



**FACULTAD DE CIENCIAS Y TECNOLOGÍAS**  
**COMPUTACIONALES**

Título: Actualización del Mercado de datos de la Fiscalía  
General de la República

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

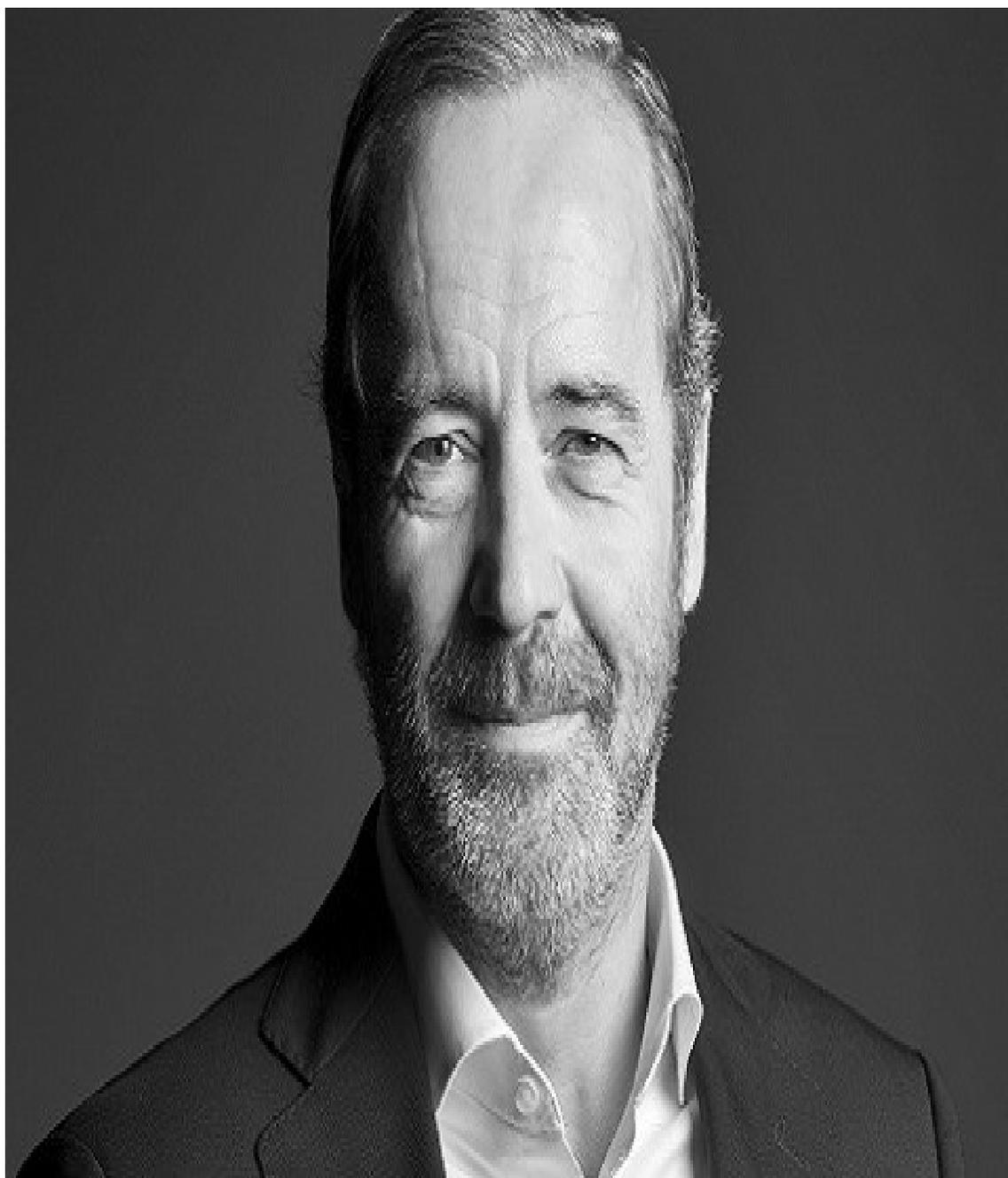
**Autor:** Elizabeth Cordero Díaz

**Tutor:** Ing. Liuba María Infante Infante

**Co-tutor:** Ing. Amarelys Alvarez Pérez

La Habana, junio de 2022

Año 63 de la Revolución



“LOS DATOS SON EL ACTIVO ESTRATÉGICO DE LAS COMPAÑÍAS POR EXCELENCIA”

Christian Gardine

**DECLARACION DE AUTORIA**

El autor del trabajo de diploma con título **“Actualización del Mercado de datos de la FGR”**, concede a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la investigación, con carácter exclusivo. De forma similar se declara como único autor de su contenido. Para que así conste firma la presente a los días 29 del mes noviembre del año 2022.

**Elizabeth Cordero Díaz**

---

Firma del Autor

Ing. Liuba Infante Infante

---

Firma del Tutor

Ing. Amarelys Alvarez Pérez

---

Firma del Co-Tutor

## **DATOS DE CONTACTO**

### **Información de Contacto del Tutor**

Nombre y Apellidos: Instructor, Ing. Liuba Infante Infante, SCC

Títulos Académicos: Ingeniería en Ciencias Informáticas

Lugar de Trabajo: Universidad de Las Ciencias Informáticas

Experiencia Profesional: 10

Líneas de Trabajo y/o Investigación: Almacenes de Datos

Correo Electrónico: lminfante@uci.cu

### **Información de contacto del Co-tutor**

Nombre y Apellidos: Amarelys Alvarez Pérez

Títulos Académicos: Ingeniería en Ciencias Informáticas

Lugar de Trabajo: Universidad de Las Ciencias Informáticas

Responsabilidades Laborales Asumidas: Especialista A

Experiencia Profesional: 10

Líneas de Trabajo y/o Investigación: Almacenes de Datos

Correo Electrónico: aaperez@uci.cu

## **AGRADECIMIENTOS**

Agradezco a mi familia por ser mis pilares a seguir y ayudarme incondicionalmente a cumplir mis sueños.

Agradezco a mis amigos que siempre estuvieron a mi lado aún más cuando todo parecía que no iba bien.

Agradezco inmensamente a mis tutoras Amarelys Alvarez Pérez y Liuba María Infante Infante por ayudarme, aconsejarme y apoyarme durante todo el proceso de desarrollo del presente proyecto.

### **DEDICATORIA**

Dedico este trabajo de diploma a mi familia que fue mi impulso y mi motor ante todos mis años de estudio y sacrificio, por no dejarme sola y siempre estar ahí cuando más los necesitaba.

## RESUMEN

La Fiscalía General de la República (FGR) es el órgano del Estado que tiene como misión fundamental ejercer el control de la investigación penal y el ejercicio de la acción penal pública en representación del Estado, así como velar por el estricto cumplimiento de la Constitución y las leyes. Actualmente la Fiscalía al tomar decisiones y llevar a cabo sus procesos fiscales necesita realizar análisis a datos tanto históricos como actuales para una correcta toma de decisiones además de utilizar potentes herramientas que permitan el análisis de datos o información más compleja con las que deba interactuar teniendo como intención que la información que guarda y que muestra sea confiable, se encuentre disponible en todo momento que se necesite y que no presente errores que afecten su integridad. A partir de estas necesidades de la fiscalía se determinó como objetivo Actualizar el Mercado de datos (MD) de la FGR. Para guiar el desarrollo de la solución se utilizó la “Metodología para el desarrollo de proyectos de Almacenes de Datos”, haciendo uso del *Visual Paradigm*, *PostgreSQL* y varias herramientas de la *Suite de Pentaho*. Después del análisis, diseño e implementación se obtuvo como solución un MD y la capa de visualización, donde se muestran los reportes correspondientes a cada indicador del proceso de Mercado de datos para la FGR al cual para verificar y valorar el funcionamiento del producto se realizaron pruebas unitarias, de integración, de sistema y aceptación.

Palabras clave:

Toma de decisiones, Mercado de Datos.

## **ABSTRACT**

*The Office of the Attorney General of the Republic (FGR) is the State body whose fundamental mission is to exercise control over criminal investigations and the exercise of public criminal action on behalf of the State, as well as to ensure strict compliance with the Constitution and laws. Currently, the Prosecutor's Office, when making decisions and carrying out its fiscal processes, needs to analyze both historical and current data for correct decision-making, in addition to using powerful tools that allow the analysis of data or more complex information with which it must interact, having as The intention is that the information it stores and displays is reliable, available at all times it is needed and that it does not present errors that affect its integrity. Based on these needs of the prosecutor's office, the objective was to update the FGR data market (DM). To guide the development of the solution, the "Methodology for the development of Data Warehouse projects" was used, making use of Visual Paradigm, PostgreSQL and several tools from the Pentaho Suite. After the analysis, design and implementation, a DM and the visualization layer were obtained as a solution, where the reports corresponding to each indicator of the Data Market process for the FGR are shown, to which tests were carried out to verify and assess the operation of the product, unitary, integration, system and acceptance.*

*Keywords:*

*Decision making, Data Market.*

## ÍNDICE

INTRODUCCIÓN.....	1
CAPÍTULO I: FUNDAMENTACIÓN TEÓRICA DE LOS ALMACENES DE DATOS.....	6
1. Almacenes de Datos.....	6
I.1 Definiciones asociadas a los Almacenes de Datos.....	6
I.2 Características de un almacén de datos.....	7
I.3 Arquitectura de un almacén de datos.....	7
I.4 Componentes que forman parte de la Arquitectura de un Almacén de datos.....	7
1.5 Ventajas e inconvenientes.....	8
I.6 Mercado De Datos.....	9
I.6.1 Definición.....	9
I.6.2 Importancia del mercado de datos.....	9
I.7 Modelos o Esquemas de los Almacenes de Datos.....	10
I.8.1 Tipos de Esquemas:.....	10

I.8 Modo de almacenamiento OLAP.....	12
I.8.1 Funcionamiento de los sistemas OLAP.....	12
I.8.2 Tipos de sistemas OLAP.....	13
I.9 Metodologías para el desarrollo de almacenes de datos.....	13
I.10 Metodología a utilizar en la solución.....	16
I.10.1. Objetivos de la Metodología.....	16
I.10.2 Fases del ciclo de vida de la metodología.....	16
I.11 Tendencias.....	18
I.11.1 Tendencias actuales de los almacenes de datos en el mundo.....	18
I.11.2 Tendencias de Almacenes de Datos en Cuba.....	19
I.12 Tecnologías para el desarrollo del Mercado de datos de la FGR.....	20
I.13 Herramienta de modelado de datos.....	20
I.14 Sistema gestor de base de datos.....	21
I.14.1 Administrador de bases de datos.....	21
I.15 Herramienta para el perfilado de datos.....	22
I.16 Herramienta para la extracción, transformación y carga de los datos.....	22
I.17 Herramientas de inteligencia de negocio.....	22
I.18 Conclusiones del capítulo.....	23
<b>CAPÍTULO II: ANÁLISIS Y DISEÑO DEL MERCATO DE DATOS DE LA FGR.....</b>	<b>24</b>
II.1 Estudio preliminar del negocio. Caracterización de los Procesos Fiscales.....	24
II.2 Requisitos de información.....	25
II.2.1 Requisitos Funcionales.....	25
II.2.2 Requisitos No Funcionales.....	26
II.2.3 Reglas del negocio.....	27
II.3 Modelo de casos de uso del sistema.....	28
II.3.1 Actores de sistema.....	28

II.3.2 Diagrama de CUS.....	29
II.3.3 Descripción de casos de uso del sistema.....	29
II.3.4 Especificación de casos de uso.....	30
II.4 Arquitectura del Mercado de datos de la FGR.....	33
II.5 Diseño de los subsistemas que componen el Mercado de datos de la FGR.....	34
II.5.1 Diseño del subsistema de almacenamiento.....	34
II.5.1.1 Hechos y medidas.....	34
II.5.1.2 Dimensiones.....	35
II.6 Matriz bus.....	35
II.6.1 Modelo de datos.....	36
II.6.2 Diseño del subsistema de integración.....	36
II.6.3 Perfilado de los datos.....	37
II.6.4 Diseño de transformaciones.....	38
II.6.4.1 Diseño para la carga de dimensiones.....	38
II.6.4.2 Diseño para la carga de hechos.....	39
II.7 Diseño del subsistema de visualización.....	39
II.7.1 Arquitectura de la Información.....	39
II.7.2 Diseño de Reportes.....	40
II.8 Esquema de Seguridad.....	41
II.8.1 Políticas de respaldo y recuperación.....	41
II.8.2 Roles y permisos.....	41
II.9 Implementación del subsistema de almacenamiento.....	41
II.9.1 Modelo de Datos Físico.....	42
II.9.2 Estándares de codificación.....	42
II.10 Implementación del subsistema de integración de datos.....	43
II.10.1 Transformaciones y trabajos.....	43

II.11 Implementación del subsistema de visualización.....	45
II.11.1 Implementación de los cubos OLAP.....	46
II.11.2 Implementación de la capa de visualización.....	46
II.11.2.1 Implementación de los reportes operacionales.....	47
II.11.2.2 Implementación de las vistas OLAP.....	48
II.11.2.3 Implementación de los cuadros de mando.....	49
II.12 Conclusiones del capítulo.....	50
CAPÍTULO III: PRUEBAS DE VALIDACIÓN DE LA SOLUCIÓN PROPUESTA.....	52
III.1 Pruebas al MD de la FGR.....	52
III.1.1 Pruebas aplicadas a la documentación del proyecto.....	52
III.1.2 Pruebas aplicadas al software.....	53
III.1.2.1 Pruebas unitarias.....	53
III.1.2.2 Pruebas de integración.....	55
III.1.2.3 Prueba de sistema.....	56
III.1.2.3 Pruebas de aceptación.....	57
III.2 Conclusiones del capítulo.....	58
CONCLUSIONES.....	59
RECOMENDACIONES.....	60
REFERENCIAS BIBLIOGRÁFICAS.....	61
Anexos.....	65

## ÍNDICE DE FIGURAS

FIG 1- ARQUITECTURA GENERAL DE UN AD TOMADA DE LA METODOLOGÍA HEFESTO.	7
FIG 2- ESQUEMA ESTRELLA [ELABORACIÓN PROPIA]	11
FIG 3- ESQUEMA EN COPO DE NIEVE [ELABORACIÓN PROPIA]	11
FIG 4- ESQUEMA EN CONSTELACIÓN O GALAXIA [ELABORACIÓN PROPIA]	12
FIG 5- CICLO DE VIDA KIMBALL TOMADO DE METODOLOGÍAS PARA DESARROLLAR ALMACÉN DE DATOS.	14
FIG 6- FASES DE LA METODOLOGÍA RAPID WAREHOUSING METHODOLOGY TOMADO DE METODOLOGÍAS PARA DESARROLLAR ALMACÉN DE DATOS	15
FIG 7- CICLO DE VIDA DE LA METODOLOGÍA DE DESARROLLO DE PROYECTOS DE ALMACENES DE DATOS.	17
FIG 8- DIAGRAMA DE CASO DE USO DEL SISTEMA.	29
FIG 9 - ARQUITECTURA DEL MD DE LA FGR.	33
FIG 10- MODELO DE DATOS.	36
FIG 11- PERFILADO DE DATOS.	38

FIG 12- DISEÑO PARA LA CARGA DE DIMENSIONES.	38
FIG 13- DISEÑO PARA LA CARGA DEL HECHO.	39
FIG 14- MAPA DE NAVEGACIÓN.	40
FIG 15- ESTRUCTURA FÍSICA DEL MERCADO DE DATOS.	42
FIG 16- PROCESO DE INTEGRACIÓN DE DATOS PARA LA DIMENSIÓN DIM_CENTRO_INFORMANTE.	44
FIG 17- PROCESO DE INTEGRACIÓN DE DATOS PARA EL HECHO PROCESO_FISCAL.	45
FIG 18 TRABAJO_GENERAL.	45
FIG 19- IMPLEMENTACIÓN DEL CUBO OLAP DEL MD	46
FIG 20- ARQUITECTURA DE VISUALIZACIÓN DE LA INFORMACIÓN.	46
FIG 21- REPORTE OPERACIONAL DEL A.A PROCESOS PENALES.	48
FIG 22-VISTA OLAP REFERENTE AL A.A PROCESOS PENALES EN LA PROVINCIA DE PINAR DEL RÍO.	49
FIG 23- CUADRO DE MANDO DE LA DIRECCIÓN DE PROCESOS PENALES.	50
FIG 24- RESULTADO DE LA APLICACIÓN DE LAS LISTAS DE CHEQUEO	53
FIG 25- RESULTADO DE LA APLICACIÓN DE LAS PRUEBAS UNITARIAS.	54
FIG 26- RESULTADO DE LA APLICACIÓN DE LAS PRUEBAS DE INTEGRACIÓN.	55
FIG 27- RESULTADO DE LA APLICACIÓN DE LAS PRUEBAS DEL SISTEMA.	57
FIG 28- RESULTADO DE LA APLICACIÓN DE LA PRUEBA DE ACEPTACIÓN.	58
FIG 29- RESULTADO DE LAS PRUEBAS.	58

## ÍNDICE DE TABLAS

TABLA 1- DESCRIPCIÓN DE LOS ACTORES DEL SISTEMA.....	28
TABLA 2- DESCRIPCIÓN DE CASOS DE USOS DEL SISTEMA.....	30
TABLA 3- MOSTRAR INFORMACIÓN SOBRE LOS PROCESOS FISCALES DE LA DIRECCIÓN DE PROCESOS PENALES.....	33
TABLA 4- DESCRIPCIÓN DE LA TABLA HECHOS.....	34
TABLA 5- DESCRIPCIÓN DE LAS TABLAS DIMENSIONES.....	35
TABLA 6- MATRIZ BUS.....	36
TABLA 7- REPORTE “CANTIDAD DE PROCESOS FISCALES” DEL AA PROCESOS FISCALES.....	40
TABLA 8- ROLES Y PERMISOS.....	41
TABLA 9- ESTÁNDARES DE CONFIGURACIÓN PARA EL MD.....	43
TABLA 10- ESCENARIO 3 DEL CASO DE USO “MOSTRAR INFORMACIÓN SOBRE LOS PROCESOS FISCALES DEL A.A PROCESOS PENALES” .....	56



## INTRODUCCIÓN

El mundo gira cada vez más rápido. La era digital y el uso de nuevas tecnologías nos han llevado a vivir en una época en donde todo va a gran velocidad. Lo mismo pasa con las nuevas tecnologías en investigación de mercados y el uso de un software para la recopilación de datos. Estas tecnologías están hondamente arraigadas en la vida social y forman parte de la manera en que las organizaciones conducen sus negocios, lo cual trae como resultado que se encuentren evolucionando constantemente, reflejándose así en el aumento del volumen de los datos que se generan como resultado de sus procesos operativos.

Es muy común, que los datos que se generan de estas operaciones se almacenen y administren a través de sistemas transaccionales en Bases de Datos (BD) relacionales. La idea central que se persigue en las organizaciones es que dejen de ser solo simples datos para convertirse en información que enriquezca la toma de decisiones. Precisamente, la inteligencia de negocios (BI), entrará a jugar un rol protagónico debido a que proporcionará las herramientas necesarias para analizar la información contenida en esos sistemas transaccionales. Para desarrollar la BI en las empresas u organizaciones, es necesario gestionar datos guardados en distintos formatos, fuentes, para luego depurarlos e integrarlos, además de almacenarlos en un solo destino o BD que permita su posterior análisis y exploración; por lo que es imperativo y de vital importancia contar con una herramienta que satisfaga todas las necesidades.

En los últimos años, Cuba se ha aproximado cada vez más a esta apertura tecnológica, desplegando varios proyectos y procesos de informatización de la sociedad. Como parte de estos procesos, se han establecido entidades encargadas de la producción de soluciones informáticas. Tal es el caso de la Universidad de las Ciencias Informáticas (UCI), la que posee entre sus misiones además de formar profesionales altamente calificados en la rama de la Informática, producir aplicaciones y servicios informáticos a partir del vínculo estudio, trabajo como modelo de formación, investigación, producción, sirviendo así de soporte a la industria cubana de la Informática [CITATION UCI22 \l 2058 ].

La universidad cuenta con diversos centros de desarrollo enfocados a soluciones informáticas entre los que está presente el Centro de Representación y Análisis de Datos (CREAD). Dicho centro está encargado de informatizar los procesos de la Fiscalía General de la República (FGR) que es el órgano encargado de vigilar por el estricto cumplimiento de la constitución, las disposiciones legales y las leyes.

