

Universidad de las Ciencias Informáticas

Facultad 6



Título: Mercado de datos Civil-Administrativo del Sistema de Información Estadística Judicial para los Tribunales Populares.

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

Autores: Anneris Eloisa Wong Jorge

Yasmany Pino Mesa

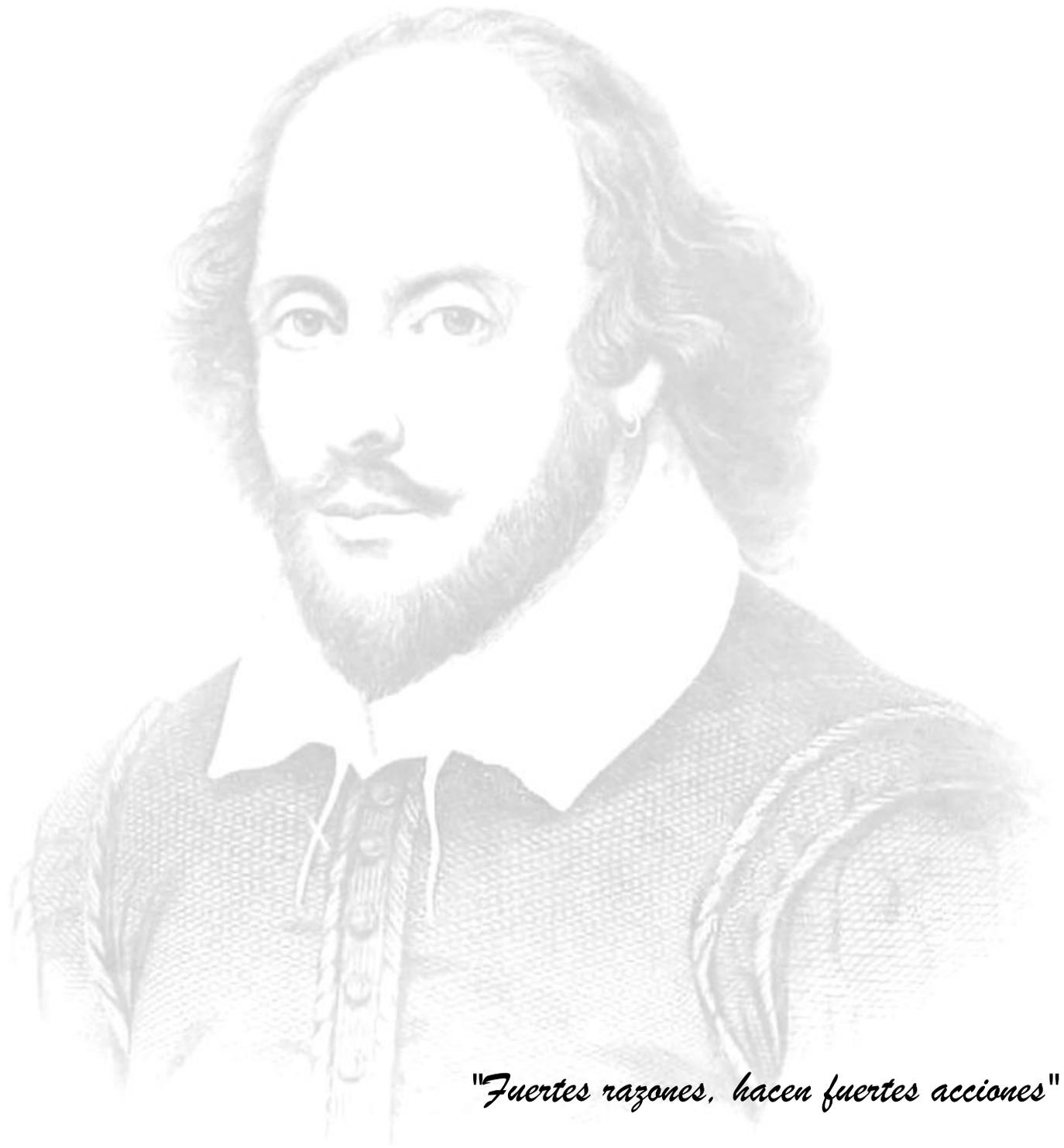
Tutoras: Ing. Themis Patricia Díaz Morales

Ing. Irina Collazo Ávila

La Habana, Cuba

Junio 2013

“Año 55 de la Revolución”



"Fuentes razones, hacen fuertes acciones"

William Shakespeare

Declaración de autoría

Declaramos ser los autores del presente Trabajo de Diploma “Mercado de datos Civil-Administrativo del Sistema de Información Estadística Judicial para los Tribunales Populares” y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales del mismo, con carácter exclusivo.

Para que así conste se firma la presente a los ____ días del mes de _____ del año 2013.

Anneris E. Wong Jorge

Firma del Autor

Yasmany Pino Mesa

Firma del Autor

Ing. Themis Patricia Díaz Morales

Firma del Tutor

Ing. Irina Collazo Ávila

Firma del Tutor

Tutoras:

Ing. Themis Patricia Díaz Morales

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Categoría docente: Instructora

Años de experiencia en el tema: 4

Años de graduado: 3

Correo Electrónico: tpdiaz@uci.cu

Ing. Irina Collazo Ávila

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Categoría docente: Instructora

Años de experiencia en el tema: 3

Años de graduado: 5

Correo Electrónico: icollazo@uci.cu

Dedico este trabajo de diploma:

A mis padres por hacer de mi la persona que soy hoy en día y educarme con todo el amor y el cariño del mundo.

