| Servicios técnicos         | 06   |                    |                     |            |                     |            |             |            |
| Software                   | 07   |                    |                     |            |                     |            |             |            |
| Paquetes y aplicaciones    | 08   |                    |                     |            |                     |            |             |            |
| Servicios informáticos     | 09   |                    |                     |            |                     |            |             |            |

|  |   |  |   |  |  |  |     |     |     |
|--|---|--|---|--|--|--|-----|-----|-----|
| Certificamos que los datos contenidos en este modelo corresponden a los anotados en nuestros registros primarios y de acuerdo a las instrucciones vigentes para la elaboración del mismo | DIRECTOR CONTABLE-FINANCIERO<br>Nombre y apellidos: _____<br>Firma: _____ | DIRECTOR GENERAL<br>Nombre y apellidos: _____<br>Nombre y apellidos: _____ | FECHA<br><table border="1"> <tr> <td>  </td> <td>  </td> <td>  </td> </tr> <tr> <td>Día</td> <td>Mes</td> <td>Año</td> </tr> </table> |  |  |  | Día | Mes | Año |
|  |   |  |   |  |  |  |     |     |     |
| Día  | Mes   | Año  |   |  |  |  |     |     |     |

Anexo 1. Modelo 0764 Ingresos por tecnologías de la informática.

|   |  |   |   |   |
|---|--|---|---|---|
| <br>OFICINA NACIONAL DE ESTADÍSTICA E INFORMACIÓN | Sistema de Información Estadístico Nacional (SIEN) | INDICADORES ESPECÍFICOS DE LOS SERVICIOS<br>1. Empresas Transportistas<br>2. Empresas no Transportistas<br>3. Empresas TICs | Modelo No. 0009-02<br>Páginas 1 de 2<br>MENSUAL/SEMESTRAL/ANUAL |   |
|   |  |   | Informe acumulado hasta: Mes: _____ Año: _____                  | Unidad de medida: Entero con un decimal |
| Centro Informante:  |  | Código Centro Informante:   |   |   |

| INDICADOR                       | U.M | Código | Plan | Real |           | Real Año Anterior |
|---------------------------------|-----|--------|------|------|-----------|-------------------|
|                                 |     |        |      | Mes  | Acumulado | Acumulado         |
| A                               | B   | C      | 1    | 2    | 3         | 4                 |
|                                 |     |        |      |      |           |                   |
|                                 |     |        |      |      |           |                   |
|                                 |     |        |      |      |           |                   |
|                                 |     |        |      |      |           |                   |
|                                 |     |        |      |      |           |                   |
|                                 |     |        |      |      |           |                   |
|                                 |     |        |      |      |           |                   |
|                                 |     |        |      |      |           |                   |
|                                 |     |        |      |      |           |                   |
| SUMA DE CONTROL (Página 2 de 2) |     | 999999 |      |      |           |                   |

|   |   |   |   |  |  |  |  |  |  |     |     |     |  |  |  |
|---|---|---|---|--|--|--|--|--|--|-----|-----|-----|--|--|--|
| Certificamos que los datos contenidos en este modelo corresponden a los anotados en nuestros registros primarios y de acuerdo a las instrucciones vigentes para la elaboración del mismo. | VICEDIRECTOR ECONÓMICO<br>Nombre y apellidos: _____<br>Firma: _____ | DIRECTOR<br>Nombre y apellidos: _____<br>Firma: _____ | FECHA<br><table border="1"> <tr> <td>  </td> <td>  </td> <td>  </td> <td>  </td> <td>  </td> <td>  </td> </tr> <tr> <td>Día</td> <td>Mes</td> <td>Año</td> <td>  </td> <td>  </td> <td>  </td> </tr> </table> |  |  |  |  |  |  | Día | Mes | Año |  |  |  |
|   |   |   |   |  |  |  |  |  |  |     |     |     |  |  |  |
| Día   | Mes   | Año   |   |  |  |  |  |  |  |     |     |     |  |  |  |

Anexo 2. Modelo 0009 Indicadores de las Tecnologías de la información y las comunicaciones.

**Glosario de términos**

**AD:** Almacén de Datos.

**CUF:** Caso de uso funcional.

**CUI:** Caso de uso de información.

**CUS:** Casos de uso del sistema.

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

**ETL:** Extracción, Transformación y Carga.

**MD:** Mercado de Datos.

**OLAP:** Procesamiento analítico en línea.

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

**RF:** Requisito funcional.

**RN:** Reglas del negocio.

**RNF:** Requisito no funcional.

**RI:** Requisito de información.

**SGBD:** Sistema Gestor de Base de Datos.

**UCI:** Universidad de las Ciencias Informáticas.