Entre las Áreas de Análisis de la FGR se encuentran: Dirección de Procesos Penales (DPP), Dirección de Verificación Fiscal (DVF), Dirección de Atención a los Ciudadanos (DAC), Dirección a la Protección a la Familia y Asuntos Jurisdiccionales (DPFAJ), Dirección de Control de la Legalidad en Establecimientos Penitenciarios (CLEP), Dirección de Cuadros (DC), Dirección de Comunes (DCO) y la Dirección de Comunicaciones (DIC). En la actualidad, la FGR capta los datos en cada una de sus direcciones a partir del Sistema de Gestión Fiscal (SIGEF), no siendo así en años anteriores, por tal motivo el Sistema de Análisis para los Partes Mensuales Informativos (SAPMI) desplegado en la FGR desde 2014, contempla información histórica de los procesos que allí se manejan, no presentando información actualizada de los últimos años, pudiendo constituir estos datos juntos a los históricos una base de conocimiento para el apoyo de la toma de decisiones.

El alcance que presenta SAPMI no incluye la visualización gráfica de los indicadores más importantes en el proceso fiscal, sin embargo, este tipo de representación es una potente herramienta para el análisis fácil de información compleja para los fiscales. Los especialistas de la FGR tienen necesidades de información que no se contemplan en el Almacén de Datos (AD) siendo necesario incluir nuevas aristas de análisis como Centros Informantes y los modelos de los que se visualiza la información significativa necesaria para que estos sirvan de apoyo a la hora de tomar una decisión que sea factible en cada proceso que se maneje.

A partir de la problemática anteriormente planteada se define como **Problema Científico**:

¿Cómo garantizar la integridad, confiabilidad y disponibilidad de los datos que se generan en la FGR?

De acuerdo al problema planteado, se define como **objeto de estudio** los almacenes de datos, enmarcado en el **campo de acción** Mercado de datos de la FGR.

Se define como **Objetivo general**: Desarrollar un Mercado de Datos (MD) que garantice la integridad, confiabilidad y disponibilidad de los datos que se generan en la FGR

Se define como **Objetivos específicos**:

1. Elaborar los fundamentos teóricos de la metodología, herramientas y tecnologías de almacenamiento a utilizar en el desarrollo del MD de la FGR.
2. Realizar el análisis y diseño del MD de la FGR.
3. Realizar la implementación y prueba del MD de la FGR.

A partir de los objetivos específicos planteados anteriormente se define como **tareas de investigación**:

1. Estudio de elementos teóricos relacionados con AD para profundizar el nivel de comprensión y conocimiento sobre los mismos.
2. Caracterización de la metodología, herramientas y tecnologías de almacenamiento a utilizar, en el desarrollo del MD para profundizar en el nivel de comprensión y dominio de las mismas.
3. Levantamiento de los requisitos del sistema para identificar las necesidades del cliente.
4. Descripción de los casos de uso para determinar cada una de las funcionalidades del MD.
5. Definición de la arquitectura para identificar los subsistemas fundamentales que componen la solución.
6. Definición de los hechos, las medidas y las dimensiones para identificar los elementos que forman parte del modelo lógico del MD.
7. Realización del modelo lógico para determinar los elementos que conforman el modelo físico
8. Realización del diseño del subsistema de almacenamiento para determinar las estructuras dimensionales del MD (hechos y dimensiones).
9. Realización del diseño del subsistema de integración para diseñar los procesos de carga de las dimensiones y hechos.
10. Realización del diseño del subsistema de visualización para realizar el diseño de los cubos OLAP.
11. Implementación del subsistema de almacenamiento para la construcción física del MD de la FGR.
12. Implementación del subsistema de integración para poblar el MD con los datos proveniente de los sistemas fuentes.
13. Implementación del subsistema de visualización para mostrar la información mediante reportes contenidos dentro de los libros de trabajo según el área de análisis identificada.

**Los Métodos Teóricos de la investigación [CITATION Lóp21 \l 21514 ]:**

- ❖ **Analítico-sintético:** Posibilita el análisis y procesamiento de la información recopilada durante la investigación sobre los AD y MD.
- ❖ **Histórico-Lógico:** Permite el estudio de los antecedentes y tendencias actuales de los AD, metodologías, tecnologías y herramientas de desarrollo.
- ❖ **Modelación:** Facilita realizar la modelación del ciclo de vida del Mercado de Datos para las distintas áreas de análisis de la FGR y la concepción de los diagramas establecidos por la metodología para el desarrollo del Almacén de datos.

**Los Métodos Empíricos de la investigación [CITATION Lóp21 \l 2058 ]:**

- ❖ **Observación:** se empleó este método para analizar la problemática de la investigación existente.
- ❖ **Entrevista:** este método permitió la recogida de información durante el proceso de ingeniería de requisitos.

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

El documento está estructurado de la siguiente manera: introducción, tres capítulos, conclusiones, recomendaciones, referencias bibliográficas y anexos.

### **Capítulo 1: Fundamentación teórica sobre el desarrollo de los AD, MD.**

En este capítulo se realiza una breve explicación de las metodologías, técnicas y herramientas que se tuvieron en cuenta para dar solución al problema; así como los principales conceptos y características de las mismas.

### **Capítulo 2: Diseño e implementación de los mercados de datos.**

En este capítulo se realiza un análisis del negocio, con el propósito de comprender el proceso de análisis del Mercado de Datos para la FGR. Se definirán los requisitos de información y las reglas de negocio según las necesidades del cliente identificadas, los requisitos funcionales y no funcionales para luego diseñar el diagrama de caso de uso del sistema. Se identificarán los hechos, medidas y dimensiones, conformando el modelo de datos de la solución. Además, se definirá la arquitectura a utilizar para el desarrollo del AD, definiéndose el diseño correspondiente a cada subsistema identificado, además de establecerse las políticas de respaldo y recuperación en caso de ocurrencia de fallos. Se diseñarán e implementarán reportes, vistas de análisis y cuadros de mando como parte para la representación gráfica de los indicadores más importantes.

### **Capítulo 3: Pruebas de validación.**

En el capítulo se realizan un conjunto de pruebas unitarias, de integración, de sistema y de aceptación con el fin de garantizar la calidad de la solución propuesta.

## CAPÍTULO I: FUNDAMENTACIÓN TEÓRICA DE LOS ALMACENES DE DATOS

En este capítulo se abordará acerca de los conceptos fundamentales que conforman los AD y MD, así como sus ventajas, desventajas entre otros. Se realiza la selección de la metodología para el desarrollo del mismo, además de cuales herramientas serán utilizadas para cumplir con el buen funcionamiento y cumplimiento de los objetivos trazados.

### 1. Almacenes de Datos

Un AD es un sistema de almacenamiento digital que conecta y armoniza grandes cantidades de datos de muchas fuentes diferentes y guardan los datos actuales e históricos en un solo lugar y actúan como la única fuente de verdad para una organización. Su propósito es alimentar la BI, informes y analíticas, y dar soporte a los requisitos regulatorios de manera que las empresas puedan convertir sus datos en información estratégica y tomar decisiones inteligentes basadas en datos [CITATION Chr18 \l 2058 ].

#### 1.1 Definiciones asociadas a los Almacenes de Datos

Existen definiciones, dadas por especialistas como:

1. **Bill Inmon** fue uno de los primeros autores en escribir sobre el tema el cual lo define como "...una colección de datos orientados a temas, integrados, no-volátiles y variables en el tiempo, organizados para soportar necesidades empresariales..." [CITATION Imo02 \l 21514 ].

**Inmon** defiende una metodología descendente (*top-down*) a la hora de diseñarlos, ya que de esta forma se considerarán mejor todos los datos corporativos. En esta metodología los *Data Marts* se crearán después de haber terminado el sistema completo de la organización [CITATION Imo02 \l 21514 ].

2. **Ralph Kimball** es otro conocido autor en el tema el cual los define como: "Un AD extrae, limpia, conforma y entrega una fuente de datos dimensional para la consulta y el análisis".

También fue Kimball quien determinó que estos no eran más que: "la unión de todos los *Data Marts* de una entidad". Defiende por tanto una metodología ascendente (*bottom-up*) a la hora de diseñar un AD [CITATION Kim04 \l 21514 ].

3. **Una definición más amplia:**

Un AD, es un repositorio central de datos almacenados a partir de fuentes internas y externas de una empresa u organización, con el fin de ofrecer una vista unificada de los datos a los analistas y usuarios comerciales para mejorar la BI, además también se utiliza con fines analíticos y de informes comerciales, lo que ayuda a mantener registros pasados y analizar datos para optimizar las operaciones comerciales.

## I.2 Características de un almacén de datos

Un almacén de datos se caracteriza por ser [CITATION Ram11 \l 2058 ]:

- ❖ **Orientado a temas:** los datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.
- ❖ **Variante en el tiempo:** los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar se reflejen en esas variaciones.
- ❖ **No volátil:** una vez almacenado un dato, este se convierte en información de sólo lectura, y se mantiene para futuras consultas, es decir, la información no se modifica ni se elimina.
- ❖ **Integrado:** contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.

## I.3 Arquitectura de un almacén de datos.

La arquitectura del AD describe todo el flujo de datos, desde que son extraídos de los sistemas fuentes, hasta su preparación, está compuesta por varios elementos que interactúan entre sí y que cumplen con una función específica dentro del sistema, como se muestra en la Figura 1 [CITATION Cas \l 21514 ]:

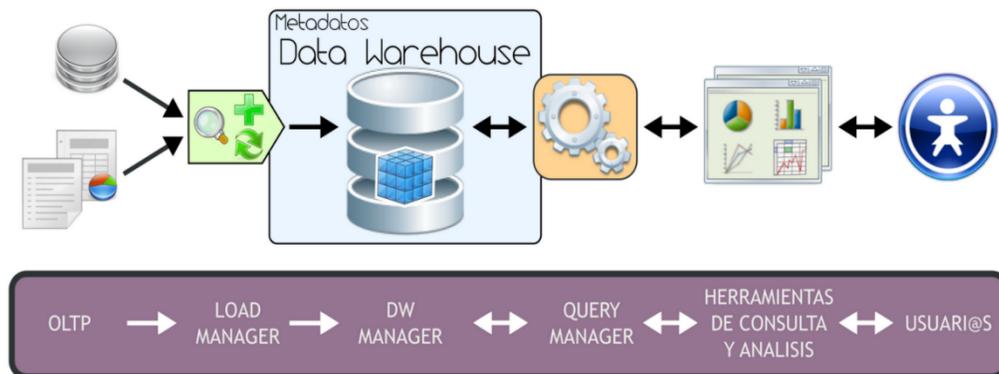


Fig 1- Arquitectura general de un AD tomada de la metodología HEFESTO.

## I.4 Componentes que forman parte de la Arquitectura de un Almacén de datos.

Los componentes según Bernabéu son [CITATION Ber17 \l 21514 ]:

- ❖ **OLTP (On Line Transaction Processing):** representa toda aquella información transaccional que genera la organización diariamente.
- ❖ **Gestor de Carga (Load Manager):** lleva a cabo el proceso ETL, es decir, se encarga de extraer datos de las diversas fuentes que se requieran, y transformarlos para resolver posibles

problemas de inconsistencias entre los mismos y finalmente, después de haberlos depurado se procede a su carga.

- ❖ **Gestor del AD (*DW Manager*):** la información del AD se estructura en cubos multidimensionales, los cuales preparan esta información para responder a consultas dinámicas.
- ❖ **Gestor de consultas (*Query Manager*):** realiza las operaciones necesarias para soportar los procesos de gestión y ejecución de consultas relacionales, tales como unión (*join*) y agregaciones y de consultas propias del análisis de datos.
- ❖ **Las herramientas de consulta y análisis:** son sistemas que permiten al usuario realizar la exploración de datos del AD utilizando diversas herramientas de consulta, exploración, análisis y reportes.
- ❖ **Los usuarios:** son aquellos que se encargan de tomar decisiones y de planificar las actividades del negocio, es por ello que se hace tanto énfasis en la integración y limpieza de datos, para poder conseguir que la información posea toda la calidad posible.

### 1.5 Ventajas e inconvenientes

El uso de almacenes de datos como sistemas de información, ha evidenciado un conjunto de ventajas que el desarrollo de este tipo de sistemas brinda a las organizaciones como las que se presentan a continuación [ CITATION Dar17 \l 21514 ]:

- Mejora la entrega de información pues la misma está completa, correcta, consistente, oportuna y accesible.
- Con un gran respaldo de información se puede tomar decisiones rápidamente; así también, la directiva de las organizaciones adquiere confianza en sus propias decisiones y logra un mayor entendimiento de los impactos de estas.

Estas ventajas demuestran que utilizar almacenes de datos facilita en las organizaciones el proceso de toma de decisiones. Sin embargo, su uso también genera algunos inconvenientes [ CITATION Dar17 \l 21514 ]:

- El desarrollo requiere de la inversión de muchos recursos y tiempo.
- El AD recibe un incremento continuo de requerimientos de usuario, necesitando escalar y evolucionar.

## **I.6 Mercado De Datos**

Los MD poseen las mismas características de un AD, pero a un nivel más específico, ya que contiene diferentes combinaciones y selecciones de los datos que se encuentran en el mismo. Ofrece una mayor personalización de los datos de un departamento, permitiendo un manejo más eficiente de la información histórica, además de una ejecución de procesamiento independiente del resto de los departamentos y un costo de almacenamiento y procesamiento inferior [CITATION Chi10 \l 21514 ].

### **I.6.1 Definición**

Existen definiciones sobre los MD, dadas por especialistas como:

- ❖ **Corey** lo define como “bases de datos multidimensionales orientadas a una materia específica” [CITATION Cor97 \l 21514 ].
- ❖ **González** lo define como “una extensión natural de los AD, siendo una BD con información de interés para un determinado sector de la organización, contribuyendo con los sistemas de apoyo a la toma de decisiones” [CITATION Gon98 \l 21514].
- ❖ **Wolff** lo define como “un conjunto de hechos y datos organizados para soporte decisional basados en la necesidad de un área o departamento específico. Los datos son orientados a satisfacer las necesidades particulares de un departamento dado teniendo solo sentido para el personal de ese departamento y sus datos no tienen por qué tener las mismas fuentes que los de otro *datamart*” [CITATION Wol00 \l 21514 ].

Partiendo de esta base, un MD constituye una tecnología de BD constituida por el repunte de los AD, el cual es un subconjunto de datos orientado a un área de negocio específica, definido por la forma en que el usuario necesita ver la información y cómo quiere que se le presente, con el objetivo de que la misma deje de ser simples datos para convertirse en información que enriquezca la toma de decisiones.

### **I.6.2 Importancia del mercado de datos**

Un MD proporciona un esquema de almacenamiento que permite realizar consultas complejas, por medio de la interacción e interrelación de múltiples entidades, manteniendo solo un repositorio con la información agregada y ordenada, según las necesidades de los usuarios. Además, provee de una *interfase* de consulta que permite al usuario la posibilidad de análisis de la información para la toma de decisiones [CITATION Chi10 \l 21514 ].

## **I.7 Modelos o Esquemas de los Almacenes de Datos**

En un AD, se utiliza un esquema para definir la forma de organizar el sistema con todas las entidades de la base de datos (tablas de hechos, tablas de dimensiones) y su asociación lógica. Los hechos

están compuestos por medidas del proceso de negocio que se pretenden analizar, mientras que las dimensiones nos permiten contextualizar los hechos, agregando diferentes perspectivas de análisis a ellos [CITATION Que22 \l 21514 ].

Las **tablas de hechos** (del inglés *fact tables*) son[CITATION Ram16 \l 21514]:

- ❖ La tabla principal del modelo dimensional.
- ❖ Contienen campos claves que se unen a las tablas de dimensión.
- ❖ Contiene métricas o también llamadas medidas y es aquello que queremos medir o analizar. Generalmente son valores numéricos que se suelen agregar.

Las **tablas de dimensión** (del inglés *dimension table*) son[CITATION Ram16 \l 21514]:

- ❖ Tablas simples des-normalizadas y se unen a las tablas de hechos a través de un campo clave.
- ❖ Los atributos de la tabla de dimensión ofrecen información característica de las tablas de hechos.
- ❖ No hay límite de tablas de dimensión y las dimensiones pueden contener una o varias

### I.8.1 Tipos de Esquemas:

#### 1. Esquema en estrella [CITATION Cas \l 21514 ] :

- ❖ Esquema más simple y efectivo.
- ❖ Una tabla de hechos en el centro rodeada por tablas de múltiples dimensiones se asemeja a una estrella en este modelo.
- ❖ La tabla de hechos mantiene relaciones de uno a muchos con todas las tablas de dimensiones.

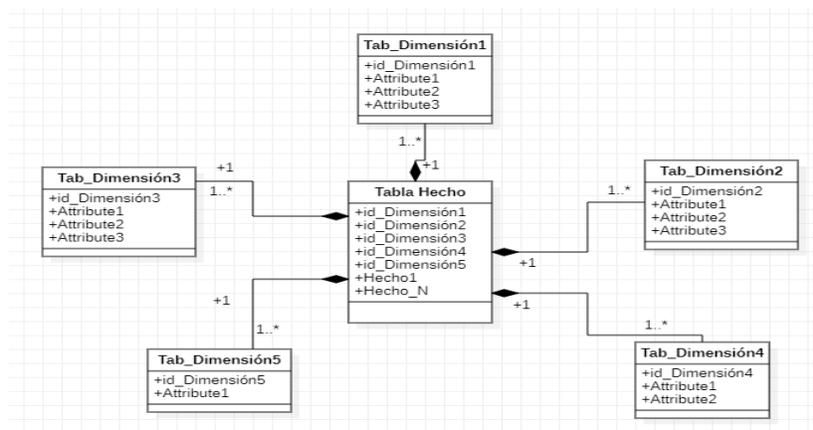


Fig 2- Esquema Estrella [Elaboración Propia]

## 2. Esquema Copo de Nieve [CITATION Cas V 21514 ] :

Este esquema es un proceso que normaliza completamente todas las tablas de dimensiones de un esquema en estrella. La disposición de una tabla de hechos en el centro rodeada por varias jerarquías de tablas de dimensiones se parece a un Copo de Nieve. Cada fila de la tabla de hechos está asociada con sus filas de la tabla de dimensiones con una referencia de clave externa.

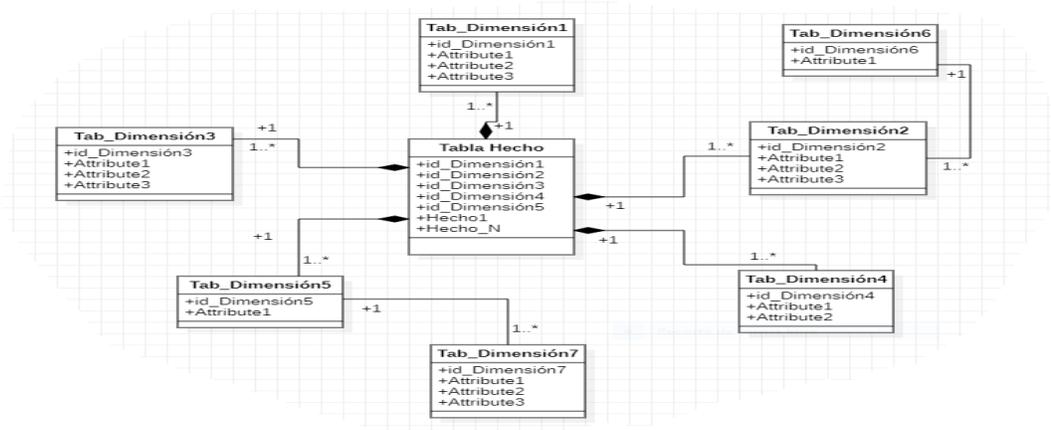
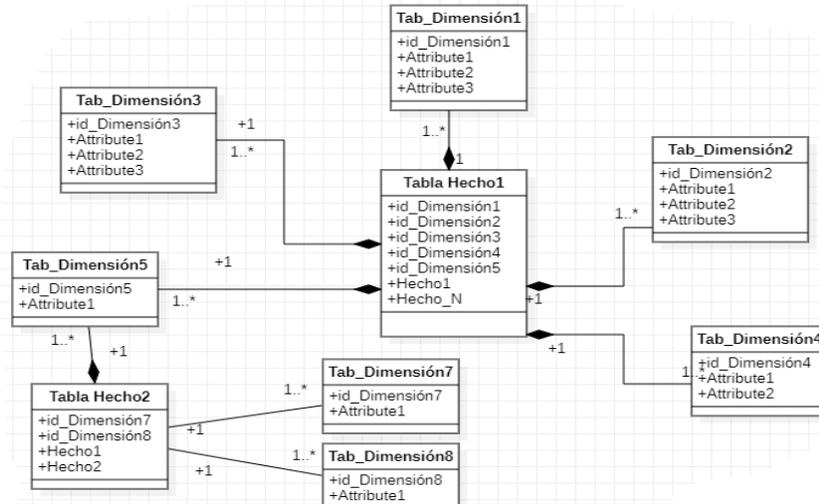


Fig 3- Esquema en Copo de Nieve [Elaboración Propia]

## 3. Esquema de constelación o Galaxia [CITATION Cas V 21514 ] :

Este tipo de esquema se utiliza para requisitos sofisticados y para tablas de hechos agregadas que son más complejas para ser compatibles con el esquema Estrella (o) el esquema Copo de Nieve. Este esquema es difícil de mantener debido a su complejidad. Permite tener más de una tabla de hechos, por lo cual se tendrá mayor capacidad analítica, contribuyendo a la reutilización de las tablas de dimensiones, ya que una misma tabla de dimensión puede utilizarse para varias tablas de hechos.