A mi hermana por ser un ejemplo a seguir.

A Alejandro Alfonso Villegas y Sadiel Rodríguez por ser mis mejores amigos y ayudarme en todos los momentos que los necesité.

A Rafael Guevara por confiar en mí y apoyarme en mis primeros años de la carrera.

A Lizy Vidal por ser mi otra hermana y estar a mi lado en los momentos buenos y malos.

A San Lázaro y la Virgencita por darme la fuerza que necesité para llegar al final de la carrera.

Agradezco:

A mis tutoras Themis e Irina porque tuvieron la paciencia y dedicación para guiarme hacia el camino correcto.

A Lianet, Yoanna, Rebeca, Irela y Anneris por ayudarme en el transcurso de los 5 años de la carrera.

Yasmany

Dedico este trabajo de diploma:

A toda mi familia, en especial a mis padres por mostrarme su apoyo y amor incondicional.

A mi esposo por haber tenido tanta paciencia conmigo y por estar siempre a mi lado en los momentos más difíciles durante este último año.

A mi abuela Mimita, que aunque no esté con nosotros fue parte de mi crecimiento como persona y como ingeniera.

A Idalia, René, Rubén, Idalita y a mi sobrinita Kuki.

Agradezco:

A mis tutoras Themis e Irina por guiarme durante esta última etapa en la universidad.

A mis amistades: Lianet, Yoanna, Rebeca, Irela, Maikel y Yasmanny por apoyarme y haber estado presente durante los cinco años de la carrera.

A los profesores que me dedicaron un tiempito cuando tuve alguna duda.

Anneris

Resumen

La presente investigación surge por la necesidad de gestionar la información del departamento de Estadística del Tribunal Supremo Popular, en el área Civil-Administrativo, donde se recopilan los datos relacionados con los procesos de divorcios y alimentos, asuntos de familia, asuntos civiles, sentencias civiles, sentencias administrativas y apelaciones. Por tal motivo en la presente investigación se desarrolló el mercado de datos Civil-Administrativo, donde la información se encuentra centralizada, estandarizada y es de fácil acceso para su consulta. Para su desarrollo fue necesario fundamentar la metodología y herramientas a utilizar. Se realizó el análisis, diseño e implementación de los subsistemas de almacenamiento, integración y visualización, así como las pruebas a la aplicación. Finalmente se le proporcionó al departamento de Estadística un mercado de datos que almacena la información del área Civil-Administrativo, el cual permite hacer análisis estadísticos históricos de los principales indicadores de esta área, facilitando la toma de decisiones a los directivos del Tribunal Supremo Popular.

Índice

| | |
|--|----|
| Resumen | V |
| Introducción | 1 |
| Capítulo1. Fundamentos teóricos para el desarrollo de un mercado de datos | 5 |
| 1.1 Sistemas de apoyo a la toma de decisiones | 5 |
| 1.2 Almacenes de datos | 6 |
| 1.3 Mercado de datos | 7 |
| 1.4 Metodologías para el desarrollo de los almacenes de datos..... | 7 |
| 1.4.1 Ciclo de vida de Kimball | 7 |
| 1.4.2 Propuesta de metodología de desarrollo de almacenes de datos en DATEC..... | 8 |
| 1.5 Herramientas | 10 |
| 1.5.1 Herramienta de modelado | 11 |
| 1.5.2 Sistema gestor de base de datos..... | 11 |
| Modos de almacenamiento | 12 |
| 1.5.4 Herramienta para el perfilado y limpieza de los datos | 14 |
| 1.5.5 Herramientas para los procesos de extracción, transformación y carga..... | 14 |
| 1.5.6 Herramientas para la inteligencia de negocio..... | 15 |
| Conclusiones del capítulo..... | 17 |
| Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo | 18 |
| 2.1 Necesidades de información..... | 18 |
| 2.2.1 Requisitos de información | 19 |
| 2.2.2 Requisitos Funcionales | 20 |
| 2.2.3 Requisitos No Funcionales..... | 20 |
| 2.2 Reglas del negocio | 21 |
| 2.3 Diagrama de Casos de uso del sistema | 22 |
| 2.3.1 Descripción de los actores | 23 |
| 2.4 Casos de uso de información | 24 |
| 2.5 Casos de uso funcionales..... | 27 |
| 2.6 Arquitectura | 29 |