## I.8 Modo de

Fig 4- Esquema en Constelación o Galaxia [Elaboración Propia]

### almacenamiento OLAP

El procesamiento analítico en línea (**online analytical processing, OLAP**) es un método informático que permite a los usuarios extraer y consultar datos de manera fácil y selectiva para analizarlos desde diferentes puntos de vista. Las consultas de inteligencia empresarial OLAP a menudo ayudan en el análisis de tendencias, informes financieros, previsión de ventas, presupuesto y otros propósitos de planificación [ CITATION OLA22 \l 21514 ].

#### I.8.1 Funcionamiento de los sistemas OLAP

Para facilitar este tipo de análisis, los datos se recopilan de múltiples fuentes de datos y se almacenan en almacenes de datos, luego se limpian y organizan en cubos de datos. Cada cubo OLAP contiene datos categorizados por dimensiones (como clientes, región geográfica de ventas y período de tiempo) derivados de tablas dimensionales en los almacenes de datos. Luego, las dimensiones se completan con miembros (como nombres de clientes, países y meses) que están organizados jerárquicamente. Los cubos a menudo se resumen previamente en todas las dimensiones para mejorar drásticamente el tiempo de consulta en las bases de datos relacionales [CITATION OLA22 \l 21514 ].

#### I.8.2 Tipos de sistemas OLAP

Los sistemas OLAP generalmente se dividen en uno de tres tipos [ CITATION Mor16 \l 2058 ]:

- ❖ **Procesamiento analítico multidimensional en línea (MOLAP)**: provee excelente rendimiento y compresión de datos. Tiene mejor tiempo de respuesta, dependiendo solo del porcentaje de las agregaciones del cubo. La estructura está muy optimizada para maximizar el

rendimiento de las consultas. En general este método, es muy apropiado para cubos con uso frecuente por su rápida respuesta.

- ❖ **Procesamiento analítico en línea relacional (ROLAP):** se utiliza para ahorrar espacio de almacenamiento cuando se trabaja con grandes conjuntos de datos que se consultan con poca frecuencia; por ejemplo, datos exclusivamente históricos, y se usa cuando los clientes desean ver los cambios inmediatamente, o cuando contamos con grandes conjuntos de datos que no son frecuentemente buscados.
- ❖ **Procesamiento analítico en línea híbrido (HOLAP):** es una combinación de ROLAP y MOLAP. Este procesamiento fue desarrollado para combinar la mayor capacidad de datos de ROLAP con la capacidad de procesamiento superior de MOLAP. Ayuda a reducir costos de hardware, ya que se necesita menos espacio en disco que las bases de datos relacionales. Sin embargo, los datos multidimensionales deben ser cargados antes de ser consultados y actualizados cuando se actualizan los datos de la organización.

Luego de conocidos los diferentes modelos de almacenamiento de OLAP y atendiendo a sus características se decide utilizar ROLAP, debido a las políticas del centro CREAD y la universidad en cuanto a la utilización de software libre.

### **I.9 Metodologías para el desarrollo de almacenes de datos**

Al elegir una metodología para desarrollar un almacén de datos no se deben utilizar metodologías que requieran fases extensas de reunión de requerimientos y análisis, fases de desarrollo monolítico que conlleve demasiado tiempo y fases de despliegue muy largas. El objetivo de cada desarrollador debe ser entregar una primera implementación que satisfaga una parte de las necesidades y motivar a los usuarios, es por eso que se debe elegir una metodología que cumpla con estos requisitos, pues el trabajo siempre debe estar dirigido a mejorar la calidad y aceptación del mismo por los usuarios que benefician [CITATION Ing13 \l 21514 ].

Varias personas e instituciones se han dado la tarea de estudiar el tema sobre la metodología correcta para la construcción de un AD, algunos han creado sus propias metodologías, otros han modificado las existentes, entre ellas se pueden mencionar:

- ❖ **Ciclo de vida Kimball** [CITATION Ing13 \l 21514 ] [CITATION Kim02 \l 21514]: presenta un marco de trabajo como el ilustrado en la Figura 5, en la cual se muestran las diferentes etapas durante todo el proceso de creación del AD.

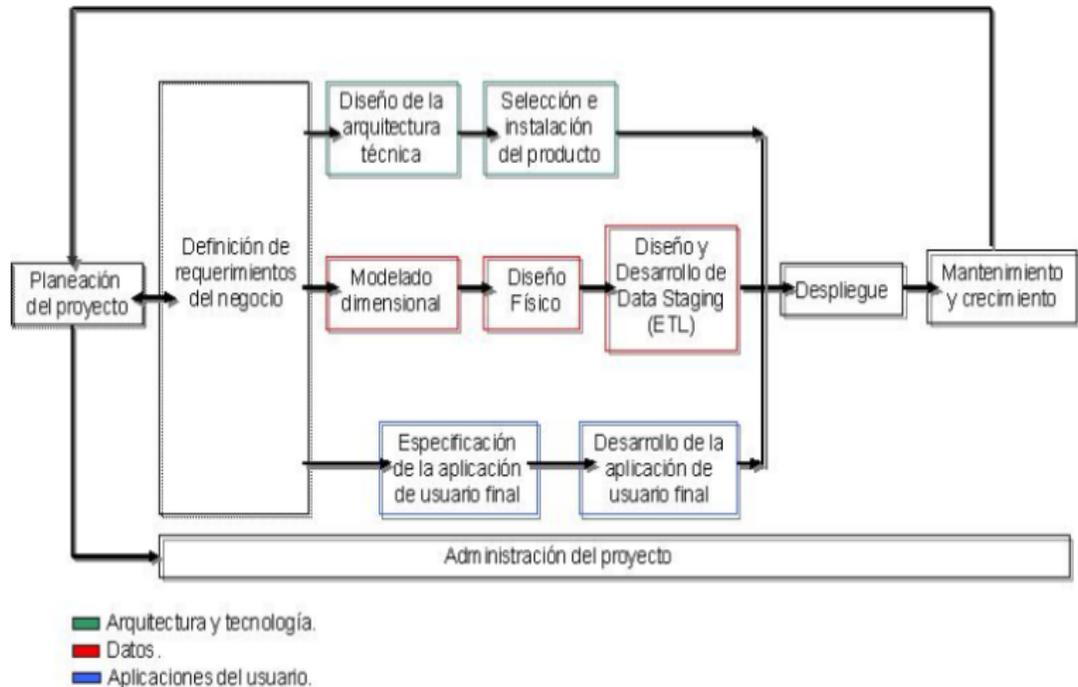


Fig 5- Ciclo de vida Kimball tomado de Metodologías para desarrollar Almacén de Datos.

- ❖ **DWEP (Data Warehouse Engineering Process)** [CITATION Luj05 \ 21514 ]: está basada en el proceso unificado (en inglés: *Unified Process UP*) estándar aceptado en el ámbito científico e industrial para el desarrollo de software.

Entre sus principales características se encuentra que es iterativo e incremental, se basa en cuatro fases de desarrollo y siete flujos de trabajo, está basado en componentes, utiliza el UML (*Unified Modeling Language* -Lenguaje Unificado de Modelado) como lenguaje para modelado gráfico [CITATION Luj06 \ 21514 ][CITATION Dha64 \ 21514 ]. Es orientada a objetos, independiente de cualquier implementación específica, ya sea relacional o multidimensional y permite la representación de todas las etapas del diseño de un AD [CITATION Her10 \ 21514 ].

- ❖ **Juan C. Trujillo** es un Catedrático de Universidad del Dpto. de Lenguajes y Sistemas Informáticos (DLSI) de la Universidad de Alicante. Desde que es Doctor en 2001 ha liderado la línea de investigación de Inteligencia de Negocio y, en 2008 crea el Grupo de Investigación *Lucentia*, que lidera desde entonces. También ha sido coeditor de 11 números especiales de diferentes revistas y, ha sido *Senior Editor* de la revista JCR del Q1 DSS (*Decision Support Systems*) hasta 2017.

- ✓ **La propuesta de Trujillo**[CITATION Tru07 \ 21514 ]: propone una metodología de modelado basada en el proceso unificado. En esta propuesta se emplean los

perfiles de UML como mecanismo para especializar este lenguaje al dominio de los almacenes de datos.

- ❖ **RWM (Rapid Warehousing Methodology)**[CITATION Ing13 \ 21514 ] : es una metodología propuesta por *SAS Institute*. Esta metodología es iterativa, y está basada en el desarrollo incremental de almacén de datos dividido en cinco fases.



*Fig 6- Fases de la metodología Rapid Warehousing Methodology tomado de Metodologías para desarrollar Almacén de Datos*

- ❖ **HEFESTO** [CITATION Dar09 \ 21514 ]: se fundamenta en una amplia investigación, comparación de metodologías existentes y experiencias propias en procesos de confección de almacenes de datos. Consta de cuatro fases: análisis de requerimientos, análisis de los OLTP, modelo lógico del almacén de datos y proceso ETL. Puede ser utilizada en cualquier ciclo de vida que no requiera fases extensas de requerimientos y análisis, con el fin de entregar una implementación que cumpla con una parte de las necesidades proporcionadas por el usuario.
- ❖ **La Metodología de Desarrollo para Proyectos de Almacenes de Datos**, de la MCs. Yanisbel González Hernández, se basa en el ciclo de vida Kimball en la que incluye los casos de uso para guiar el proceso de desarrollo. Cubre las fases por las que transita la construcción de un AD. Los rasgos distintivos que presenta esta metodología es la identificación de requisitos de información y a su vez, la trazabilidad que tienen estos en todo el ciclo de desarrollo del mismo. También se incluye una fase de pruebas que mejora la calidad con que se desarrolla la solución definida [CITATION Gon14 \ 21514 ].

## **I.10 Metodología a utilizar en la solución**

En la presente investigación se define como metodología a utilizar “Metodología de Desarrollo de Proyectos de Almacenes de Datos” propuesta por la MCs. Yanisbel González Hernández [CITATION Gon14 \l 21514 ], debido a que presenta las siguientes características:

- ❖ Un ciclo de vida para el desarrollo de almacenes de datos, claramente definido a partir de la descripción del flujo de trabajo y las fases que presenta.
- ❖ Una descripción de los procesos y actividades que se realizan a lo largo de todo el ciclo de vida de la metodología, relacionándolos con los roles, artefactos y herramientas necesarios para su ejecución, según las características de este tipo de soluciones.
- ❖ Es la utilizada en CREAD para el desarrollo de proyectos de almacenes de datos.

### **I.10.1. Objetivos de la Metodología**

- ❖ Proveer una guía referencial de la forma en que debe organizarse el proceso de desarrollo y las actividades que deben ser realizadas para garantizar la correcta realización de un proyecto de AD [CITATION Gon14 \l 21514 ].
- ❖ Facilitar la adaptación y formación de los especialistas y estudiantes que se enfrentan por primera vez al desarrollo o liderazgo de un proyecto.
- ❖ Establecer un marco para la gestión del proyecto como parte del proceso de desarrollo de las soluciones.

### **I.10.2 Fases del ciclo de vida de la metodología**

El ciclo de vida de la metodología se divide en siete fases y un flujo de trabajo que se puede observar en la Figura 7. Algunas fases podrán ser implementadas de forma paralela como es el caso de la fase de Requisitos y Arquitectura, además durante la fase de Diseño e Implementación podrán desarrollarse varios componentes al mismo tiempo, esto permite un desarrollo más ágil. El flujo de trabajo de Gestión del Proyecto se ejecuta durante todo el ciclo de vida del proyecto [CITATION Gon14 \l 21514 ].

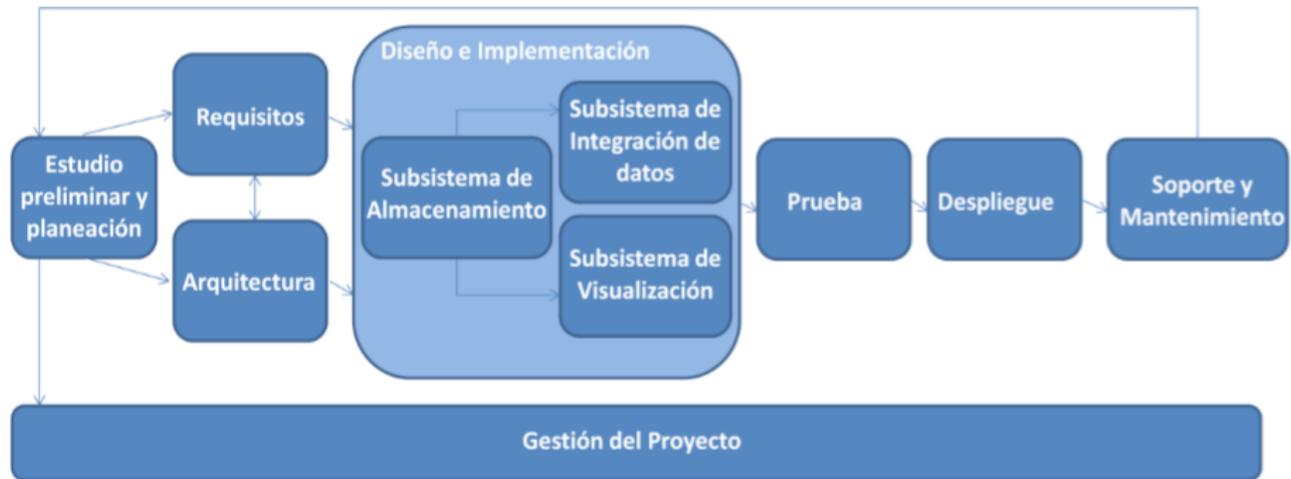


Fig 7- Ciclo de vida de la Metodología de Desarrollo de Proyectos de Almacenes de Datos.

- ❖ **Estudio preliminar y planeación:** se realiza un estudio integral de la organización, que se basa en el diagnóstico del negocio, los datos y la infraestructura tecnológica. Además, se definen una serie de aspectos importantes en la gestión de proyecto, tales como el alcance del mismo, los riesgos, la calidad, cronograma, presupuesto, costo, etc.
- ❖ **Requisitos:** se identifican las necesidades del cliente, para luego realizar el levantamiento de los requisitos tanto de información, como funcionales y no funcionales.
- ❖ **Arquitectura:** se definen las vistas arquitectónicas de la solución y aspectos como la seguridad, la tecnología, los subsistemas y la comunicación entre ellos, etc.
- ❖ **Diseño e implementación:** se diseñan e implementan los tres subsistemas que conforman un AD, definiéndose el diseño de las estructuras de almacenamiento de datos, los procesos de integración, como el mapa lógico, los cubos OLAP, y se implementa el repositorio de datos, la integración de datos y la presentación de los mismos.
- ❖ **Prueba:** se realizan pruebas de tipo unitarias, integración y sistema con el objetivo de validar la calidad y el funcionamiento del producto.
- ❖ **Despliegue:** esta fase se divide en dos etapas. Una primera etapa, que consiste en un despliegue piloto, donde se configuran los servidores, se instala las herramientas en base a la arquitectura definida y se carga una muestra de los datos en un ambiente controlado que le demostrará al cliente la funcionalidad de la solución, para luego realizar la carga histórica de los datos. La segunda etapa consiste en la capacitación y transferencia tecnológica de la solución al cliente. Esta etapa tiene como finalidad, una solución desplegada en un entorno real y con un correcto funcionamiento.

- ❖ **Soporte y mantenimiento:** esta fase comienza cuando la solución está implementada y en explotación. Tiene como objetivo evitar que el sistema quede fuera de servicio debido a fallos en su funcionamiento u obsoleto.
- ❖ **Gestión de proyecto:** es un flujo de trabajo que se ejecuta durante todo el ciclo de vida del proyecto, constituyendo la columna vertebral del mismo, pues es aquí donde se controla, gestiona y chequea el desarrollo del AD, el cronograma y otras actividades de gestión y administración de proyecto.

### **I.11 Tendencias**

Tomar mejores decisiones de negocios rápidamente es la llave para el éxito en el mercado competitivo que se vive actualmente. Comprendiendo esto, las organizaciones están buscando mejorar sus sistemas de toma de decisiones ya que pueden ser rebasados por el volumen y la complejidad de los datos disponibles provenientes de sus sistemas de producción y transaccionales. El permitir que todos estos datos estén disponibles para toda la audiencia a lo largo de la empresa es hoy por hoy uno de los retos más significativos para los profesionales involucrados en las tecnologías de información[CITATION Mic98 \l 3082 ].

Las diferentes compañías que existen hoy en el mundo se han vuelto más digitalizadas, siendo prácticamente imposible evitar el uso de AD. Los sectores donde principalmente se ha implantado almacenes de datos son las empresas de telecomunicaciones, empresas de transporte aerolíneas, transporte de cargas, transporte de pasajeros, empresas de fabricación de bienes de consumo masivo, empresas aseguradoras [CITATION Rey15 \l 3082 ].

#### **I.11.1 Tendencias actuales de los almacenes de datos en el mundo**

El mundo se está digitalizando a gran velocidad, así que la búsqueda de diferentes técnicas para tratar datos heterogéneos es un desafío para las empresas que buscan sacar el máximo partido a los avances más recientes de tecnologías. Los nuevos tipos de bases de datos se admitirán diferentes modelos de tablas con columnas más anchas o series de tiempo, almacenamiento y análisis de datos estructurados, semi-estructurados y no estructurados. Todo ello, tendrá como resultado un rendimiento mayor de los dispositivos inteligentes que requieran almacenamiento de datos de múltiples modelos, el análisis de metadatos, datos de series de tiempo y registros de dispositivos[CITATION Sof21 \l 3082].

Tal es el caso de:

- ❖ La oficina nacional de estadística (ONE) en la República Dominicana pone a disposición de la ciudadanía el sub-portal del almacén de datos del sistema estadístico nacional (SEN), valiosa herramienta que concentra las estadísticas, publicaciones y bases de datos relevantes, así

como indicadores para la toma de decisiones y la elaboración de políticas públicas[CITATION Pab13 \l 2058].

- ❖ Almacén de datos para la prestación del servicio público de información estadística, Ricardo Luján Salazar, instituto nacional de estadística, geografía e informática de México. El Instituto nacional de estadística, geografía e informática (INEGI) tiene la responsabilidad de coordinar los sistemas nacionales estadístico y de información geográfica de México [ CITATION Sal091 \l 2058 ].

### **I.11.2 Tendencias de Almacenes de Datos en Cuba**

Cada día se hace más evidente en el mundo la brecha entre las naciones desarrolladas y subdesarrolladas, debido a los obstáculos para acceder a las tecnologías de punta. Para el país; anticiparse al desarrollo tecnológico es un reto, es por ello que son varias las empresas que pueden contar con un almacén de datos para ayudar a mantener un historial de su información y apoyar a la toma de decisiones de las mismas, entre estas se encuentran:

- ❖ Mercados de datos para el análisis estadístico de la información. Su desarrollo facilita la toma de decisiones en el área de Cooperación intencional de la Universidad de las Ciencias Informáticas, identificando y prediciendo tendencias futuras a partir de datos acumulados [CITATION Pom14 \l 3082 ].
- ❖ Mercado de Datos para los departamentos de Procesos Penales y Gestión de Cuadros y Personal de Apoyo de la Fiscalía General de la República de Cuba. Su desarrollo surge con el objetivo de facilitarle a la FGR una forma más eficiente de poder almacenar la información que se gestiona en los departamentos antes mencionados para su explotación y análisis [CITATION Bea15 \l 3082 ].
- ❖ Sistema para la gestión y análisis de datos de una red de sensores inalámbricos basado en un almacén de datos. Este sistema permitió la estandarización de los datos y su almacenamiento en una estructura multidimensional. Permitted analizar los datos desde diferentes escenarios históricos, y evaluar su comportamiento y evolución en un ambiente multidimensional, o sea, mediante la combinación de diferentes perspectivas o dimensiones [CITATION Dia19 \l 3082 ].

Estos sistemas que fueron desplegados, han formado parte del gran avance informático que ha ocurrido en Cuba, pero a su vez han quedado obsoletos con el pasar del tiempo pues los lenguajes de programación, las herramientas y las formas de representación de información para toma de decisiones entre otros, evolucionan día a día. Debido a estos motivos la presente investigación lleva a cabo el proceso de Actualización del MD para la FGR, con el objetivo de que los casos que en él se

desarrollen tengan los mejores resultados a través del uso de herramientas que cuentan con tecnología más actualizada que apoye de manera satisfactoria al proceso de toma de decisiones.

### **I.12 Tecnologías para el desarrollo del Mercado de datos de la FGR**

Para el desarrollo de un MD es necesario el soporte de un conjunto de herramientas donde se establezca una cooperación entre ellas, para transitar por las diferentes etapas de este proceso, garantizando un producto de calidad. A través de los años, empresas como *Microsoft* y *Oracle Corporation* se han dado la tarea de crear herramientas que facilitan el proceso de desarrollo de almacenes de datos, colocándose estas en la cima del mercado mundial.

Desde el punto de vista económico, la utilización de software privativo en empresas cubanas, tendría un impacto negativo, dado los inaccesibles precios de sus licencias. Teniendo en cuenta lo antes mencionado, la estrategia a seguir para la selección de las tecnologías se basará en el uso de software libre, cuyo código fuente puede ser estudiado, modificado, y utilizado libremente con cualquier finalidad y redistribuido con cambios o mejoras sobre ellas y otorga a los usuarios de manera adecuada las libertades.

### **I.13 Herramienta de modelado de datos**

**Visual Paradigm for UML 16.3:** es una herramienta CASE que soporta el modelado mediante UML y proporciona asistencia a los analistas, ingenieros de *software* y desarrolladores, durante todos los pasos del ciclo de vida de desarrollo de un *Software*.

Es la propuesta por la metodología de desarrollo seleccionada. Permite modelar lógicamente la estructura de un AD, representar las jerarquías con sus niveles asociados, generar el modelo físico a partir del modelo lógico, generar un XML con el diseño preliminar de los esquemas OLAP a partir del diseño lógico, y generar el script de base de datos para cargar el diseño físico en el SGBD [CITATION Gal16 \ 21514].

### **I.14 Sistema gestor de base de datos**

Un SGBD (Sistema Gestor de Base de Datos) realiza las funciones de modificar, extraer y almacenar información de una BD, además de poseer herramientas con funciones de eliminar, modificar, analizar, etc. datos de estas. Realiza la función concreta de interfaz entre la base de datos y los usuarios finales o los programas correspondientes, organizando los datos y permitiendo su acceso [CITATION Car22 \ 21514 ].

Algunas de sus principales características son:

- ❖ Permite una vista muy centralizada y clara de los datos para que sean accedidos de la mejor manera posible.
- ❖ Se encargan de gestionar adecuadamente los datos, evitando a los usuarios o programas que les requieren, tener que entender dónde se encuentran físicamente los datos.
- ❖ Estos sistemas disponen de un lenguaje de programación llamado SQL.

**PostgreSQL14 v4.3:** Es multiplataforma. Soporta distintos tipos de datos, además del soporte para los tipos base, también soporta datos de tipo fecha, monetarios, etc. También permite la creación de tipos propios. Utiliza un modelo cliente/servidor y usa multiprocesos en vez de multihilos para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto y el sistema continuará ejecutándose. Permite el trabajo con grandes volúmenes de datos y admite una alta concurrencia de usuarios accediendo al sistema.

#### **I.14.1 Administrador de bases de datos**

**PgAdmin:** es la principal herramienta de código abierto para la gestión gráfica de PostgreSQL, es decir, es una aplicación gráfica para trabajar con el diseño y administración total de bases de datos PostgreSQL, además está disponible en más de 30 idiomas y es capaz de gestionar diferentes versiones de PostgreSQL, incluyendo también un constructor de consultas, un editor de SQL y un servidor del lado del editor de código.

**PgAdmin 4 v6. 9:** Es una herramienta de administración de código abierto para PostgreSQL, y está diseñado para satisfacer las necesidades de los usuarios y proporcionar una poderosa interfaz gráfica que simplifica la creación, mantenimiento y uso de objetos de base de datos [CITATION pgA20 \l 3082 ].

#### **I.15 Herramienta para el perfilado de datos**

El perfilado de los datos es la actividad de analizar el estado de los datos. La manera de realizar el perfilado depende del nivel de conocimiento que se tenga acerca de la calidad de los datos [CITATION Dat15 \l 21514 ].

**DataCleaner v5.8.1:** es una herramienta que sirve para analizar la calidad de los datos obtenidos, con capacidad para encontrar patrones y supervisar los valores de los datos. Está construida para poder manejar pequeñas y grandes cantidades de datos. Es posible diseñar nuestras propias reglas de limpieza de datos y componerlas en múltiples escenarios distintos o bases de datos objetivo, dichas reglas pueden ser: reglas de búsqueda y/o reemplazo, expresiones regulares, coincidencia de patrones (*pattern matching*) o transformaciones totalmente personalizadas.

### **I.16 Herramienta para la extracción, transformación y carga de los datos**

El proceso ETL, permite a las organizaciones mover datos desde múltiples fuentes, transformarlos y cargarlos en otra base de datos, MD o AD para su análisis [CITATION Cas17 \l 21514].

***Pentaho Data Integration (PDI) v 9.3:*** es una herramienta cuyo nombre en clave es *Kettle*, es una de las herramientas o componentes de *Pentaho Suite* que permite que se utilicen técnicas ETL, es decir, poder implementar procesos de extracción, transformación y carga de datos. *Kettle*, además, ofrece datos analíticos muy precisos, eliminando las complejidades involucradas en la codificación al proporcionar bibliotecas en profundidad para el mismo [CITATION Que221 \l 21514 ].

### **I.17 Herramientas de inteligencia de negocio**

Las herramientas de inteligencia de negocio (BI), facilitan el análisis de los datos, apoyando así el proceso de toma de decisiones en las organizaciones. A continuación, se presentan las herramientas de BI que se utilizarán en la investigación:

***Pentaho Schema Workbench (PSW) v9.3:*** esta es una herramienta e interfaz de diseño visual que permite crear esquemas de cubos OLAP. Permite modelar un XML (en inglés *Extensible Markup Language*) con el diseño del cubo a través de opciones lógicas e intuitivas que no requieren de un manejo avanzado de este formato de archivo y crear, editar [CITATION Sch16 \l 2058 ].

***Pentaho BI Server v9.3:*** esta herramienta permite elaborar reportes de manera dinámica según las necesidades de usuarios finales. Provee el soporte y la infraestructura necesarios para crear soluciones de BI ante problemas de negocio. Presta servicios de autenticación, registro, auditoría entre otros [CITATION Par16 \l 2058].

***Pentaho Report Desing (PRD) v9.3:*** es una herramienta de diseño *desktop*, que permite definir y construir informes, y luego publicarlos en el portal de BI para que puedan ser ejecutados por los usuarios. El resultado de los informes que se vayan diseñando se puede ver con las opciones de pre-visualización, y permite la salida de resultados en diferentes formatos como PDF, HTML, XLS, RTF y CSV [CITATION Pen18 \l 2058 ].

### **I.18 Conclusiones del capítulo**

En este capítulo se definieron los conceptos de AD y MD, así como sus principales características, además de sus ventajas y desventajas, lo que permitió un mejor entendimiento de que la diferencia entre ellos es solamente en cuanto al alcance, debido a que mientras un AD es un sistema centralizado con datos globales de la empresa y de todos sus procesos operacionales, un MD es un subconjunto temático de datos, orientado a un proceso o un área de negocio específica.

Se realizó un estudio de las principales metodologías existentes para el desarrollo de los mercados

de datos, el cual permitió identificar a la “Metodología de Desarrollo para Proyectos de Almacenes de Datos” como la más adecuada para la realización del proyecto, debido a que solo se dispone de la información referente a las Áreas de Análisis de la FGR.

Se conocieron los distintos tipos de sistemas OLAP y según las características de cada uno se decidió utilizar ROLAP, donde se decide utilizar como sistema gestor de base de datos (SGBD) la herramienta *PostgreSQL* para almacenar la información pues este además de ser de código abierto, permite modelar bases de datos relacionales.

Se identificaron las herramientas para el desarrollo de los mercados de datos ya que teniendo en cuenta sus características y funcionalidades se puede asegurar que son adecuadas para implementar de manera eficiente, la solución propuesta para resolver los problemas que tienen hoy en día las fiscalías del país.

## **CAPÍTULO II: ANÁLISIS Y DISEÑO DEL MERCATO DE DATOS DE LA FGR**

En el presente capítulo se realiza un análisis del negocio, con el propósito de comprender los principales aspectos de relevancia para los procesos fiscales que se desarrollan en la FGR. Se definirán los requisitos de información y también según las necesidades del cliente los requisitos funcionales y no funcionales para luego diseñar el diagrama de caso de uso del sistema con las especificaciones correspondientes. Se identificarán los hechos, medidas y dimensiones para las Áreas de Análisis: Dirección de Procesos Penales (DPP), Dirección de Verificación Fiscal (DVF), Dirección de Atención a los Ciudadanos (DAC), Dirección a la Protección a la Familia y Asuntos Jurisdiccionales (DPFAJ), Dirección de Control de la Legalidad en Establecimientos Penitenciarios (CLEP), Dirección de Cuadros (DC), Dirección de Comunes (DCO), Dirección de Comunicaciones (DIC), teniendo en cuenta las necesidades de información de los usuarios, conformando el modelo lógico de datos de la solución. Se definirá la arquitectura a utilizar para el desarrollo del MD, definiéndose el diseño correspondiente a cada subsistema identificado, además de establecerse las políticas de respaldo y recuperación en caso de ocurrencia de fallos. Se diseñarán e implementarán reportes, vistas de análisis y cuadros de mando como parte para la representación gráfica de los indicadores más importantes.

### **II.1 Estudio preliminar del negocio. Caracterización de los Procesos Fiscales**

La Fiscalía General de la República es el órgano del Estado que tiene entre sus misiones velar por el estricto cumplimiento de la Constitución, las leyes, entre otros, con el objetivo de llevar a cabo cada una de sus misiones es necesario que cada proceso fiscal se realice de una manera correcta. El SAPMI es una herramienta que responde a las necesidades de la FGR, permitiendo la recolección de información utilizando mecanismos seguros para la transferencia de la misma, la cual, a través del uso del sistema de encuestas o reportes los fiscales actualizan cada caso o proceso que se ejecute para así tener constancia de los mismos.

Los especialistas mediante el uso de reportes operacionales cargan y analizan la información que se actualiza a través del sistema de encuesta para que a la hora de una toma de decisión esta sea la correcta según sea necesario en próximos o futuros casos que se lleven a cabo en la misma. Teniendo en cuenta los requerimientos de la FGR, se hace necesario añadir nuevas aristas, crear reportes además de implementar vistas de análisis y cuadros de mandos que permiten la utilización de gráficos para realizar análisis a los datos más significativos en un lenguaje claro y sencillo. Por otra parte, es fundamental realizar comparaciones de la información existente en cada una de las provincias del país, es por ello que se lleva a cabo el proceso de Actualización del MD para la FGR.

## II.2 Requisitos de información

Los Requisitos de Información (RI) son descripciones de lo que el sistema debe hacer: el servicio que ofrece y las restricciones en su operación [CITATION Som11 \l 2058]. Se define como requisito de información solo uno, el cual abarca todos los procesos necesarios para el control de la información de todas las áreas de análisis de la FGR con el objetivo de obtener un perfecto desarrollo de los procesos que se manejan en ellas, este requisito es:

- ❖ **RI-1:** Obtener el valor solicitado por fiscalía, modelo, centro informante, indicador, temática y tiempo.

### II.2.1 Requisitos Funcionales

Los Requisitos Funcionales (RF) según Sommerville son enunciados acerca de servicios que el sistema debe proveer, cómo debería reaccionar el sistema a entradas particulares y de cómo debería comportarse el sistema en situaciones específicas [CITATION Som11 \l 2058]. En esta etapa se identifican las necesidades de información de los usuarios en el desarrollo de la solución.

A continuación, se muestran los RF definidos para el MD:

- ❖ **RF-1:** Extraer datos: mediante consultas al sistema fuente se realiza la extracción de los datos.
- ❖ **RF-2:** Transformar y cargar los datos: se realiza el proceso de transformación para integrar de los datos y luego se insertan en la base de datos destino.
- ❖ **RF-3:** Adicionar reporte: adiciona los reportes que estarán disponibles en el MD.
- ❖ **RF-4:** Eliminar reporte: elimina reportes que están disponibles en el MD.
- ❖ **RF-5:** Visualizar reporte: visualiza los reportes que están disponibles en el MD.
- ❖ **RF-6:** Modificar reporte: modifica los reportes que están disponibles en el MD.
- ❖ **RF-7:** Exportar a PDF: permite llevar a formato PDF el reporte que esté visualizando.
- ❖ **RF-8:** Exportar a Excel: permite llevar a formato Excel el reporte que esté visualizando.
- ❖ **RF-9:** Visualizaciones Gráficas o Vistas de Análisis: permite llevar a cabo un mejor análisis de la información disponible en el MD.
- ❖ **RF-10:** Mostrar Cuadros de Mandos de las áreas de análisis de la FGR: permite la representación gráfica de los indicadores más importantes.
- ❖ **RF-11:** Autenticar usuario: consiste en introducir un usuario y contraseña válidos en el sistema e ingresar al mismo con los permisos concedidos.
- ❖ **RF-12:** Mostrar información sobre las Fiscalías: Permite visualizar los reportes definidos sobre cada fiscalía.

## II.2.2 Requisitos No Funcionales

Los Requisitos No Funcionales (RNF) según Sommerville son limitaciones sobre servicios o funciones que ofrece el sistema. Incluyen restricciones tanto de temporización y del proceso de desarrollo, como impuestas por los estándares. Los requerimientos no funcionales se suelen aplicar al sistema como un todo, más que a características o a servicios individuales del sistema [CITATION Som11 \l 2058].

### 1. Disponibilidad

- ❖ **RNF-13:** Garantizar el cumplimiento de actualización del MD: Se debe actualizar el MD todos viernes a las 5 de la tarde, a través de tareas programadas en el sistema operativo para asegurar que la información esté disponible para el usuario.

### 2. Usabilidad

- ❖ **RNF-14:** El MD deben garantizarle al usuario el rápido acceso a los datos contenidos en las distintas áreas de análisis. De esta manera el usuario podrá analizar de manera rápida la información que solicita.

### 3. Restricciones de diseño

- ❖ **RNF-15:** Utilizar las herramientas para la implementación de la capa de BI definidas durante la investigación. De la suite Pentaho, se usarán los siguientes componentes. *Schema Workbench v9.3, Pentaho BI Server v, Pentaho Report Design v9.3.*

### 4. Interfaz de hardware

- ❖ **RNF-17:** Proporcionar características mínimas de hardware a las estaciones de trabajo. Los PC clientes deben contar con al menos los siguientes requerimientos de hardware:

#### ✓ Servidores

Memoria RAM 8 GB o superior.

Disco duro de 1 TB o superior.

Intel(R) Core(TM) i3-8130U CPU @ 2.20GHz 2.21 GHz o superior.

Tarjeta de red Ethernet.

### 5. Eficiencia

- ❖ **RNF-18:** En cuanto a los tiempos de respuesta para acceder a la información contenida en los mercados de datos, serán como promedio de veinte segundos. Para un trabajo óptimo de los mercados de datos se desglosan los recursos tanto de software como de hardware necesarios.

### 6. Disponibilidad del sistema y Confiabilidad:

- ❖ **RNF-19:** El MD estará disponible durante el horario laboral, efectuándose en períodos de

tiempos definidos el proceso de actualización de la información del servidor local de base de datos con el centro de datos.

#### **7. Exactitud de las respuestas:**

- ❖ **RNF-20:** Las salidas que ofrece el MD como resultado del manejo y procesamiento de la información estarán adecuadamente validadas durante los diferentes hitos de prueba, siendo estas fidedignas y con total exactitud a los valores esperados.

#### **8. Interfaces de software**

- ❖ **RNF 21:** Instalar en las estaciones de trabajo el software necesario para el correcto funcionamiento del sistema.
  - ✓ Las configuraciones de software de las máquinas clientes deben contar al menos con:  
*Firefox 90.0* o superior.
  - ✓ Las configuraciones de software del servidor deben contar al menos con:  
*PostgreSQL 14.3*, como servidor de base datos.  
*PgAdmin IV 4.6*, como administrador de base datos .  
*Pentaho Data Integration 9.3*, donde se utiliza el *PAN* para ejecutar las transformaciones y el *Kitchen* para ejecutar los trabajos.  
*Pentaho BI Server 9.3* como servidor de aplicaciones para la visualización de las vistas de análisis y los reportes operacionales.

#### **II.2.3 Reglas del negocio**

Existen varias definiciones de Reglas de Negocio (RN), algunas sencillas y compactas, otras más amplias y detalladas. Según Ronald G. Ross es simplemente una regla que está bajo jurisdicción del negocio, lo cual significa que estas pueden ser creadas, revisadas y eliminadas cuando el negocio lo estime conveniente [CITATION ROS10 \l 2058 ]. Es decir, las RN son declaraciones que definen aspectos del negocio, a su vez operan y restringen las funcionalidades del sistema [CITATION Cha14 \l 2058 ].

Se definieron un conjunto de reglas del negocio agrupadas en reglas de variables, almacenamiento transformación y visualización. A continuación, se describen algunas de ellas:

- ❖ **Reglas de variables:**

- RN-1:** El valor va a ser igual a la sumatoria de todos los id registrados por fiscalía, modelo, centro informante, temática, indicador y tiempo.

- ❖ **Regla de almacenamiento:**

- RN-2:** Los valores de tipo *varchar (String o Big Number)* tendrán una capacidad de almacenamiento en un rango de 0 a 255 caracteres.

❖ **Reglas de transformación:**

**RN-3:** Cuando las columnas tienen campos vacíos se les debe declarar “Desconocido”, en caso de ser valores de tipo *String* o *Big Number* y “0”, en caso de ser valores de tipo *Integer*.

**RN-4:** Cuando las columnas tienen palabras completas en mayúsculas se cambia cada palabra a la forma correcta y si tienen errores ortográficos estos se corrigen.

❖ **Regla de visualización:**

**RN-5:** El análisis estadístico es realizado por los directivos de la FGR.

### II.3 Modelo de casos de uso del sistema

El modelo de casos de uso del sistema es una forma de representar gráficamente la relación existente entre los actores y casos de uso del sistema (CUS), de manera que sirva de guía para reflejar las metas y funcionalidades que persigue el negocio[CITATION Pre10 \l 2058 ]. Los CUS se identificaron a partir del agrupamiento de los requisitos informacionales y los requisitos funcionales, para un total de 11 CUS y 3 actores que se relacionan con ellos, según sus responsabilidades.

A continuación, se presenta la descripción de las responsabilidades de cada uno de los actores en el sistema y el diagrama de CUS en la tabla 1 y la figura 8 respectivamente. Para la confección del diagrama se utilizaron los patrones de CRUD Total, Múltiples actores-Rol común y Concordancia de Adición.

#### II.3.1 Actores de sistema

Actores	Descripción
Especialista	Se encarga del análisis y la realización de los pedidos de información, así como de la visualización gráfica de los mismos.
Administrador ETL	Se encarga de la realización de los procesos de extracción, transformación y carga de los datos.
Administrador del Sistema	Se encarga de gestionar los cuadros de mandos y los reportes e informes disponibles en el MD, además de gestionar los usuarios del sistema y los roles asignados a los usuarios.

*Tabla 1- Descripción de los actores del sistema.*

### II.3.2 Diagrama de CUS

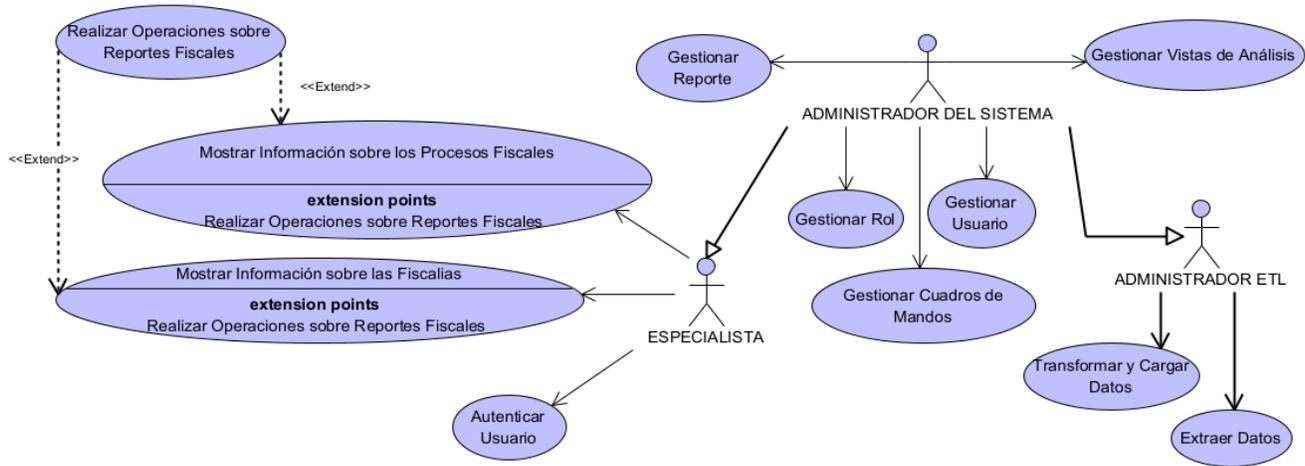


Fig 8- Diagrama de Caso de Uso del Sistema.