| | |
|---|-----------|
| 2.7 Diseño del subsistema de almacenamiento..... | 30 |
| 2.7.1 Dimensiones..... | 30 |
| 2.7.2 Hechos..... | 30 |
| 2.7.3 Medidas | 31 |
| 2.7.4 Matriz buz | 31 |
| 2.7.5 Topología de los almacenes de datos..... | 32 |
| 2.7.6 Modelo de datos | 35 |
| 2.8 Diseño del subsistema de integración | 36 |
| 2.8.1 Diseño de los procesos de integración..... | 38 |
| 2.9 Diseño del subsistema de visualización..... | 38 |
| 2.10 Esquema de seguridad..... | 41 |
| 2.10.1 Políticas de respaldo y recuperación..... | 42 |
| Conclusiones del capítulo..... | 42 |
| Capítulo 3: Implementación y prueba del MD Civil-Administrativo..... | 43 |
| 3.1 Implementación del subsistema de almacenamiento..... | 43 |
| 3.1.1 Implementación del modelo de datos físico..... | 43 |
| 3.1.2 Estándares de codificación | 44 |
| 3.2 Implementación del subsistema de integración | 44 |
| 3.2.1 Implementación de las transformaciones | 46 |
| 3.2.2 Implementación de los trabajos..... | 48 |
| 3.3 Implementación del subsistema de visualización..... | 50 |
| 3.3.1 Implementación de los cubos OLAP | 51 |
| 3.3.2 Arquitectura de la información..... | 52 |
| 3.4 Implementación de la seguridad de los usuarios | 53 |
| 3.5 Pruebas aplicadas al MD Civil-Administrativo..... | 53 |
| 3.6 Pruebas de calidad de software..... | 54 |
| 3.7 Herramientas de pruebas | 54 |
| 3.8 Resultados de las pruebas | 56 |
| Conclusiones del capítulo | 58 |

| | |
|---|----|
| Conclusiones Generales | 59 |
| Recomendaciones | 60 |
| Referencias bibliográficas | 61 |
| Anexos | 65 |
| Glosario de términos | 67 |

Índice de figuras

| | |
|---|--------------------------------------|
| Figura 1: Ciclo de vida dimensional del negocio..... | 8 |
| Figura 2: Ciclo de vida de la Propuesta de metodología de desarrollo de almacenes de datos en DATEC.. | 9 |
| Figura 3: Diagrama de casos de uso del sistema..... | 23 |
| Figura 4: Arquitectura de información del MD Civil-Administrativo. | 29 |
| Figura 5: Representación del esquema en estrella. | 33 |
| Figura 6: Representación del esquema copo de nieve..... | 34 |
| Figura 7: Representación del esquema constelación de hechos..... | 35 |
| Figura 8: Modelo de datos dimensional. Representa una topología de constelación de hechos. | 36 |
| Figura 9: Distribución en por ciento de los campos nulos y no nulos..... | 37 |
| Figura 10: Distribución de los tipos de datos..... | ¡Error! Marcador no definido. |
| Figura 11:Diseño general de las transformaciones para la carga del hecho asuntos de familia. | 38 |
| Figura 12: Mapa de navegación del MD Civil-Administrativo..... | 39 |
| Figura 13: Representación de la carga del hecho asuntos de familia provincial. | 46 |
| Figura 14: Representación de la carga del hecho asuntos de familia provincial. | 47 |
| Figura 15: Trabajo que permite la ejecución de las transformaciones para la carga de las dimensiones. .. | 48 |
| Figura 16: Trabajo que permite la ejecución de las transformaciones para la carga de los hechos..... | 48 |
| Figura 17: Representación de los hechos del MD Civil-Administrativo. Relación del hecho indicadores asuntos de familia provincial con sus dimensiones y su medida. | 51 |
| Figura 18: Arquitectura de información del MD Civil-Administrativo. | 52 |
| Figura 19: Representación de una vista de análisis del resumen nacional por provincias de los asuntos de familia provinciales..... | 53 |

Índice de tablas

| | |
|---|----|
| Tabla 1: Descripción de los casos de uso de información. | 24 |
| Tabla 2: Especificación de CUI Presentar información de indicadores de asuntos de familia..... | 25 |
| Tabla 3: Especificación de CUF autenticar usuario. | 28 |
| Tabla 4: Matriz bus. | 32 |
| Tabla 5: Comportamiento de los asuntos de familia en los TPP..... | 41 |
| Tabla 6: Relación de las tablas por esquemas del MD Civil-Administrativo..... | 43 |
| Tabla 7: Estándares de codificación del mercado de datos Civil-Administrativo..... | 44 |
| Tabla 8: Resultado general de las pruebas aplicadas al MD Civil-Administrativo según su complejidad.... | 58 |

Introducción

El vertiginoso avance de las Tecnologías de la Información y las Comunicaciones (TICs) ha permitido el incremento acelerado de los resultados alcanzados en la producción, los servicios y la investigación. Su incidencia en las diferentes áreas del conocimiento conlleva al efectivo almacenamiento, tratamiento, registro y presentación de la información en forma de voz, imágenes y datos. Es por ello que a nivel mundial la mayoría de las empresas han automatizado sus procesos, en una época en que la información es la clave para obtener una ventaja competitiva en el mundo de los negocios, buscando soluciones tecnológicas que brinden mayor rapidez en el manejo de sus datos.