### II.3.3 Descripción de casos de uso del sistema

En la tabla 2 se muestra la descripción de los casos de uso definidos en el Diagrama de Caso de Uso del Sistema.

No	Caso de uso del Sistema	Descripción
1	Extraer datos.	Mediante consultas al sistema fuente se realiza la extracción de los datos.
2	Transformar y cargar los datos.	Se realiza el proceso de transformación para que los datos sean correctos y confiables y luego se insertan en la base de datos destino.
3	Mostrar información de los procesos fiscales.	Permite visualizar los reportes definidos sobre los procesos fiscales.
4	Realizar operaciones sobre reportes.	Permite realizar todas las operaciones sobre los reportes definidos.
5	Autenticar usuario.	Permite introducir un usuario y contraseña válidos en el sistema e ingresa al mismo con los permisos concedidos.
6	Gestionar usuario.	Permite adicionar, modificar, eliminar y visualizar un usuario.
7	Gestionar reporte.	Permite adicionar, modificar, eliminar y visualizar un

		reporte.
8	Gestionar rol.	Permite adicionar, modificar, eliminar y visualizar un rol.
9	Gestionar Cuadros de Mandos.	Permite adicionar, modificar, eliminar y visualizar los cuadros de mandos necesarios.
10	Gestionar Vistas de Análisis.	Permite visualizar de la información más significativa de cada proceso.
11	Mostrar Información sobre las Fiscalías.	Permite visualizar los reportes definidos sobre cada Fiscalía.

*Tabla 2- Descripción de Casos de Usos del Sistema.*

### II.3.4 Especificación de casos de uso

En la tabla 3 se muestra la especificación del CU "Mostrar información sobre los Procesos Fiscales de la Dirección de Procesos Penales".

<b>Objetivo</b>	<b>Visualizar reportes sobre los procesos fiscales de la Dirección de Procesos Penales.</b>	
<b>Actores</b>	Especialista (Inicia)	
<b>Resumen</b>	El CU inicia cuando el especialista necesita consultar la información referente a los procesos fiscales. El usuario selecciona la opción Examinar archivos, para luego en la estructura de carpetas mostrada acceder al A.A.G / A.A Procesos Penales. A continuación, selecciona el reporte que desea visualizar y el sistema muestra la información contenida en él. El CU finaliza una vez que el especialista termina el análisis de la información relacionada con los procesos fiscales necesitados.	
<b>Complejidad</b>	Media	
<b>Prioridad</b>	Alta	
<b>Pre-condiciones</b>	<ul style="list-style-type: none"> <li>❖ El especialista tiene que estar autenticado.</li> <li>❖ El MD tiene que estar poblado.</li> <li>❖ Los reportes relacionados con los procesos fiscales de la Dirección de Procesos Penales fueron creados.</li> </ul>	
<b>Post-condiciones</b>	Los reportes correspondientes a la Dirección de Procesos Penales fueron consultados por el especialista.	
<b>Flujo de Eventos</b>		
<b>Flujo básico: Mostrar información sobre los Procesos Fiscales.</b>		
<b>No</b>	<b>Actores</b>	<b>Sistema</b>
1	Selecciona el Área de Análisis General (A.A.G).	

2		Muestra la carpeta contenida en el Área de Análisis General (A.A.G).
3	Selecciona Área de Análisis (A.A).	
4		Muestra la carpeta contenida en el Área de Análisis (A.A).
5	Selecciona el A.A Procesos Penales.	
6		Muestra las carpetas contenidas en A.A Procesos Penales.
7	<p>Selecciona la carpeta deseada:</p> <p>1- Cuadros de Mando (Ver Evento 1).</p> <p>2-Reportes Operacionales (Ver Evento 2)</p> <p>4- Vistas OLAP (Ver Evento 3)</p>	
<b>Evento 1: Cuadros de Mando</b>		
7.1.1	Selecciona la carpeta Cuadros de Mando.	
7.1.2		Muestra las vistas contenidas dentro de la carpeta.
7.1.3	Selecciona la vista que desea analizar.	
7.1.4		Muestra la información contenida en la vista seleccionada y brinda diferentes opciones al especialista para visualizar el cuadro de mando durante su análisis.
7.1.5		Finalizar el CU.
<b>Evento 2: Reporte Operacionales</b>		
7.2.1	Selecciona la carpeta Reporte Operacionales.	
7.2.2		Muestra las vistas contenidas dentro de la carpeta.
7.2.3	Selecciona las vistas contenidas dentro de la carpeta.	
7.2.4		Muestra la información contenida en la vista seleccionada y brinda diferentes opciones al especialista para visualizar los reportes durante su análisis.
7.2.5		Finaliza el CU.
<b>Evento 3: Vistas OLAP</b>		

<b>7.3.1</b>	Selecciona la carpeta Vistas OLAP.	
<b>7.3.2</b>		Muestra las vistas contenidas dentro de la carpeta.
<b>7.3.3</b>	Selecciona la vista que desea analizar.	
<b>7.3.4</b>		Muestra la información contenida en la vista seleccionada y brinda diferentes opciones al especialista para visualizar las vistas OLAP durante su análisis.
<b>7.3.5</b>		Finaliza el CU.

#### Flujos alternos

*No aplica*

Perspectivas de análisis	Posibles resultados	
<b>Variables de entrada relacionada con el caso de uso:</b>	Medidas	Periodicidad
<ul style="list-style-type: none"> <li>✓ Fiscalía.</li> <li>✓ Modelo.</li> <li>✓ Centro Informante.</li> <li>✓ Indicador.</li> <li>✓ Temática.</li> <li>✓ Tiempo (año, mes y fecha).</li> </ul>	Variables de salida disponibles: <ul style="list-style-type: none"> <li>✓ El valor obtenido por fiscalía, modelo, centro informante, indicador, temática, tiempo.</li> </ul>	Rango de tiempo en que se solicitan las variables de salida: <ul style="list-style-type: none"> <li>✓ Semanal.</li> </ul>
<b>Relaciones</b>	CU Incluidos.	No aplica.
	CU Extendidos.	Realizar operaciones sobre Reportes Fiscales.
<b>Requisitos no funcionales</b>	RNF-13, RNF-14, RNF-15, RNF-16, RNF-17, RNF-18, RNF-19, RNF-20	
<b>Asuntos pendientes</b>	----	

#### II.4 Arquitectura del Mercado de datos de la FGR

Definir una arquitectura adecuada influye positivamente en el proceso de desarrollo del MD, por lo que teniendo en cuenta las características generales de un MD se definen los componentes que intervienen en la arquitectura o ambiente en la figura 9:

*Tabla 3- Mostrar información sobre los Procesos Fiscales de la Dirección de Procesos Penales.*



Fig 9 - Arquitectura del MD de la FGR.

La arquitectura del MD está formada por varios elementos que interactúan entre sí y que cumplen una función específica que se resumen de la siguiente manera:

Los datos son extraídos de la Base de Datos sigef la cual se comunica a través del protocolo TCP/IP con el subsistema de integración, que se encarga del proceso de perfilado de los datos mediante el uso de la herramienta *DataCleaner*, este subsistema se comunica de forma bidireccional a través del protocolo TCP/IP con el subsistema de almacenamiento en la etapa de ETL, utilizando la herramienta *Pentaho Data Integration*. Los datos se almacenan en la base de datos destino con apoyo del gestor *PostgreSQL 14* y del administrador *PgAdmin 4*. El subsistema de visualización realiza pedidos de información al de almacenamiento a través del protocolo TCP/IP, este último envía los datos solicitados utilizando el mismo protocolo, para luego mostrar la información a los usuarios finales, para ello utiliza la herramienta *Pentaho BI Server* a través del protocolo de comunicación HTTP. De igual manera los usuarios se comunican a través de este protocolo con el subsistema de visualización para realizar las operaciones sobre los reportes, para ello se utiliza la herramienta *Pentaho Report Design* para la creación de estos reportes.

### II.5 Diseño de los subsistemas que componen el Mercado de datos de la FGR

El diseño del subsistema de almacenamiento comprende la identificación de dimensiones, hechos y medidas del MD, así como la elaboración de la matriz dimensional y la selección de la topología a utilizar en el diseño de la solución. Posteriormente, en el diseño del subsistema de integración se realiza el perfilado de los datos y el diseño de los procesos de integración, sentando las bases para

una correcta implementación de dichos procesos. Finalmente, en el diseño del subsistema de visualización se crean las vistas de análisis quedando diseñada la propuesta de solución.

### II.5.1 Diseño del subsistema de almacenamiento.

En el diseño del subsistema de almacenamiento se identifican los esquemas, las tablas dimensiones y las tablas hechos, dando lugar a la creación del modelo de datos.

#### II.5.1.1 Hechos y medidas

Las tablas de hechos son las tablas primarias en el modelo dimensional. Generalmente, almacenan medidas numéricas, las que representan valores de las dimensiones. La llave de la tabla de hecho, es una llave compuesta, debido a que se forma de la composición de las llaves primarias de las tablas dimensionales a las que están unidas.

En la tabla que se muestra a continuación se presenta la descripción de los hechos:

No	Hecho	Descripción
1	proceso fiscal	Contiene las medidas correspondientes a la cantidad de procesos fiscales llevados a cabo en la FGR.

*Tabla 4- Descripción de la Tabla hechos*

#### II.5.1.2 Dimensiones

Las tablas de dimensiones definen como están los datos organizados lógicamente y proveen el medio para analizar el contexto del negocio. La tabla 5 se muestran las dimensiones se obtuvieron mediante el análisis de los reportes de las Áreas de Análisis de la FGR.

No	Dimensiones	Descripción
1	dim_fiscalia	Representa la información referente a las fiscalías del país, dando lugar sus provincias con sus respectivos municipios y sus nombres.
2	dim_modelo	Representa el contenido de cada proceso fiscal.
3	dim_centro_informante	Representa los diferentes centros informantes de cada fiscalía del país de las cuales se tiene sus nombres y direcciones respectivamente.
4	dim_indicador	Representa los indicadores de cada fiscalía del país, o sea, define el tipo de documento y los indicadores por los que se desea analizar.
5	dim_tematica	Representa las diferentes temáticas que se llevan a cabo en las fiscalías, o sea, establece la temática sobre el proceso fiscal que se desea analizar.

<b>6</b>	dim_tiempo	Define una línea de tiempo específica donde el nivel más bajo es el día y el más alto es el año, en esta dimensión se almacenan los años, meses y días.
----------	------------	---

*Tabla 5- Descripción de las Tablas Dimensiones.*

## II.6 Matriz bus

A partir del análisis de las tablas anteriores se puede construir la herramienta de la metodología denominada matriz bus. La matriz bus representa la relación que existe entre las dimensiones y los hechos del negocio. Constituye una validación del análisis realizado evitando que exista solapamiento entre los hechos. Permite además, identificar los procesos críticos para poder priorizarlos [CITATION Kim13 \ 2058 ].

La matriz diseñada para el Mercado de datos de la FGR, contiene en sus filas y columnas, los hechos y dimensiones identificados respectivamente. Cada X en la intersección de las filas y columnas significa la relación que existe entre los hechos y las dimensiones. En la siguiente tabla, se muestra esta matriz:

H/D	dim_fiscalia	dim_modelo	dim_centro_informante	dim_indicador	dim_tematica	dim_tiempo
proceso_fiscal	x	x	X	x	x	x

*Tabla 6- Matriz Bus*

### II.6.1 Modelo de datos

El Modelo de datos representa una descripción de la estructura de los datos además de proporcionar un modelo independiente de cualquier almacenamiento de datos y métodos de acceso que permita tomar decisiones objetivas de las técnicas de implementación y coexistencia con sistemas antiguos, donde se definen las relaciones que se establecen entre las dimensiones y los hechos que lo componen [CITATION Add \ 2058 ]. Para el diseño del modelo lógico del MD se utiliza la topología de estrella siendo este el más adecuado a las necesidades de los procesos de negocio de la FGR tal y como se muestra en la figura 10:

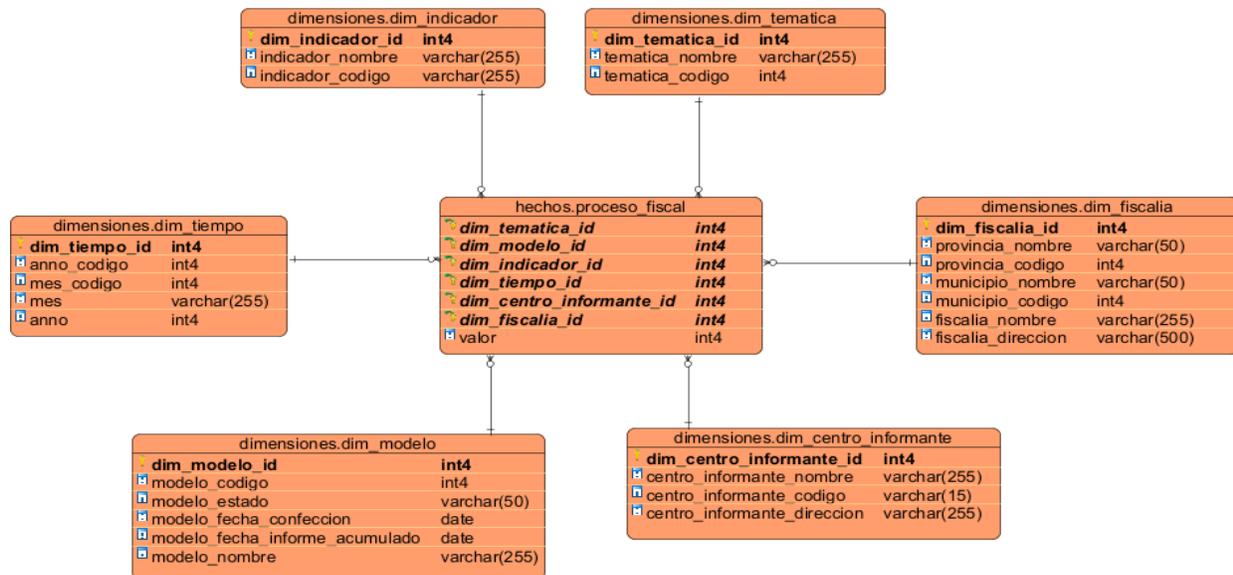


Fig 10- Modelo de Datos.

## II.6.2 Diseño del subsistema de integración

El subsistema de integración comprende el perfilado de los datos, que permite realizar un análisis profundo de los datos provenientes de la fuente para conocer el estado en que se encuentran, así como su calidad y estructura; y la extracción de los mismos desde los sistemas fuentes, los cuales sufren un conjunto de transformaciones.

Para este proceso existen distintas técnicas, las que se definen en el artículo [CITATION Oli14 \l 2058 ] y se describen a continuación:

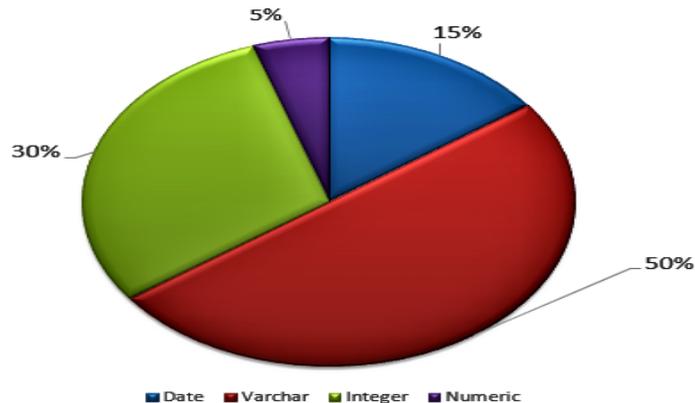
- ❖ **Consolidación de datos:** captura la información de las diferentes fuentes de datos y la almacena en un único repositorio central. Este repositorio puede ser utilizado para la generación de reportes, para los procesos de análisis de la información y toma de decisiones, además, puede actuar como una fuente de datos para diversas aplicaciones. Una de las tecnologías que implementan esta técnica es ETL, que consiste en que la información es extraída desde un origen de datos y pasa por una secuencia de transformaciones antes de ser cargado en un repositorio central.
- ❖ **Federación de datos:** provee una vista virtual única de la información recuperada de diferentes fuentes de datos. El acceso a los datos se realiza en tiempo real y que los mismos se mantienen en su lugar de origen. EII (Integración de Información Empresarial), es una de las tecnologías utilizadas para implementar esta técnica.
- ❖ **Propagación de datos:** consiste en la distribución de datos desde una fuente de información hacia otra, lo que posibilita que la información de ambas fuentes se encuentre siempre

sincronizada. Las fuentes de datos deben ser constantemente actualizadas y este proceso consiste en mover grandes volúmenes de datos de un sistema a otro.

Luego de un análisis de las distintas técnicas de integración, se utilizará la consolidación de datos mediante la tecnología ETL. Esta combinación brinda la posibilidad integrar grandes volúmenes de datos, a través de una serie de transformaciones sobre ellos y el almacenamiento en un repositorio. Este puede ser utilizado para la generación de reportes, para los procesos de análisis de la información y toma de decisiones.

### II.6.3 Perfilado de los datos

El perfilado de los datos permite lograr una mejor comprensión de la información y verificar la existencia de valores nulos, distintos, duplicados, entre otros; permitiendo así, definir nuevas reglas del negocio que posteriormente pasan a ser las reglas de transformación aplicadas en la implementación de la solución.



*Fig 11- Perfilado de Datos.*

En el perfilado realizado a los datos de la fuente se detectó que el 50 % de datos son de tipo Varchar, el 30 % de tipo Integer, el 15 % son de tipo Date, y el 5% restante son del tipo Numeric como se evidencia en la figura 11.

### II.6.4 Diseño de transformaciones

El proceso ETL es el encargado de impulsar el flujo de datos haciendo transformaciones para lograr una integración exitosa. Este componente permite recopilar datos desde múltiples fuentes, formatearlos, limpiarlos y cargarlos en la base de datos, protegiendo su integridad.

#### II.6.4.1 Diseño para la carga de dimensiones

Para este proceso primeramente se verifica la conexión con la base de datos, posteriormente se extraen los datos de la fuente "sigef" que no han sido insertados en el MD y se les realiza la limpieza

y transformación necesaria a los datos, para luego insertarlos en la base de datos destino “mart\_fgr” junto con los valores creados de forma manual.



Fig 12- Diseño para la carga de dimensiones.

#### II.6.4.2 Diseño para la carga de hechos

Después de haber cargado todas las dimensiones se verifica la conexión con la base de datos, posteriormente se extraen los datos de la fuente “SIGEF” que no han sido insertados en el AD y se les realiza la limpieza y transformación que estos necesiten, de forma paralela a la transformación de los datos se calculan las medidas de cantidad para validarlos. Luego de la transformación se buscan las llaves primarias o id de cada una de ellas en la fuente para obtener el nombre del campo. En caso de que las llaves primarias no coincidan con las llaves de la base de datos destino, se les hace las transformaciones necesarias para que sean iguales. En caso correcto se insertan en la base de datos destino “FGR”.



Fig 13- Diseño para la carga del hecho.

#### II.7 Diseño del subsistema de visualización

El diseño del Subsistema de visualización se realiza con el objetivo de organizar las vistas de análisis, reportes y *dashboard* por áreas de análisis además de la realización de los cubos OLAP.

### II.7.1 Arquitectura de la Información

La arquitectura de la información o mapa de navegación, se compone según las necesidades de los usuarios finales, se ha identificado un área de análisis general (A.A.G), el cual contiene las análisis (A.A) de la FGR las cuales a su vez contiene las Vistas OLAP, Cuadros de Mando y los Reportes Operacionales. Estos elementos componen la estructura de navegación de la información presentada en la capa de visualización del MD de la FGR, como se puede apreciar en la figura 14:

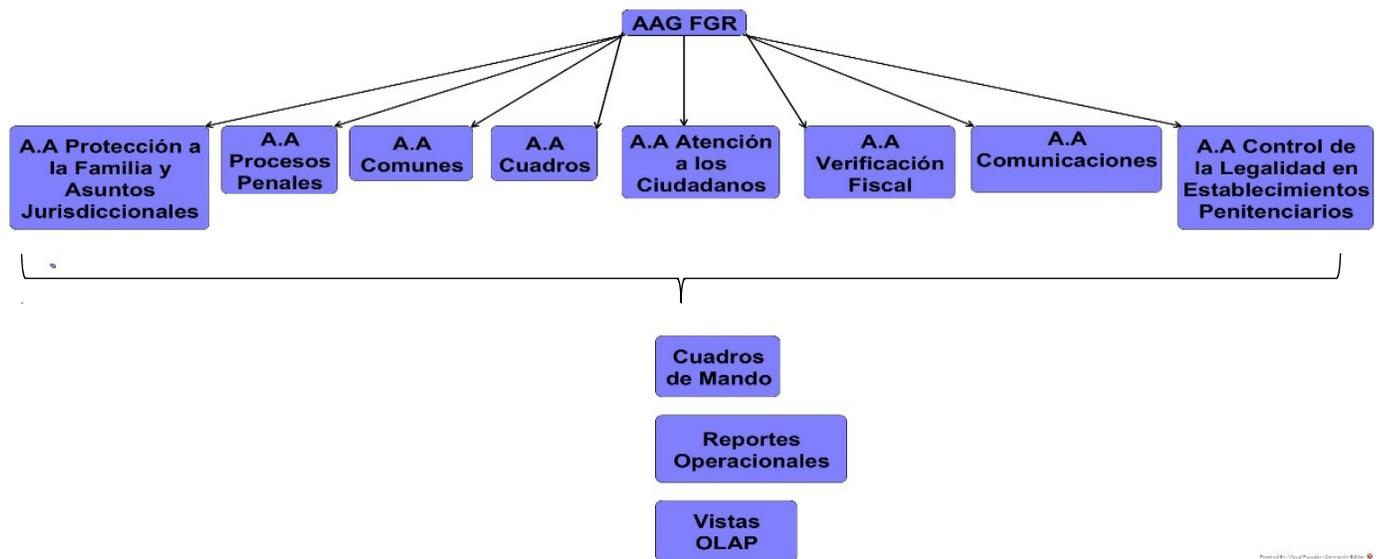


Fig 14- Mapa de navegación.

### II.7.2 Diseño de Reportes

Los reportes candidatos describen detalladamente los reportes preliminares que se deben desarrollar para los usuarios. En la tabla 7, se muestra un ejemplo de reporte candidato en el A.A Procesos Penales, el cual debe reflejar la Cantidad de Procesos Fiscales que llevan a cabo por fiscalía, modelo, centro informante, indicador, temática y tiempo (año y mes).

1	Área de análisis (A.A)	Procesos Penales.
2	Reporte (Tabla de Salida-TS)	Cantidad de Procesos Fiscales llevados a cabo.
3	Descripción	Representa la cantidad de procesos fiscales que se llevan a cabo en las diferentes fiscalías.
4	Elementos de Reportes	Fiscalía, Modelo, Centro Informante, Indicador, Temática, Tiempo (año y mes).
5	Frecuencia de realización	Semanal
6	Funciones	Suma

7	Gráfico	Barras
---	---------	--------

*Tabla 7- Reporte "Cantidad de Procesos Fiscales" del AA Procesos Fiscales.*

## II.8 Esquema de Seguridad

En el MD de la FGR es necesaria la seguridad de la información, pues los datos que maneja son de vital importancia para el país. Con este objetivo se establecieron un conjunto de políticas de respaldo y recuperación ante fallos, además de definirse roles para darle permisos cada uno de los usuarios que interactúan directamente con el sistema.

### II.8.1 Políticas de respaldo y recuperación

En caso de ocurrir fallos en el sistema, ya sea por causas internas o externas, el mismo debe estar preparado para no perder la información que contiene. Se realizarán salvadas a la información contenida en la base de datos del MD de la FGR. Las copias de seguridad se realizarán en dos momentos: semanalmente a través de tareas programadas en el sistema operativo y luego que termine el proceso de consulta popular sobre cada documento definido por la directiva.

### II.8.2 Roles y permisos

Los niveles de acceso al sistema, se limitan por los permisos y roles que a los usuarios se les han sido otorgados para interactuar MD. En la siguiente tabla se muestra la descripción de cada uno de ellos:

Roles	Permisos			
	Aplicación		Base de Datos	
	Lectura	Escritura	Lectura	Escritura
<b>Administrador del Sistema</b>	x	X	x	
<b>Administrador ETL</b>			x	x
<b>Especialista</b>	x			

*Tabla 8- Roles y Permisos.*

## II.9 Implementación del subsistema de almacenamiento

Una vez planteado el modelo dimensional siguiendo una estandarización de los nombres, se dio lugar al modelo físico, permitiendo describir el almacenamiento de los datos y la relación entre las tablas. Además, fueron creados los esquemas, así como las tablas correspondientes a cada uno de ellos. Por lo que, en la implementación del subsistema de almacenamiento, se realiza el desarrollo de la estructura física del MD, además se definen todos los estándares de codificación que van a poseer sus estructuras, para facilitar la comprensión por parte del cliente.

### II.9.1 Modelo de Datos Físico

En la investigación realizada se definieron dos esquemas: hechos y dimensiones; el esquema dimensiones contendrá las dimensiones del MD de la FGR, mientras que el de hechos contendrá el hecho del mismo. Para la implementación de la estructura física se partió del modelo de datos definido en la etapa del diseño del subsistema de almacenamiento, con apoyo de las herramientas Visual Paradigm para la generación del script y PgAdmin 4 para la creación de la base de datos del MD.

En la figura 15 se muestra la estructura del modelo físico del MD de la FGR:

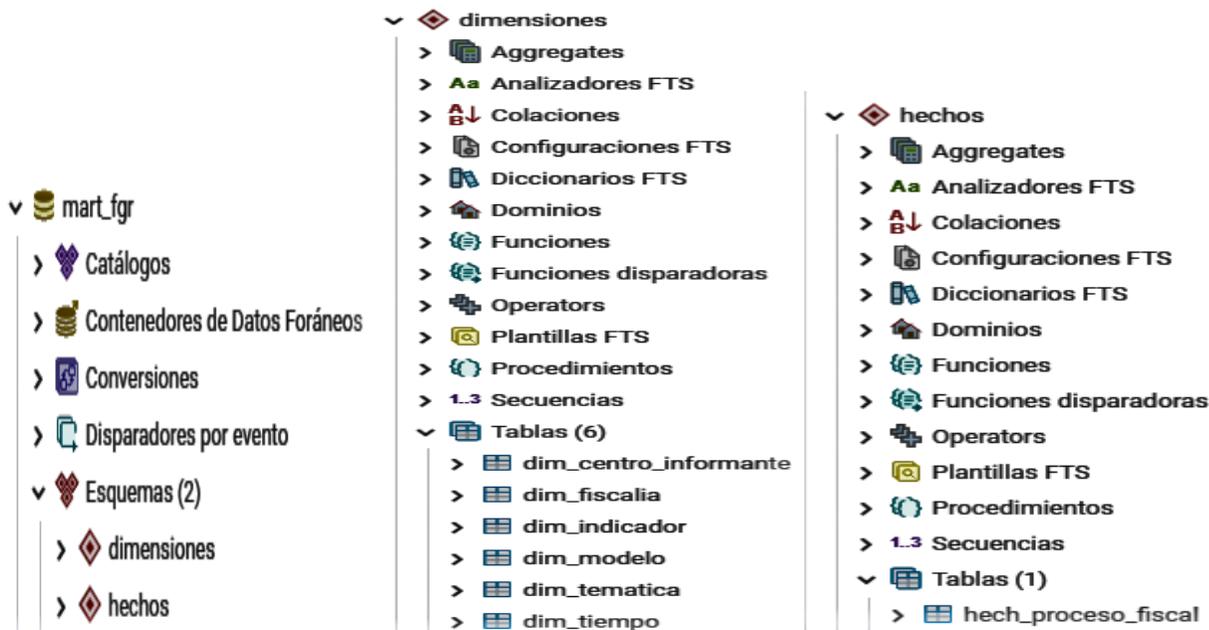


Fig 15- Estructura física del Mercado de datos.

### II.9.2 Estándares de codificación

Los estándares de codificación persiguen el objetivo de organizar la forma en que se nombran las estructuras con el fin de lograr un patrón que contribuya a la correcta normalización de los términos utilizados. Esta codificación está más bien dirigida a los desarrolladores, para que exista un vocabulario común en todo el MD para la FGR, que permita un entendimiento claro.

En la tabla 9 se muestra la descripción de los estándares de codificación utilizados:

Tipo de Objeto	Función	Nomenclatura	Descripción
Esquemas	Hechos	Hechos	Esquema donde se encuentran las tablas de hechos.

	Dimensiones	Dimensiones	Esquema donde se encuentran las tablas de dimensiones.
<b>Tablas</b>	Hechos	hech_[nombre]	Tabla de hechos que contienen los principales procesos con sus medidas asociadas.
	Dimensiones	dim_[nombre]	Tablas de dimensiones utilizadas como perspectivas de análisis.

*Tabla 9- Estándares de configuración para el MD.*

## **II.10 Implementación del subsistema de integración de datos**

El proceso de ETL consiste en extraer los datos de las fuentes y se seleccionan los campos necesarios conforme al modelo de datos, luego estos datos se transforman, limpian y estandarizan para eliminar inconsistencias y posibles errores que pudieran llegar a existir. Se realiza la carga de las dimensiones y hechos que componen el MD para la FGR a través de un grupo de componentes que se encuentran en la herramienta definida en el capítulo uno, teniendo como salida la tabla correspondiente en la base de datos.

También según Bernabéu el proceso ETL consiste en [CITATION Ber17 \l 2058 ]:

- ❖ Extracción: incluye técnicas enfocadas por ejemplo a obtener desde diversas fuentes solamente los datos relevantes.
- ❖ Transformación: incluye técnicas encargadas de compatibilizar formatos, filtrar y clasificar datos, relacionar diversas fuentes, etc.
- ❖ Carga: incluye técnicas propias de la carga de datos y actualización del AD.

### **II.10.1 Transformaciones y trabajos**

Como parte del proceso de ETL, se realizaron las transformaciones correspondientes a las dimensiones identificadas. Esta transformación comienza con la extracción de datos de los centros informantes que tiene la fiscalía, tal como su nombre, código y dirección. Luego se utiliza el componente llamado Renombrar filas para cambiar el tipo de datos del código de los centros informantes de *Character Varying* a *String*. Se crea manualmente una nueva fila para el tratamiento de valores null y se añade al flujo de datos. Por último, se inserta en la base de datos destino (dim\_centro\_informante).

A continuación, en la figura 16, se muestra un ejemplo de transformación para la dimensión dim\_centro\_informante:

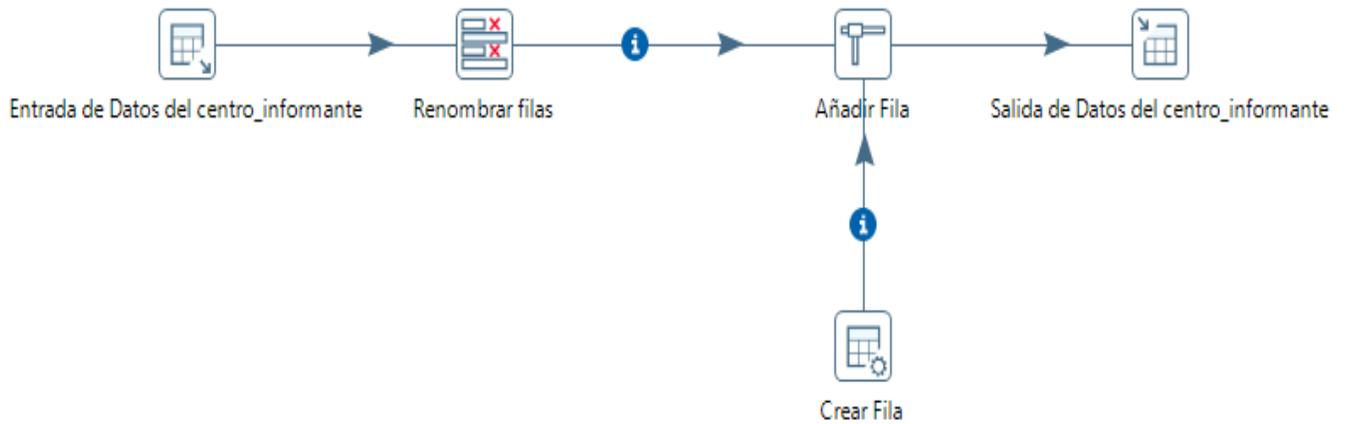


Fig 16- Proceso de integración de datos para la dimensión dim\_centro\_informante.

En la figura 17, se realiza la transformación del hecho identificado en el modelo lógico. En esta transformación, se extrae de la fuente los valores que aún no han sido insertados. Para el tratamiento de los valores null se utiliza el componente Valor Nulo se le da el valor "0" a los códigos, los id y al valor; al año y al mes se les da el valor de "9999 y 12" respectivamente. Luego se utiliza el componente llamado Cambiar Datos para cambiar el tipo de datos del codcentroinformante, idfecha-delinformeacumulado, idmodeloestado, codmodelo, codfila y codome de *Varchar* a *Integer*. Se obtienen los id de las dimensiones. Por último, se ordenan y se agrupan las filas, para luego cargar los datos en la base de datos destino (hech\_proceso\_fiscal).

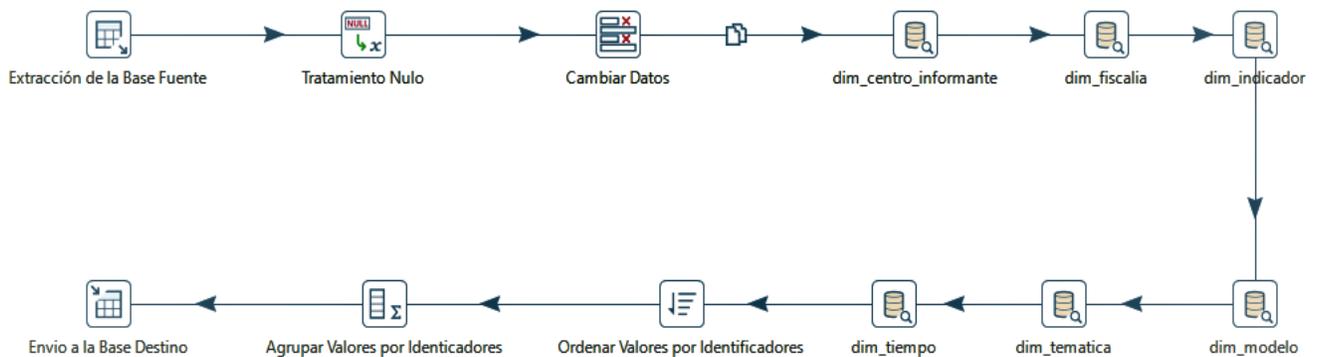


Fig 17- Proceso de integración de datos para el hecho proceso\_fiscal.

Para automatizar el proceso de carga, se implementaron 2 trabajos, uno que ejecuta todas las transformaciones que hacen posible la carga las dimensiones (trab\_dimensiones), el segundo que ejecuta el trabajo anteriormente explicado junto con la carga del hecho (trabajo\_general). Este último, se muestra en la figura 18:

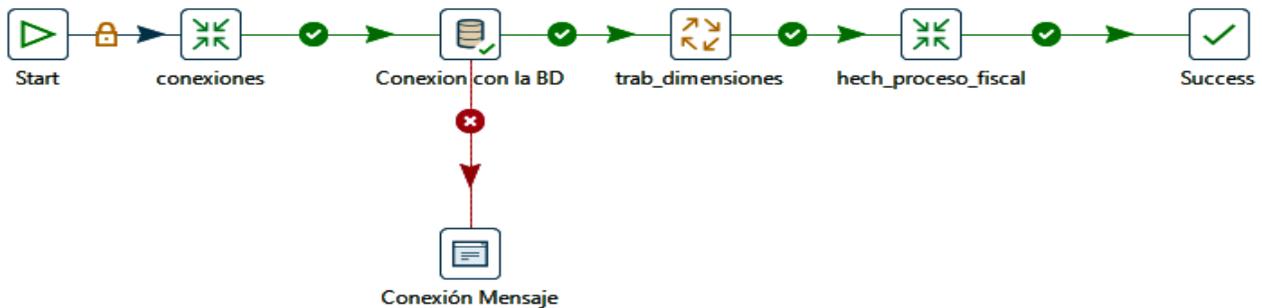


Fig 18 Trabajo\_General.

### II.11 Implementación del subsistema de visualización.

Luego de haber concluido con los subsistemas anteriores, se procede a la implementación del Subsistema de Visualización, que es el responsable de mostrar el resultado final a los usuarios. Para esto se realizan los cubos OLAP en la herramienta *Schema-Worbench*, donde se definieron las dimensiones, medidas y niveles de jerarquía que conforma el Mercado de Datos de la FGR. Se utilizó como base la arquitectura de la información que contiene el Área General, y las Áreas de Análisis de las FGR este se encuentra formado por las vistas de análisis, cuadros de mandos y reportes operacionales.

#### II.11.1 Implementación de los cubos OLAP.

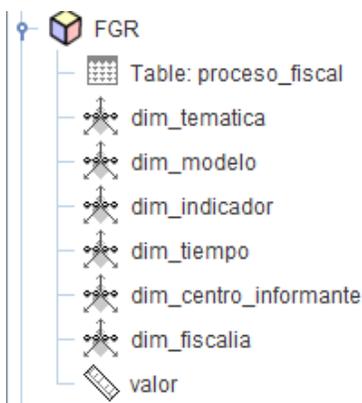
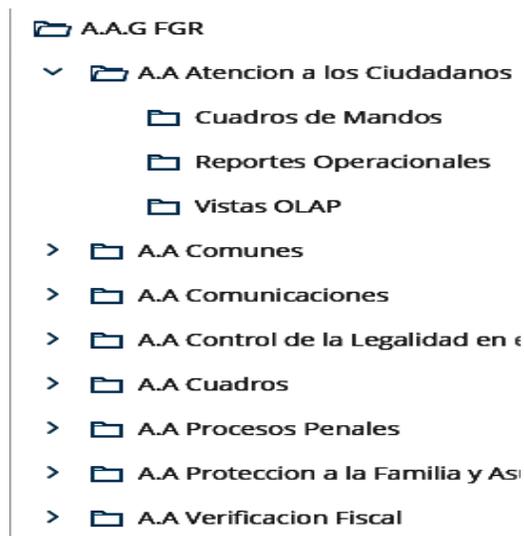


Fig 19- implementación del Cubo OLAP del MD

Como se puede ver en la figura 19, para la implementación de los cubos OLAP del MD de la FGR, se definió un cubo para el hecho identificado, así como las dimensiones, medidas y niveles de jerarquía que conforman el esquema para el MD.

### II.11.2 Implementación de la capa de visualización

El mapa de navegación está diseñado de manera que permita tener una mejor visualización de cómo estará estructurada la información en el MD. A continuación, en la figura 20, se detallan las características de los elementos que componen la arquitectura:



*Fig 20- Arquitectura de visualización de la información.*

- ❖ **A.A.G Fiscalía:** agrupa la información referente a los procesos que se llevan a cabo en la FGR.
- ❖ **A.A Procesos Fiscales:** contiene la información referente a la cantidad de procesos que se llevaron en las áreas de la FGR.
- ❖ **A.A Procesos Penales:** contiene las vistas de análisis referentes a los procesos fiscales que se realizan en la Dirección de Procesos Penales.
- ❖ **A.A Verificación Fiscal:** contiene las vistas de análisis referentes a los procesos fiscales que se realizan en la Dirección de Verificación Fiscal.
- ❖ **A.A Atención a los Ciudadanos:** contiene las vistas de análisis referentes a los procesos fiscales que se realizan en la Dirección de Atención a los Ciudadanos.

- ❖ **A.A Protección a la Familia y Asuntos Jurisdiccionales:** contiene las vistas de análisis referentes a los procesos fiscales que se realizan en la Dirección a la Protección a la Familia y Asuntos Jurisdiccionales.
- ❖ **A.A Control de la Legalidad en Establecimientos Penitenciarios:** contiene las vistas de análisis referentes a los procesos fiscales que se realizan en la Dirección de Control de la Legalidad en Establecimientos Penitenciarios.
- ❖ **A.A Cuadros:** contiene las vistas de análisis referentes a los procesos fiscales que se realizan en la Dirección de Cuadros.
- ❖ **A.A Comunes:** contiene las vistas de análisis referentes a los procesos fiscales que se realizan en la Dirección de Comunes.
- ❖ **A.A Comunicaciones:** contiene las vistas de análisis referentes a los procesos fiscales que se realizan en la Dirección de Comunicaciones.
- ❖ **Cuadros de Mando:** contiene los cuadros de mando referentes de cada A.A.
- ❖ **Vistas OLAP:** contiene las vistas OLAP referentes de cada A.A.
- ❖ **Reportes Operacionales:** contiene los reportes operacionales referentes de cada A.A.

#### **II.11.2.1 Implementación de los reportes operacionales.**

Como parte de la implementación de la capa de visualización, se elaboraron reportes operacionales. Según [ CITATION Dar17 \l 3082 ], estos brindan la posibilidad de generar informes avanzados y detallados del tema de interés que se esté analizando también son informes que organizan y exhiben la información contenida en una base de datos. Su función es aplicar un formato determinado a los datos para mostrarlos por medio de un diseño atractivo y que sea fácil de interpretar por los usuarios. Los reportes fueron elaborados mediante consultas SQL con la herramienta PRD.