El esfuerzo que se ha realizado en Cuba por experimentar estos avances tecnológicos es incuestionable, siendo la informática un factor primario para su desarrollo. En aras de lograr la informatización del país y en medio de un sin número de sucesos, surge la Universidad de las Ciencias Informáticas (UCI), centro que vincula la docencia y la producción para formar profesionales capaces de fortalecer la industria del software en Cuba. Para lograr una mejor organización en la producción de dicha universidad, se han creado diferentes centros de desarrollo, entre los que se encuentra el Centro de Tecnologías de Gestión de Datos (DATEC), que tiene como misión proveer soluciones integrales, productos y servicios relacionados con las tecnologías de gestión de datos, en el cual el departamento Almacenes de Datos se encuentra trabajando en conjunto con el Tribunal Supremo Popular.

El Tribunal Supremo Popular es la institución encargada de ejercer la máxima autoridad judicial en la República de Cuba. A través de su Consejo de Gobierno ejerce la iniciativa legislativa en materia concerniente a la administración de justicia y la potestad reglamentaria. Dentro de los departamentos que constituyen esta entidad se encuentra el de Estadística, cuya finalidad consiste en presentar cifras estadísticas e indicadores que reflejen el comportamiento y desarrollo de la actividad jurídica del país. Para un mejor control y organización, este departamento divide la información que se trata en diferentes áreas, una de ellas es la Civil-Administrativo.

Toda la información referente a esta área, es recopilada mediante una serie de modelos estadísticos que provienen de las diferentes provincias y municipios, los cuales son recogidos mensualmente y almacenados en formatos "dbf", de difícil acceso para su consulta. Algunos de estos ficheros están codificados de forma que si no se es conocedor de la terminología utilizada pueden no comprenderse correctamente. Posteriormente dichos modelos son revisados de forma manual por los diferentes especialistas del área, lo cual provoca un engorroso trabajo a la hora de realizar diferentes análisis, ya que

se interactúa con un gran volumen de datos que se han ido acumulando con el transcurso de los años; lo que trae consigo el riesgo de perder información útil y valiosa dificultando así el proceso de toma de decisiones.

Por la situación anteriormente descrita, se plantea como **problema de la investigación**: ¿Cómo contribuir a la toma de decisiones en el área Civil-Administrativo del Sistema de Información Estadística Judicial para los Tribunales Populares?

La presente investigación tiene como **objeto de estudio** los almacenes de datos, enmarcado en el **campo de acción**: mercado de datos Civil-Administrativo para el Sistema de Información Estadística Judicial para los Tribunales Populares.

Para darle solución al problema de la investigación se define como **objetivo general** desarrollar el mercado de datos Civil-Administrativo para el Sistema de Información Estadística Judicial para los Tribunales Populares, que contribuya a la toma de decisiones.

En correspondencia con el objetivo general, se plantean como **objetivos específicos**:

- ✓ Fundamentar la selección de la metodología y herramientas a utilizar en el desarrollo de los almacenes de datos.
- ✓ Realizar el análisis y diseño del mercado de datos Civil-Administrativo del Sistema de Información Estadística Judicial para los Tribunales Populares.
- ✓ Realizar la implementación y prueba del mercado de datos Civil-Administrativo del Sistema de Información Estadística Judicial para los Tribunales Populares.

Para alcanzar los objetivos propuestos se plantean las siguientes **tareas de la investigación**:

- ✓ Análisis de la metodología y herramientas a utilizar en el desarrollo de almacenes de datos.
- ✓ Levantamiento de requisitos para definir las necesidades del cliente.
- ✓ Descripción de los casos de uso del mercado de datos para determinar cada una de las funcionalidades del sistema.
- ✓ Definición de la arquitectura del mercado de datos identificando los subsistemas fundamentales que componen la solución.

- ✓ Definición de los hechos, las medidas y las dimensiones del mercado de datos para identificar los elementos que forman parte del modelo lógico de datos.
- ✓ Diseño del modelo de datos para determinar los elementos que conforman el modelo físico.
- ✓ Diseño del subsistema de integración para definir cómo se realizará la carga de las dimensiones y los hechos al mercado de datos.
- ✓ Diseño del subsistema de visualización definiendo la estructura de la capa de presentación y realizando el diseño de los cubos OLAP.
- ✓ Diseño de los casos de prueba para describir los elementos que tienen que estar disponibles en el almacén una vez culminada la implementación y que servirán como apoyo en la realización de las pruebas.
- ✓ Implementación del modelo de datos para garantizar la disponibilidad de las estructuras de la base de datos.
- ✓ Implementación del subsistema de integración para poblar el mercado de datos.
- ✓ Implementación del subsistema de visualización para lograr que se muestre la información correspondiente a los requerimientos planteados por el cliente.
- ✓ Aplicación de las listas de chequeo para garantizar el correcto funcionamiento del sistema.
- ✓ Aplicación de los casos de prueba para avalar la disponibilidad de cada uno de los elementos del mercado de datos.

Estructura del trabajo de diploma

El trabajo de diploma está estructurado de la siguiente manera: resumen, introducción, tres capítulos, conclusiones, recomendaciones, glosario de términos, referencias bibliográficas, bibliografía y anexos.

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

Está referido al análisis del estado del arte de los almacenes de datos y de los mercados de datos, con sus principales características, metas y elementos que los componen. Se define cómo se realiza el proceso de gestión de la información correspondiente al área Civil-Administrativo y se caracteriza la metodología y herramientas a utilizar para el desarrollo de la solución.

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo del Sistema de Información Estadística Judicial para los Tribunales Populares.

En este capítulo se realiza un estudio preliminar del negocio con el fin de obtener las reglas del negocio, identificar los requisitos y los casos de uso con sus relaciones, describiendo brevemente los actores que van a interactuar con el sistema. Se define la arquitectura del mercado de datos y se diseñan los subsistemas de almacenamiento, integración y visualización. Además se establece el esquema de seguridad para la interacción con la base de datos y la aplicación.

Capítulo 3: Implementación y prueba del mercado de datos Civil-Administrativo del Sistema de Información Estadística Judicial para los Tribunales Populares.

En este capítulo se hace referencia a la implementación de la solución, abordando específicamente cómo se realiza la implementación de los subsistemas de almacenamiento, integración y visualización, teniendo en cuenta los requisitos y necesidades del negocio. De igual forma se hace referencia a las pruebas, mediante la utilización de las listas de chequeo y de los casos de pruebas realizados al mercado de datos.

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

En este capítulo se aborda el estudio de los sistemas de apoyo a la toma de decisiones entre los que se encuentran los almacenes de datos (AD) y los mercados de datos (MD). Se abordan las definiciones y conceptos más importantes referidas a estos, además de sus principales características. Se realiza un estudio de la metodología y las principales herramientas para el desarrollo de la investigación.

1.1 Sistemas de apoyo a la toma de decisiones

Los sistemas de apoyo a la toma de decisiones han evolucionado conforme lo han hecho las empresas. En el mundo existen diversos sistemas que permiten el apoyo a la toma de decisiones, entre los que se encuentran:

- ✓ Sistemas de información gerencial (MIS¹): contiene la información financiera de la empresa, organizada y programada de tal manera que produce informes periódicos sobre las operaciones para todos los niveles de gestión de la misma (1).
- ✓ Sistemas de información ejecutiva (EIS²): proveen a los gerentes un acceso sencillo a información interna y externa de su compañía, la cual es relevante para sus factores claves de éxito (2).
- ✓ Sistemas de expertos (SE³): utilizan redes neuronales para simular el conocimiento de un experto y utilizarlo de forma efectiva para resolver un problema concreto (2).
- ✓ Sistemas de apoyo a decisiones de grupo (GDSS⁴): es un sistema desarrollado para apoyar a grupos de personas que tienen una tarea u objetivo común (2).
- ✓ Sistema de soporte a la toma de decisiones (DSS⁵): es un conjunto de programas y herramientas que permiten obtener oportunamente la información requerida durante el proceso de toma de decisiones. Dentro de los elementos que componen los DSS se encuentran los almacenes de datos, mercados de datos, herramientas de extracción de datos y herramientas de procesamiento analítico en línea (OLAP⁶) (3).

¹ MIS: Management Information Systems

² EIS: Executive Information System

³ SE: Expert Systems

⁴ GDSS: Group Decision Support Systems

⁵ DSS: Decision Support System

⁶ OLAP: On-Line Analytical Processing

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

Después de analizar los diferentes sistemas de apoyo a la toma de decisiones, en la presente investigación se decide desarrollar un mercado de datos como DSS, teniendo en cuenta que es el que más se adapta a las condiciones del centro DATEC y debido a que presenta las siguientes características (2):

- ✓ Disponibilidad de información histórica.
- ✓ Cada usuario dispone de la información específica de su perfil.
- ✓ Integración entre todos los sistemas/departamentos de la compañía.

1.2 Almacenes de datos

Un almacén de datos es una colección de datos integrados que recoge información de múltiples sistemas fuentes u operacionales dispersos, que centra su actividad en el análisis de la información en lugar de su captura y guarda los datos reunidos, permitiendo el acceso histórico a los mismos. Proporciona una interfaz de usuario consolidada facilitando la realización de consultas para la toma de decisiones (4).

Existen dos grandes autores con respecto al tema de los AD: Bill Inmon y Ralph Kimball.

Bill Inmon plantea que un AD es una colección de datos orientado a temas, integrado, variable en el tiempo y no volátil para ayudar al proceso de toma de decisiones gerenciales (5)."

Mientras que Ralph Kimball en su libro "*The Data Warehouse ETL Toolkit*" define un almacén de datos como un sistema que extrae, limpia, ajusta e integra los datos del origen en un almacenamiento de datos dimensionales y luego apoya e implementa la consulta y el análisis con el fin de la toma de decisiones.

Las principales características de un AD son (6):

Orientados a temas: en los AD los datos se estructuran por temas de interés para facilitar su análisis por parte de los usuarios.

Integrados: en la construcción de los AD, los datos deben ser integrados en una única BD, este proceso de integración implica costosas tareas de limpieza y transformación de los datos.

Variables en el tiempo: el tiempo adquiere un valor importante, cuando se analizan los datos es importante conocer "la variación de los datos en el tiempo", por este motivo en un AD los datos deben estar ligados siempre a un instante específico de tiempo o a un intervalo.

No volátil: los AD existen para ser consultados y no para ser modificados. La información es por tanto permanente y la actualización del AD consiste exclusivamente en la incorporación de datos correspondientes al último período de tiempo.