La figura 21 muestra un reporte operacional correspondiente al A.A Procesos Penales:

## CANTIDAD DE PROCESOS FISCALES

### Dirección de Procesos Penales (DPP)

FISCALÍA	MODELO	CENTRO INFORMANTE	INDICADOR	MES	AÑO	CANTIDAD
Fiscalía Municipal de Sandino	[DPP-DECI] Del Proceso Penal Ordinario	Fiscalía Municipal de Sandino	Expedientes de Fase Preparatoria radicados	Marzo	2,022	165
Fiscalía Municipal de Sandino	[DPFAJ] - Revisiones	Fiscalía Municipal de Sandino	Expedientes de Fase Preparatoria radicados	Marzo	2,022	227
Fiscalía Municipal de Sandino	[DPP-DECI-DVF-DAC-DCLEP] Formulario sobre hechos contra el ganado mayor (tauro)	Fiscalía Municipal de Sandino	Expedientes de Fase Preparatoria radicados	Marzo	2,022	64
Fiscalía Municipal de Sandino	[DC]-Cuadros	Fiscalía Municipal de Sandino	Expedientes de Fase Preparatoria radicados	Marzo	2,022	16
Fiscalía Municipal de Sandino	[DPP-DECI] Del Proceso Penal Ordinario	Fiscalía Municipal de Sandino	Imputados presentados en EFP	Marzo	2,022	630
Fiscalía Municipal de Sandino	[DPFAJ] - Revisiones	Fiscalía Municipal de Sandino	Imputados presentados en EFP	Marzo	2,022	723
Fiscalía Municipal de Sandino	[DPP-DECI-DVF-DAC-DCLEP] Formulario sobre hechos contra el ganado mayor (tauro)	Fiscalía Municipal de Sandino	Imputados presentados en EFP	Marzo	2,022	482
Fiscalía Municipal de Sandino	[DC]-Cuadros	Fiscalía Municipal de Sandino	Imputados presentados en EFP	Marzo	2,022	14
Fiscalía Municipal de Sandino	[DPP-DECI] Del Proceso Penal Ordinario	Fiscalía Municipal de Sandino	Personas naturales	Marzo	2,022	193
Fiscalía Municipal de Sandino	[DPFAJ] - Revisiones	Fiscalía Municipal de Sandino	Personas naturales	Marzo	2,022	127
Fiscalía Municipal de Sandino	[DPP-DECI-DVF-DAC-DCLEP] Formulario sobre hechos contra el ganado mayor (tauro)	Fiscalía Municipal de Sandino	Personas naturales	Marzo	2,022	51
Fiscalía Municipal de Sandino	[DC]-Cuadros	Fiscalía Municipal de Sandino	Personas naturales	Marzo	2,022	13
Fiscalía Municipal de Sandino	[DPP-DECI] Del Proceso Penal Ordinario	Fiscalía Municipal de Sandino	Mujeres	Marzo	2,022	227
Fiscalía Municipal de Sandino	[DPFAJ] - Revisiones	Fiscalía Municipal de Sandino	Mujeres	Marzo	2,022	160
Fiscalía Municipal de Sandino	[DPP-DECI-DVF-DAC-DCLEP] Formulario sobre hechos contra el ganado mayor (tauro)	Fiscalía Municipal de Sandino	Mujeres	Marzo	2,022	59
Fiscalía Municipal de Sandino	[DC]-Cuadros	Fiscalía Municipal de Sandino	Mujeres	Marzo	2,022	12
Fiscalía Municipal de Sandino	[DPP-DECI] Del Proceso Penal Ordinario	Fiscalía Municipal de Sandino	Menores de 18 años	Marzo	2,022	414
Fiscalía Municipal de Sandino	[DPFAJ] - Revisiones	Fiscalía Municipal de Sandino	Menores de 18 años	Marzo	2,022	308
Fiscalía Municipal de Sandino	[DPP-DECI-DVF-DAC-DCLEP] Formulario sobre hechos contra el ganado mayor (tauro)	Fiscalía Municipal de Sandino	Menores de 18 años	Marzo	2,022	236
Fiscalía Municipal de Sandino	[DC]-Cuadros	Fiscalía Municipal de Sandino	Menores de 18 años	Marzo	2,022	3
Fiscalía Municipal de Sandino	[DPP-DECI] Del Proceso Penal Ordinario	Fiscalía Municipal de Sandino	Mayores de 60 años	Marzo	2,022	35

1 / 7464

*Fig 21- Reporte Operacional del A.A Procesos Penales.*

### II.11.2.2 Implementación de las vistas OLAP.

Una vista es una tabla virtual cuyo contenido está definido por una consulta. Al igual que una tabla, una vista consta de un conjunto de columnas y filas de datos con un nombre. Actúa como filtro de las tablas subyacentes a las que se hace referencia en ella [ CITATION Dar17 \l 3082 ].

Tiempo	Fiscalia	Indicador	Medidas
			• Valor
2022	Pinar del Río	Indicador	47.203.804
		Decisiones sobre la denuncia recibidas	1.480
		Por archivo definitivo al aplicarse criterio de oportunidad	883
		Menores de 18 años de edad a los que se aplicó criterio de oportunidad	18.290
		Otros motivos	636
		Decisiones adoptadas por el fiscal	2.872
		Archivos ratificados	971
		Archivos revocados	1.424
		Pendientes de decidir	2.036
		EFP radicados	4.687
		EFP puestos en curso	2.676
		Imputados presentados al fiscal	1.634
		Personas naturales	847
		Mujeres	2.374
		Menores de 18 años de edad	1.040
		Mayores de 60 años de edad	978
		Extranjeros o cubanos residentes en el exterior	2.180
		Personas jurídicas	875
		Entrevistas realizadas a imputados presentados	1.818

*Fig 22-Vista OLAP referente al A.A Procesos Penales en la Provincia de Pinar del Río.*

En la figura 22 se muestra la vista OLAP implementada que muestra la cantidad de personas reportadas por los indicadores de los procesos fiscales del A.A Procesos Penales ejecutados en la fiscalía de la provincia de Pinar del Río en el año 2022 así como

### **II.11.2.3 Implementación de los cuadros de mando.**

Un cuadro de mando es un modelo de gestión, con un soporte de información periódica para la dirección de la organización, capaz de facilitar una toma de decisiones adecuada sabiendo el grado de cumplimiento de los objetivos, previamente definidos mediante indicadores de control y otras informaciones que lo soporte. Es un medio innovador para comunicar la estrategia y alinear a los miembros de la organización a la ejecución de ésta.

Para presentar la información a los usuarios finales, fueron creados los cuadros de mando correspondiente a cada Área de Análisis para facilitar la toma de decisiones. En la figura 23 se muestra el cuadro de mando del A.A Procesos Penales donde se representa la cantidad de personas reportadas en los 10 indicadores más relevantes en la selección de la provincia de Pinar del Río, en la Fiscalía Municipal de Pinar del Río en el año 2022

## Dirección de Procesos Penales

Año a seleccionar:

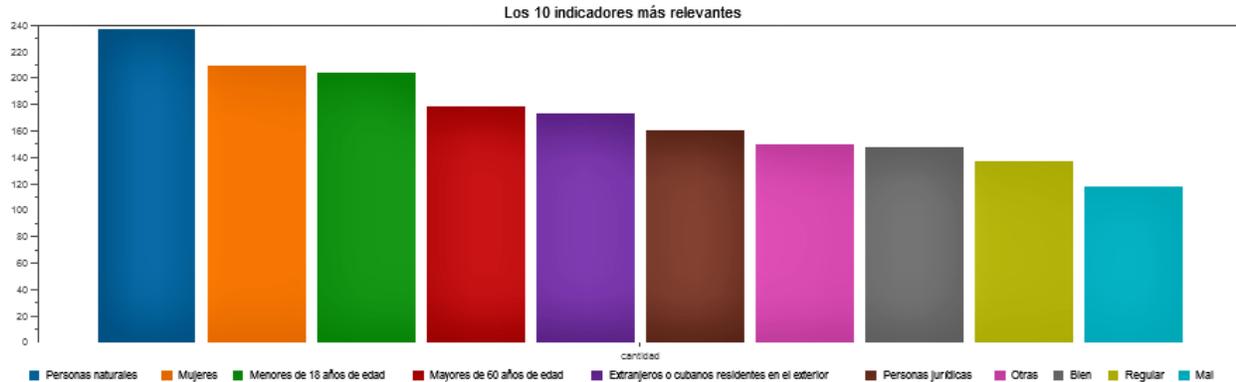
2022

Provincia a seleccionar:

Pinar del Río

Fiscalía a seleccionar:

Fiscalía Municipal de Pinar del Río



*Fig 23- Cuadro de Mando de la Dirección de Procesos Penales.*

En este cuadro de mando, primero se selecciona el año, la provincia y después la fiscalía deseada en la parte superior izquierda mostrándose en un gráfico de barra la cantidad de persona reportadas en los 10 indicadores más relevantes de los elementos seleccionados.

### II.12 Conclusiones del capítulo

Después de haber realizado el análisis y diseño del MD de la FGR, se arribaron a las siguientes conclusiones:

- ❖ Se realizó un estudio de las necesidades de información permitiendo la identificación de 1 RI, 12 RF sirviendo de base para elaborar el diagrama de CUS. De igual manera, se identificaron 8 RNF, agrupadas en 5 Reglas de Variables.
- ❖ En la etapa del modelado de datos, se identificaron 6 tablas dimensionales y 1 tabla de hechos, que garantizan el correcto funcionamiento del sistema.
- ❖ El perfilado de datos realizado a la base de datos "sigef" permitió obtener una noción del estado de la misma, así como el establecimiento de nuevas reglas del negocio aplicables durante el proceso de transformación.
- ❖ Durante el diseño de los subsistemas de integración y visualización, quedó definido el diseño general para las transformaciones, la arquitectura de información, el diseño de los reportes, de los cuadros de mando y las vista OLAP del MD. Todo esto servirá de guía para la

implementación de dichos subsistemas.

- ❖ Se realizarán salvadas a la información semanalmente, como parte de las políticas de respaldo y recuperación para el MD. Además, se definieron niveles de acceso al sistema con la especificación de los roles y permisos.
- ❖ Durante la implementación del subsistema de almacenamiento quedaron definidas las estructuras físicas del MD de la FGR con 7 tablas y 2 esquemas.
- ❖ En la implementación del subsistema de integración, se ejecutaron un total de 7 transformaciones: 6 para la carga de las dimensiones, 1 para la carga de los hechos y 2 trabajos: 1 que carga cada una de las dimensiones y el otro contempla la carga del hecho.
- ❖ A través de la implementación del subsistema de visualización quedaron definidas un A.A.G y 8 A.A y cada una va a contener una carpeta definida para los Cuadros de Mandos, las Vistas OLAP y los Reportes Operacionales, los cuales fueron implementados, permitiendo la visualización de la información.

## **CAPÍTULO III: PRUEBAS DE VALIDACIÓN DE LA SOLUCIÓN PROPUESTA**

En este capítulo se hace referencia a las pruebas aplicadas al MD para verificar que cumpla con las funcionalidades y características definidas.

### **III.1 Pruebas al MD de la FGR.**

Las pruebas de software son un conjunto de actividades que se realizan para descubrir errores que se cometieron de manera inadvertida, conforme se diseñó y construyó el sistema. Las primeras etapas se enfocan sobre cada componente de forma individual. Sobre estos se aplican pruebas para descubrir errores en los datos y en la lógica de procesamiento. Después de probar estos componentes, deben integrarse hasta que se construya el sistema completo. En este punto, se ejecuta una serie de pruebas de orden superior para descubrir errores en la satisfacción de los requerimientos del cliente. Conforme se descubren los errores deben diagnosticarse y corregirse [ CITATION Rog10 \l 3082 ].

#### **III.1.1 Pruebas aplicadas a la documentación del proyecto.**

La verificación de la documentación se realizó con el empleo del artefacto Listas de chequeo, el cual consiste en un conjunto de preguntas que verifican la correcta estructura y semántica del documento, así como los indicadores definidos en los artefactos a evaluar; estos son: Especificación de casos de uso, Especificación de requisitos de software, Reglas de negocio y transformación y Mapa lógico de los datos.

Algunas de las no conformidades (NC) críticas encontradas:

- ❖ Falta por especificar el RF referido a los cuadros de mando.
- ❖ No se define el rango de almacenamiento de los valores de tipo *Integer* y *String*.
- ❖ Incorrecto el nombre de la tabla fuente referida a los centros\_informantes.
- ❖ Falta por especificar el RF referido a la autenticación del usuario.

En la figura 24 se puede evidenciar la relación de cantidad de NC por indicador:



*Fig 24- Resultado de la aplicación de las listas de chequeo*

Teniendo en cuenta los indicadores a evaluar, fueron detectadas y resueltas en una primera iteración, un total de 11 no conformidades (NC), en una segunda iteración 5 NC y para una tercera iteración se detectaron cero NC.

### III.1.2 Pruebas aplicadas al software

Una vez implementado el software, se realizan las pruebas necesarias para verificar la correcta implementación de cada función específica. Su adecuada ejecución arrojaría a resultados útiles para que los responsables del diseño e implementación de la aplicación, puedan detectar donde deben revisar y corregir los errores identificados. Para la realización de las pruebas al MD de la FGR, se tuvieron en cuenta las pruebas definidas por la metodología utilizada, por lo que se realizaron pruebas unitarias, de integración, de sistema y de aceptación.

#### III.1.2.1 Pruebas unitarias

Las pruebas unitarias se encargan de probar una clase en concreto, testeando cada uno de sus métodos y viendo si dado unos parámetros de entrada, la salida es la esperada [ CITATION Gon14 \l 3082 ].

En la aplicación de las pruebas de integración al MD de la FGR, por parte del equipo de desarrollo del proyecto y especialistas del centro CREAD, arrojaron a la detección de las NC que se describen a continuación:

#### **NC detectadas en el subsistema de almacenamiento:**

- ❖ El nombre de la base de datos del MD no se ajusta al proceso de negocio.

### NC detectadas en el subsistema de integración:

- ❖ No se realiza el tratamiento de errores para los valores null en la transformación hech\_proceso\_fiscal.
- ❖ No existe un componente que elimine los espacios en blanco en la transformación dim\_modelo.
- ❖ No se normaliza a mayúsculas los valores de tipo texto en la transformación dim\_centro\_informante.
- ❖ La estructura de carpetas del repositorio de ETL no está correcta.
- ❖ Errores lógicos en la implementación de las transformaciones.

### NC detectadas en el subsistema de visualización:

- ❖ La información graficada en el cuadro de mando Cuadro de Mando de la Dirección de Procesos Penales no se ubica en el tiempo (mes, año).
- ❖ En la realización de los cubos OLAP a la dim\_modelo le falta especificar el campo *primary key*.
- ❖ En la realización del cubo OLAP al dim\_centro\_informante no se le definió correctamente el orden de los elementos que contiene.
- ❖ Los gráficos en los cuadros de mandos no corresponden con la información solicitada.

En la figura 25, se muestra el total de NC encontradas por subsistemas:

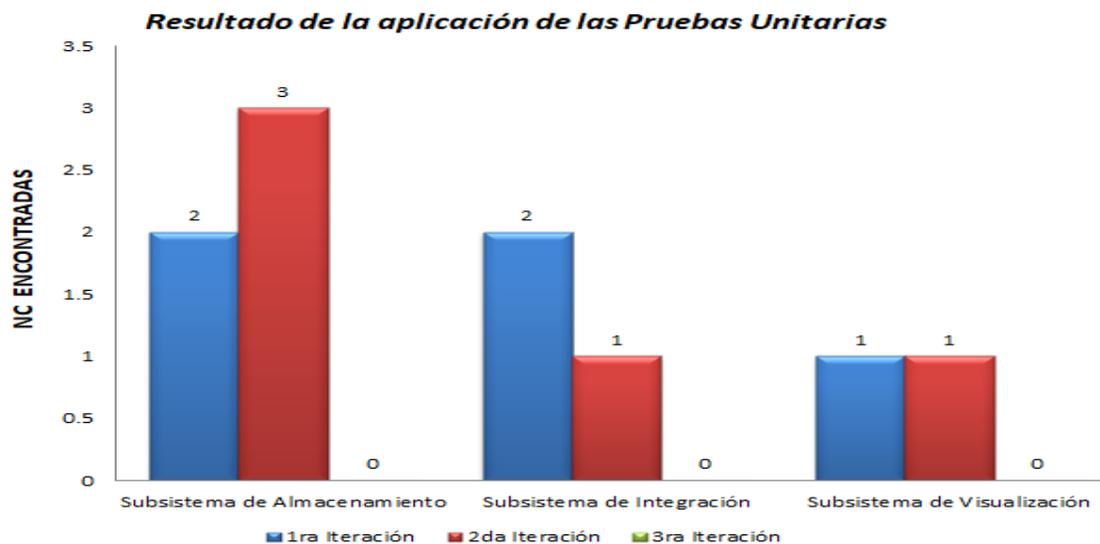


Fig 25- Resultado de la aplicación de las pruebas unitarias.

Teniendo en cuenta los indicadores a evaluar, fueron detectadas y resueltas en una primera iteración, un total de 5 no conformidades NC, en una segunda iteración 3 NC y para una tercera iteración se detectaron cero NC.

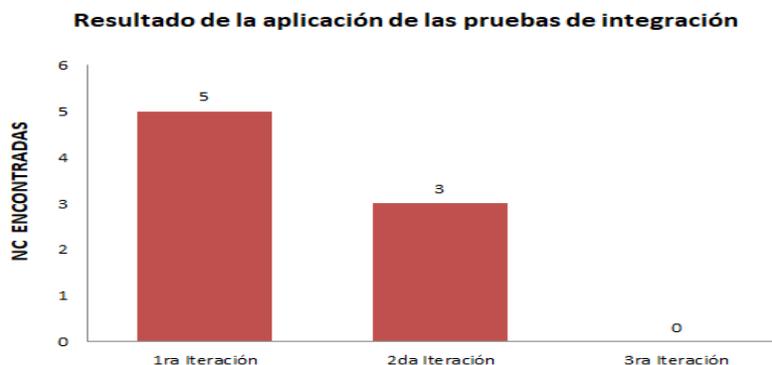
### III.1.2.2 Pruebas de integración

Las pruebas de integración permiten verificar la correcta integración de los componentes y subsistemas que conforman la solución. Ponen a prueba algunas de las vistas arquitectónicas del sistema [ CITATION Gon14 \l 3082 ].

En la aplicación de las pruebas de integración al MD de la FGR, por parte del equipo de desarrollo del proyecto y especialistas del centro CREAD, arrojaron a la detección de las NC que se describen a continuación:

- ❖ En la transformación `hech_proceso_fiscal` realizada en el subsistema de integración no se calcula la medida valor correctamente, reflejándose este error de cálculo en la vista OLAP “V.A Procesos Penales” del subsistema de visualización.
- ❖ En la transformación `dim_modelo` del subsistema de integración, no se realizó un tratamiento para los valores nulos, por lo que estos datos no pueden ser cargados en el MD del subsistema de almacenamiento, debido a que la base de datos destino contiene la restricción *not null* para todos los campos.
- ❖ El valor del parámetro de conexión usuario en el subsistema de almacenamiento, no coincide con el definido en el subsistema de visualización, por lo que no se pueden visualizar los datos.

En la figura 26, se muestra el total de NC encontradas:



*Fig 26- Resultado de la aplicación de las Pruebas de Integración.*

Teniendo en cuenta los indicadores a evaluar, fueron detectadas y resueltas en una primera iteración, un total de 5 no conformidades NC, en una segunda iteración 3 NC y para una tercera iteración se detectaron cero NC.

### III.1.2.3 Prueba de sistema

Las pruebas de sistema permiten validar el cumplimiento de los requisitos de información y funcionales definidos por los clientes[ CITATION Gon14 \l 3082 ]. La aplicación de las pruebas de sistema al MD de la FGR, fueron realizadas por el equipo de desarrollo del proyecto.

A continuación en la tabla 10 se muestra el escenario 3 del CU “Mostrar información sobre los procesos fiscales del A.A Procesos Penales”, seguido de las NC detectadas con el mismo:

EC 3: Muestra información sobre los procesos fiscales archivados.	Muestra la cantidad de procesos fiscales por fiscalía, modelo, centro informante, indicador, temática y tiempo(año, mes).	Fiscalía	Procesos Fiscales	El sistema muestra todas las variables por las cuales se puede filtrar la información para obtener los procesos fiscales archivados del A.A Procesos Penales.	El usuario se autentica, selecciona el A.A.G / A.A / A.A Procesos Penales y selecciona el reporte, la vista o cuadro de mando según necesite el mismo.
---	---	----------	-------------------	---	--

*Tabla 10- Escenario 3 del caso de uso “Mostrar información sobre los procesos fiscales del A.A Procesos Penales”*

#### **Clasificación de la NC: Correspondencia**

- ❖ El reporte operacional “Cantidad de Procesos Fiscales de A.A Procesos Penales” no muestra el valor de las Fiscalías.

#### **Clasificación de la NC: Funcionalidad**

- ❖ La opción de exportar a pdf de la barra de herramientas de la vista OLAP, no se ejecuta.
- ❖ La opción de exportar a excel de la barra de herramientas de la vista OLAP, no se ejecuta.

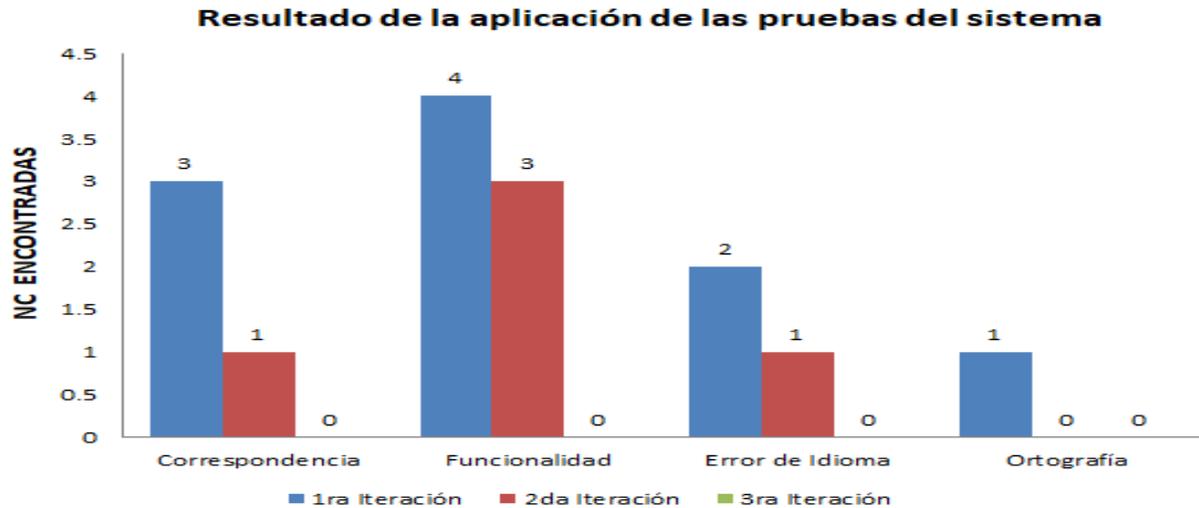
#### **Clasificación de la NC: Error de idioma**

- ❖ En la interfaz de logueo del sistema, los textos se muestran en idioma inglés en ves del idioma español.

#### **Clasificación de la NC: Ortografía**

- ❖ Error ortográfico en los nombres de los cuadros de mando y reportes.

En la figura 27, se muestra el total de NC encontradas por clasificación:



*Fig 27- Resultado de la aplicación de las pruebas del sistema.*

Teniendo en cuenta los indicadores a evaluar, fueron detectadas y resueltas en una primera iteración, un total de 10 no conformidades NC, en una segunda iteración 6 NC y para una tercera iteración se detectaron cero NC.

### III.1.2.3 Pruebas de aceptación

Las pruebas de aceptación son pruebas funcionales, pero vistas directamente desde el cliente, son aquellas pruebas que demuestran al cliente que la funcionalidad está terminada y funciona de forma correcta [ CITATION Gon14 \l 3082 ].

Para verificar que el MD de la FGR cumple con los objetivos propuestos, se realizaron pruebas de aceptación por parte de los especialistas del centro CREAD, las cuales se ejecutaron en una máquina de escritorio que cuenta con las siguientes prestaciones:

- ❖ 8 GB RAM
- ❖ Procesador i3

En la ejecución de las pruebas de aceptación se encontraron 2 NC en una primera iteración, las cuales fueron resueltas. En una segunda iteración, se encontraron cero NC, quedando los especialistas satisfechos con los resultados obtenidos y dando su aprobación final. Las NC encontradas se describen a continuación:

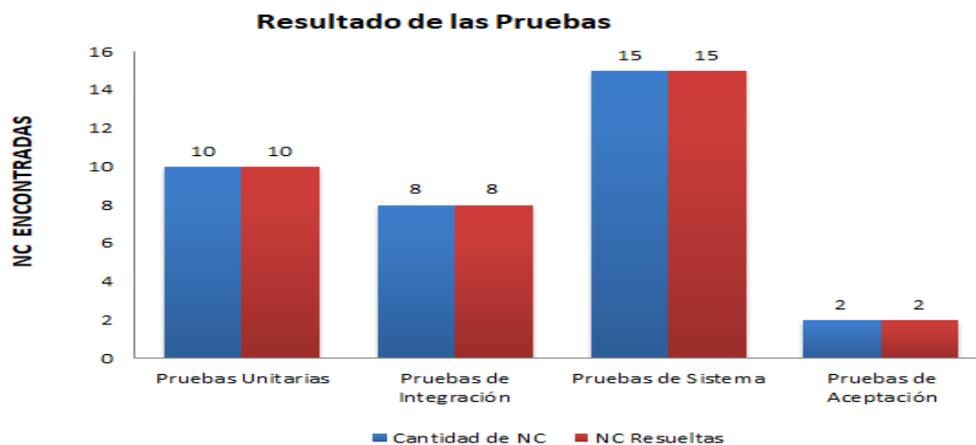
- ❖ La fuente de letra definida no es muy legible y los colores son muy opacos.
- ❖ El color de la barra de navegación de la aplicación, no coincide con el formato del logo de la organización.

A continuación en la figura 28, se muestra un resumen de las NC detectadas:



*Fig 28- Resultado de la aplicación de la Prueba de Aceptación.*

En la figura 29, se muestra un resumen de las NC detectadas en la etapa de prueba:



*Fig 29- Resultado de las pruebas.*

La gráfica representada en la figura anterior muestra un resumen de las NC detectadas en los niveles de pruebas: unidad, integración, sistema y aceptación, en las que se detectaron, 10, 8, 10 y 2 NC respectivamente que fueron resueltas satisfactoriamente.

### III.2 Conclusiones del capítulo

Después de haber realizado las pruebas del MD de la FGR, se obtuvieron las siguientes conclusiones:

- ❖ Se realizaron pruebas al MD con la aplicación de las Listas de chequeo y pruebas de software. Como resultado de estas pruebas se encontraron y corrigieron las NC en la documentación y NC en la etapa de diseño e implementación del MD.

## CONCLUSIONES

Al finalizar el proceso de desarrollo del MD de la FGR se puede afirmar que se le ha dado cumplimiento de forma satisfactoria al objetivo general del presente trabajo de diploma, obteniéndose los resultados que se esperaban. Es por ello que se llega a las siguientes conclusiones:

- ❖ El estudio de los fundamentos teóricos permitió comprender mejor el tema tratado.
- ❖ Se seleccionó para guiar el proceso de desarrollo, la Metodología para el desarrollo de proyectos de AD, la que se adapta a las características de trabajo específicas del centro CREAD y permite un desarrollo más ágil.
- ❖ La selección de las herramientas basadas en software libre permitió la correcta implementación de la solución con interfaces amigables para el usuario.
- ❖ El proceso de entrevistas con el cliente permitió definir los requisitos de la solución, identificándose así un RI, 12 RF, 8 RNF y 5 RN y se definió la arquitectura de la solución identificándose tres subsistemas.
- ❖ El subsistema de almacenamiento permitió identificar una tabla de hechos y 6 tablas de dimensiones dando lugar al diseño del modelo lógico de datos que permitió la implementación del modelo físico.
- ❖ El subsistema de integración permitió definir las estrategias que deben seguirse para el desarrollo de los procesos de integración de datos, en el que fueron realizadas 7 transformaciones y 2 trabajos.
- ❖ El subsistema de visualización permitió diseñar los cubos OLAP, las vistas de análisis, los reportes operacionales y los cuadros de mando que fueron implementados como parte de la solución.
- ❖ La aplicación de pruebas unitarias, de integración, sistema u de aceptación permitió comprobar que el sistema cumple con las necesidades del cliente.

## **RECOMENDACIONES**

- ❖ Desplegar el Mercado de Datos desarrollado en la FGR.
- ❖ Proporcionar cursos de capacitación a los especialistas de la FGR, para que puedan trabajar con el sistema satisfactoriamente

## REFERENCIAS BIBLIOGRÁFICAS

[En línea] <https://www.uci.cu/universidad/mision>.

*pgadmin4/6.9/release\_notes\_6\_9*. [En línea] [Citado el: 4 de junio de 2022.] [https://www.pgadmin.org/docs/pgadmin4/6.9/release\\_notes\\_6\\_9.html](https://www.pgadmin.org/docs/pgadmin4/6.9/release_notes_6_9.html) .

**2020**. ¿Qué es un almacén de datos? [En línea] 11 de 2020. [Citado el: 22 de 05 de 22.] <https://www.sap.com/latinamerica/insights/what-is-a-data-warehouse.html>.

**Avella, Angela. 2011**. Calameo. [En línea] 2011. [Citado el: 3 de 7 de 2022.] <http://es.calameo.com/books/002299301667571c7ab05>.

**Bernabeu, Darío. 2022**. Guía completa de aplicación teórico-práctica; metodología Data Warehouse. [En línea] 24 de 07 de 2022. [Citado el: 1 de 09 de 2022.] <https://fliphtml5.com/oqraz/eiwg/basic>.

—. **2017**. *Hefesto Datawarehousing Guía completa de aplicación teórico-práctica*. 2017.

**Bernabeu, Dario. 2009**. *HEFESTO: Metodología propia para la Construcción de un Data Warehouse*. Córdoba, Argentina : s.n., 2009.

**Calvo, Diego. 2016**. Tablas de hechos y Dimensiones (inteligencia de negocio). [En línea] Analisis de DATOS, 29 de noviembre de 2016. [Citado el: 1 de junio de 2022.] <https://www.diegocalvo.es/tablas-de-hechos-y-dimensiones-inteligencia-de-negocio/>.

**Carlos, González Alvarado. 1998**. *Depósito de datos*. Costa Rica : Instituto Tecnológico de Costa Rica, 1998.

**Castillo, Juan Manuel. 2017**. LinkedIn. [En línea] 25 de 10 de 2017. [Citado el: 26 de 09 de 2018.] <https://linkedin.com/pulse/que-es-una-etl-juan-manuel-catillo-pinto>. . [En línea] 2017. [Citado el: 4 de junio de 2022.] <https://linkedin.com/pulse/que-es-una-etl-juan-manuel-catillo-pinto>. .

**Cleaner, Data. 2019**. DataCleaner. [En línea] 2015. [Citado el: 02 de 05 de 2019.]. [En línea] 2019. [Citado el: 4 de junio de 2022.] <https://datacleaner.org/docs..>

**Community, Pentaho. 2018**. Pentaho Documentation. [En línea] 10 de 10 de 2018. [Citado el: 26 de 05 de 2022.] <http://help.pentaho.com/Documentation/6.0/Products>.

**Community, Penyaho. 2016**. Schema Workbench. [En línea] 2016. [Citado el: 1 de JULIO de 2022.] <http://community.pentaho.com/>.

**Corey, Michael. 1997**. *Oracle Data Warehousing: guía práctica para analizar, construir e implantar con éxito un sistema data warehouse*. Madrid : McGrawHill, 1997.

**Dario, Ing. Bernabeu Ricardo. 2010**. *HEFESTO: Metodología para la Construcción de un Data Warehouse*. Córdoba, Argentina : s.n., 2010.

**Dhawan B, Gosain. 2009**. *Extending Uml for Multidimensional Modeling in Data Warehouse*. s.l. : International Journal of Computer & Communication Technology, 2009.

**Fatima, Nida. 2019**. Astera Software. *¿Qué es un almacén de datos? Definición, ejemplo y beneficios* . [En línea] 3rd de junio de 2019. [Citado el: 1 de junio de 2022.] <https://www.astera.com/es/type/blog/data-warehouse-definition/>.

**Galeano, Luz. 2016.** 2016. Prezi <https://prezi.com/m/j84ywfzvit/visual-paradigm/>. . 2016.

**García Leopoldo, Ocharan Octavio. 2017.** *Diseño de un AD basado en DATA WAREHOUSE Engineering Process (DWEP) y HEFESTO.* 2017.

**Goire, Alejandro Diaz . Maria Castilla. 2019.** *Sistema para la gestión y análisis de datos de una red de sensores inalámbricos basado en un almacén de datos.* Universidad de Oriente ,Cuba : Revista Cubana de Ciencias Informáticas, 2019. 2227-1899.

**González, Yisel Pompa. 2014.** *MERCADOS DE DATOS PARA EL ANÁLISIS ESTADÍSTICO DE LA INFORMACIÓN.* 2014.

**Guide, Data. 2020.** El sistema de gestión de base de datos. [En línea] 16 de 03 de 2020. [Citado el: 3 de junio de 2022.] <https://www.kyoceradocumentsolutions.es/es/smarter-workspaces/businesschallenges/procesos/caracteristicas-avanzadas-de-un-sgbd.html>.

**Hernandez, Ing Yanisbel González. 2014.** *Metodología para el desarrollo de proyectos de almacenes de datos.* s.l. : <http://repositorio.uci.cu/jspui/handle/ident/8094.>, 2014.

**Herrera Eduard. 2010.** *Data Warehouse Engineering Process (DWEP) with U.M.L. 2.1.1.* Panama : Universidad Tecnológica de Panama., 2010.

<https://spa.myservername.com/schema-types-data-warehouse-modeling-star-snowflakeschema> . [En línea] [Citado el: 1 de Junio de 2022.] <https://spa.myservername.com/schema-types-data-warehouse-modeling-star-snowflakeschema> .

**Imon, William. 2002.** *W. Building the data warehouse.* New York : J. Wiley & Sons, 2002. 978-0764599446.

**Juan, Reyes. 2015.** [En línea] 2015. [Citado el: 20 de 06 de 2022.] <http://es.slideshare.net/juangerardomatareyes/data-warehouse-16842696>..

**Kimball R, Ross M. 2002.** *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling.* s.l. : Second ed: John Wiley & Sons, 2002.

**Kimball Ralph. 2013.** *The Data Warehouse Toolkit, The Definitive Guide to Dimensional Modeling, 3rd.* s.l. : Idianapolis : Jhon Wiley & Sons,, 2013. 978-1-118-53080-1.

**Kimball, Ralph. 2004.** *The Data Warehouse ETL Toolkit (en inglés).* New York : Wiley. p. 23., 2004. ISBN 0-76457923-1.

**Klein Christian. 2018.** Systemanalyse Programmentwicklung. [En línea] SAP, 2018. [Citado el: 24 de 07 de 2022.] <https://www.sap.com/latinamerica/insights/what-is-a-data-warehouse.html>.

lanzamiento-de-postgresql-14. [En línea] [Citado el: 3 de junio de 2022.] <https://systemguards.com.ec/postgresql/lanzamiento-de-postgresql-14/> .

**Leonard, Eric. 2013.** *Metodologías para desarrollar Almacén de Datos.* Universidad de Granma, Bayamo, Granma, Cuba, : s.n., 2013.

**López, Falcón A. L., & Ramos Serpa, G. 2021.** *ACERCA DE LOS MÉTODOS TEÓRICOS Y EMPÍRICOS DE INVESTIGACIÓN: SIGNIFICACIÓN PARA LA INVESTIGACIÓN EDUCATIVA.* 2021.

**Lujan Sergio. 2005.** *Data Warehouse Design with UML.* España : Universidad de Alicante, 2005.

**Luján Sergio, Trujillo J, Song IY. 2006.** *A UML profile for multidimensional modeling in data warehouses.* 2006.

**Martín, Paula. 2019.** Que-es-pentaho-data-integration-pdi-y-para-que-sirve. [En línea] 24 de 01 de 2019. [Citado el: 4 de junio de 2022.] <https://itop.academy/blog/item/que-es-pentaho-data-integration-pdi-y-para-que-sirve.html>.

**Martínez, Castro José Mario Morales Feliciano. 2016.** *Procesamiento Analítico con Minería de Datos.* Mexico : Revista Iberoamericana de las Ciencias Computacionales e Informática, 2016. 2007-9915.

**Microsoft. 1998.** The MS Data Warehousing Strategy : A platform for improved decision-making through easier data access and analysis. [En línea] Microsoft, 1998. [Citado el: 21 de 06 de 2022.] <http://www.microsoft.com/TechNet/sql/Technote/datawhst.asp>.

**Morales, Angelino, Feliciano, Valencia, René Edmundo Cuevas y Castro, José Mario Martínez. 2016.** Mining., *Procesamiento Analítico con Minería de Datos / Analytical Processing with Data.* Revista Iberoamericana de las Ciencias Computacionales e Informática, 2016, Vol. 5.

**Morfa, Beatriz Hidalgo . Yasmani Otero. 2015.** *Mercado de Datos para los departamentos de Procesos Penales y Gestión de Cuadros y Personal de Apoyo de la Fiscalía General de la República de Cuba.* La Habana, Cuba : Universidad de las Ciencias Informáticas, 2015.

**Morgan Tony. 2002.** *Business Rules and Information Systems: Aligning IT with Business Goals.* s.l. : Addison Wesley, 2002.

**Olivero Luis. 2014.** *Reglas de negocio: un nuevo enfoque de sistemas informáticos para la gestión dinámica empresarial.* Colombia : Investigación, Innovación, Ingeniería, : s.n., 2014.

**pgAdmin. 2020.** pgAdmin. [En línea] 2020. [Citado el: 21 de 06 de 2022.] Obtenido de <https://www.pgadmin.org/download/>.

**Pressman, Roger. 2010.** *Ingeniería del Software. Un enfoque práctico, 7ma edición.* s.l. : McGraw-Hill, 2010. ISBN: 9786071503145.

—. 2010.. *Ingeniería del Software. Un enfoque práctico, 7ma edición.* McGraw-Hill, : s.n., 2010. 9786071503145.

*Propuesta de herramientas para la integración de Datos.* **Oliva, Débora. 2012.** Cuba : Revista Cubana de Ingeniería, 2012, Vol. 3. ISSN:2223-1781.

**Ramos Salvador. 2011.** Ramos, Salvador. 2011. *Microsoft Business Intelligence: vea el cubo medio lleno.* Microsoft Business Intelligence: vea. España : Solid Quid, 2011.

**Ramos, Salvador. 2016..** *Data Warehouse, Datamart y Modelos Dimensionales. Un pilar fundamental para la toma de decisiones.* . Alicante: : SolidQ Press, 2016. ISBN:978-84-9407019-28.

**Ricardo, Chinchilla Arley. 2010.** *Mercado de datos: conceptos y metodologías de desarrollo.* 09 : s.n., 2010.

**Ross Ronald. 2010.** *What Is a Business Rule? Business Rules Journal, Vol. 11, No. 3.* 2010.

**Salazar, Daniel Arturo. 2018.** *Pruebas de Software.* 2018.

**Salazar, Ricardo Luján. 2009.** *Almacén de datos para la prestación del servicio público de información estadística.* . Mexico : s.n., 2009.

**Sofftek. 2021.** SOFFTEK. [En línea] IN DATA ECOSISTEM, 2 de september de 2021. [Citado el: 20 de junio de 2022.] <https://sofftek.eu/tech-magazine/data-ecosystem/las-nuevas-tendencias-de-las-bases-de-datos/>.

**Sommerville, Ian. 2011.** *Ingeniería de Software*. México : PEARSON EDUCACIÓN, 2011. 978-607-32-0603-7.

**Tactuk, Pablo. 2013..** Oficina Nacional de Estadística. Oficina Nacional de Estadística. . [En línea] 2013. [Citado el: 25 de junio de 2022.] <http://www.one.gob.do>.

**Trujillo Juan, Soler E, Zubcoff J, Mazón JN, Glorio O, Pardillo J. 2007.** *Desarrollo de almacenes de datos dirigido por modelos*. Universidad de Alicante : s.n., 2007.

Trujillo Mondejar Juan Carlos. [En línea] [Citado el: 2 de junio de 2022.] <https://cvnet.cpd.ua.es/curriculum-breve/es/trujillo-mondejar-juan-carlos/6151> .

Universidad de las Ciencias Informáticas. [En línea] [Citado el: 22 de 05 de 2022.] <https://www.uci.cu/universidad/mision>.

**Universidad de las Ciencias Informáticas. 2018.** UCI. *Universidad de las Ciencias Informáticas*. [En línea] 2018. [Citado el: 12 de 7 de 2022.] <https://www.uci.cu/universidad/mision>.

**Victor, Parra. 2016.** *International Journal of Advanced Computer Science and Applications*. Vol. 7, 10. 2016.

**Wesley Addison, W. 1994.** *El modelo entidad relacion Case Method*. Madrid, España : Ediciones Diaz Santos S.A, 1994. 0-201-60111-7.

**Wikimedia, Proyectos. 2022.** Wikipedia : la Enciclopedia Libre. *Almacén de datos*. [En línea] 25 de enero de 2022. [Citado el: 1 de junio de 2022.] [https://es.wikipedia.org/wiki/Almac%C3%A9n\\_de\\_datos](https://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos).

**Wolff, Carmen. 2011.** Implementando un Data Warehouse. *Revista Ingeniería Informática: revista electrónica del DIICC*, Edición 5, año 3, 2011.

## Anexos