1.3 Mercado de datos

Un MD es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer una estructura de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento (7).

Los MD se caracterizan por (8):

- ✓ Proporcionar un esquema de almacenamiento que permita realizar consultas complejas, manteniendo solo un repositorio con la información agregada y ordenada.
- ✓ Permitir al usuario la posibilidad de análisis de la información para la toma de decisiones.
- ✓ Solucionar problemas derivados de la dispersión de los datos.

1.4 Metodologías para el desarrollo de los almacenes de datos

Una metodología se refiere al entorno que se usa para estructurar, planificar y controlar un proceso determinado. En el caso del desarrollo del software, su definición no se aleja de este marco, a esto se le añade la definición de las etapas, los roles y las tareas para la generación de los artefactos finales de productos informáticos (9).

Para la elaboración y desarrollo de un AD, existen diferentes metodologías que se enmarcan en dos enfoques o vertientes principales. La propuesta de Ralph Kimball (principal promotor del enfoque dimensional para el diseño de AD) y la de Bill H. Inmon (conocido por muchos como el padre del AD). Ambos son las personalidades más influyentes en el área de los AD. La visión de Inmon se basa principalmente en un enfoque descendente (del inglés top-down), donde plantea la creación de un repositorio de datos corporativo como fuente de información consolidada, persistente, histórica y de calidad. Al ser construido descendentemente, el MD se nutre del AD corporativo convirtiéndose en un complejo empresarial de BD relacional (10).

La propuesta de Ralph Kimball se basa principalmente en un enfoque ascendente (del inglés bottom-up), debido a que plantea que se debe crear por departamentos un conjunto de MD independientes, orientados a los temas que estén relacionados con él (10).

1.4.1 Ciclo de vida de Kimball

Kimball propone una metodología con su Ciclo de vida dimensional del negocio el cual comienza con la planificación del proyecto, en donde se define el alcance, se identifican y programan las tareas, se planifica el uso de los recursos, conformando todo esto el plan de proyecto. En la segunda etapa se

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

definen los requisitos del negocio, para luego enfocar el proyecto en tres líneas concurrentes: tecnología, datos y aplicaciones de la inteligencia de negocios (ver figura 1).

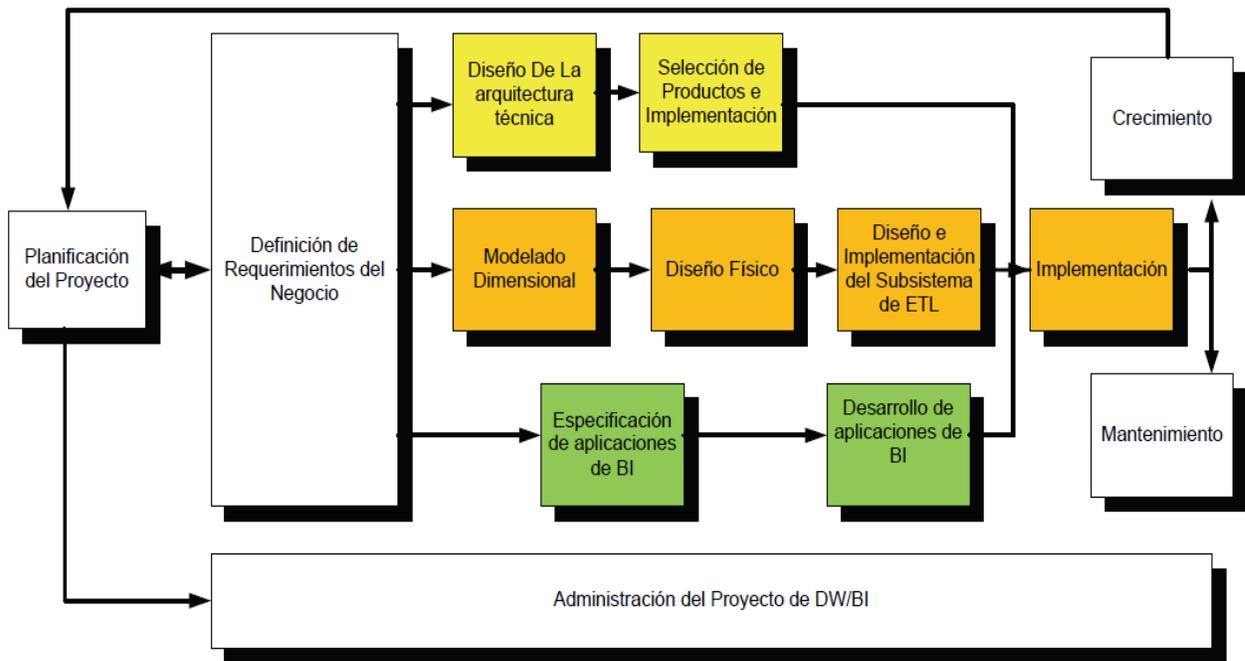


Figura 1: Ciclo de vida dimensional del negocio.

1.4.2 Propuesta de metodología de desarrollo de almacenes de datos en DATEC

Tomando como base el ciclo de vida de Kimball, DATEC propone una metodología que abarca todas las fases del desarrollo de un AD, la misma se nombra: Propuesta de metodología de desarrollo de almacenes de datos en DATEC.

Para complementar la metodología y fortalecer la etapa de levantamiento de requisitos se toma lo planteado por Leopoldo Zenaido Zepeda Sánchez en su tesis de doctorado, incluyendo de esta forma los casos de uso para guiar el proceso de desarrollo. En su concepción también se añade el Programa de Mejora llevado a cabo en la UCI incorporándose una etapa de prueba que fortalece en gran medida la calidad de la solución a desplegar. A continuación se muestra el ciclo de vida de la Propuesta de metodología de desarrollo de almacenes de datos en DATEC (Ver figura 2).

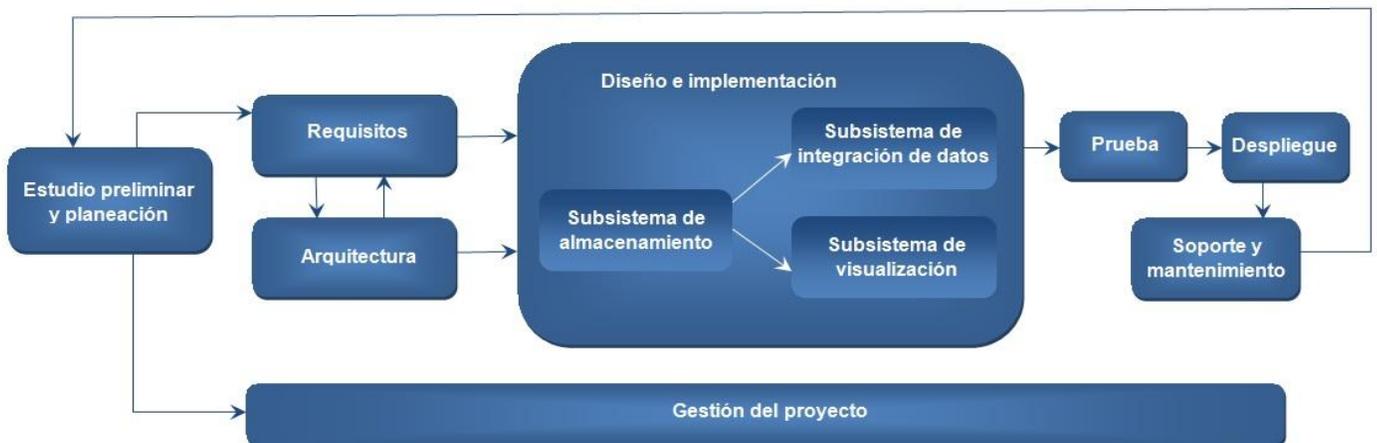


Figura 2: Ciclo de vida de la Propuesta de metodología de desarrollo de almacenes de datos en DATEC.

- **Estudio preliminar y planeación:** se realiza un estudio minucioso en la entidad cliente. Esto incluye un diagnóstico integral de la organización con el fin de determinar qué es lo que se desea construir y qué condiciones existen para el desarrollo y montaje de la misma. Además se llevan a cabo las tareas de planeación del proyecto (10).
- **Requisitos:** se realiza el proceso de entrevistas al cliente para determinar los requisitos de información. Se realiza un levantamiento detallado de las fuentes de datos para validar la disponibilidad de la información. Además se definen los requisitos funcionales y no funcionales de la solución y se hace el análisis de los requisitos que dan paso al diseño e implementación (10).
- **Arquitectura:** se definen las vistas arquitectónicas de la solución, aspectos como: los subsistemas y componentes, la seguridad, la comunicación y la tecnología a utilizar (10).
- **Diseño e implementación:** se define el diseño de las estructuras de almacenamiento de datos, se diseñan los procesos de integración de datos como el mapa lógico de datos, los cubos OLAP para la presentación de la información, así como el diseño gráfico de la aplicación definido por el cliente. Después se implementan cada uno de los subsistemas (repositorio de datos, integración de datos, presentación de datos) (10).
- **Prueba:** se realizan las pruebas que validan la calidad del producto, comenzando por las pruebas de unidad, las pruebas de integración y de sistema, hasta llegar a las pruebas de aceptación del cliente. Esta fase no es la única en la que se realizan pruebas, durante el desarrollo del proyecto en todas las fases hay actividades de aseguramiento de la calidad (10).

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

- **Despliegue:** Se realiza el despliegue piloto, donde se configuran los servidores necesarios y se instalan las herramientas según la arquitectura definida, se carga una muestra de los datos en un ambiente controlado, con el fin de mostrarle al cliente final el sistema en funcionamiento. Una vez aceptada la solución por el cliente, se realiza la carga histórica de los datos, puede ser en el mismo entorno que el despliegue piloto u otro, todo depende de las condiciones que establezca el cliente. Además se realiza la capacitación y transferencia tecnológica de la solución a los clientes. El resultado fundamental es la solución desplegada en el entorno real y en correcto funcionamiento (10).
- **Soporte y mantenimiento:** comienza cuando la solución está implantada y en explotación, y se ejecuta según el contrato firmado y las condiciones de soporte establecidas. Puede realizarse a través de variados servicios, que pueden ser soporte en línea, vía telefónica, web, correo u otros y el acompañamiento al cliente. Además se realizan las tareas de manteniendo de la aplicación tan necesarias para este tipo de desarrollo y que garantiza el adecuado funcionamiento y crecimiento del AD (10).
- **Gestión del proyecto:** esta fase se ejecuta a lo largo de todo el ciclo de vida del proyecto. Es aquí donde se controla, gestiona y chequea el desarrollo, los gastos, las utilidades, los recursos, las adquisiciones, los planes y cronogramas entre otras actividades relacionadas con la gestión de proyectos. Esta fase es el pilar base del proyecto y si no se ejecuta de forma continua y correcta el proyecto puede fracasar (10).

En la presente investigación se decide utilizar como metodología a seguir, la Propuesta de metodología de desarrollo de almacenes de datos en DATEC debido a las características antes mencionadas. Se debe señalar que solo se utilizarán las cinco primeras fases en el desarrollo de la solución, debido a que las restantes fases serán llevadas a cabo por los especialistas del Centro.

1.5 Herramientas

Existen disímiles herramientas para la automatización del diseño, construcción, implementación y mantenimiento de los AD. En los siguientes epígrafes se muestran las características de las herramientas utilizadas para la realización del MD Civil-Administrativo, las cuales fueron definidas en su mayoría por los arquitectos del proyecto.

1.5.1 Herramienta de modelado

Existen en el mundo diferentes herramientas de modelado para el desarrollo de un software, las cuales se utilizan para agilizar y facilitar el trabajo, debido a que proveen de métodos, técnicas y utilidades que ayudan al perfeccionamiento y desarrollo del MD. Como herramienta para el modelado de la solución se empleará Visual Paradigm para UML 8.0 por las cualidades que a continuación se presentan.

Visual Paradigm para UML 8.0

Visual Paradigm para UML⁷ es una herramienta CASE⁸ 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 un software. Esta herramienta tiene características gráficas muy cómodas que facilitan la realización de los diagramas de modelado que siguen el estándar de UML. Tiene la capacidad de crear el esquema de clases a partir de una BD y crear la definición de BD a partir del esquema de clases (11).

Otras de sus características son (12):

- Entorno de creación de diagramas para UML 2.0.
- Diseño centrado en casos de uso y enfocado al negocio que genera un software de mayor calidad.
- Uso de un lenguaje estándar común a todo el equipo de desarrollo, facilitando la comunicación.
- Disponibilidad de integración con los principales IDEs⁹ de desarrollo.
- Multiplataforma.
- Compatibilidad entre ediciones.
- Exportación de imágenes en formato jpg, png y svg.

1.5.2 Sistema gestor de base de datos

Un sistema gestor de base de datos (SGBD) es una colección de programas que permite a los usuarios crear y mantener una BD, con el propósito de facilitar los procesos de definición, construcción y manipulación de la BD para distintas aplicaciones (13).

⁷UML: Unified Modeling Language (Lenguaje Unificado de Modelado).

⁸CASE: Computer Aided Software Engineering (Ingeniería de Software Asistida por Computadora).

⁹IDE: Integrated Development Environment (Entorno de Desarrollo Integrado).

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

Entre los gestores de base de datos más utilizados se encuentra PostgreSQL el cual se utilizará en la solución propuesta por las características que a continuación se describen.

PostgreSQL 9.1

Constituye el SGBD objeto-relacional de código abierto más desarrollado del mundo y opera bajo licencia BSD. Alcanza un conjunto de características tales como: herencia, tipos de datos, funciones, restricciones, disparadores, reglas e integridad transaccional (14) .

Algunas de las ventajas que presenta son (15):

- Soporta distintos tipos de datos.
- Multiplataforma.
- Documentación muy bien organizada.
- Incluye herencia entre tablas.

pgAdmin III

Como interfaz de administración de PostgreSQL 9.1, se utiliza la herramienta **pgAdmin III** en su versión 1.14.0. La misma es una herramienta de software libre, que accede a casi todos los objetos del PostgreSQL y responde a las necesidades de todos los usuarios, es la más completa y popular con licencia de código abierto (14).

El pgAdmin III está escrito en C++ usando la librería gráfica multiplataforma, lo que permite que se pueda utilizar en Linux, Mac OS y Windows. Está diseñado para responder a las necesidades de todos los usuarios, desde escribir consultas SQL simples hasta desarrollar BD complejas. La aplicación también incluye un editor SQL con resaltado de sintaxis, un editor de código de la parte del servidor y un agente para lanzar scripts programados (14).

Modos de almacenamiento

Los MD almacenan información para su posterior búsqueda y análisis, es por ello que se requiere de la tecnología OLAP (Procesamiento analítico en línea) la cual se define como una tecnología que se basa en el análisis multidimensional de los datos y que le permite al usuario tener una visión más rápida e interactiva de los mismos (16).

En general, los sistemas OLAP deben (17):

- ✓ Soportar requisitos complejos de análisis.

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

- ✓ Analizar datos desde diferentes perspectivas.
- ✓ Soportar análisis complejos contra un volumen de datos.

Existen tres clasificaciones de OLAP:

- **MOLAP** (Procesamiento analítico en línea multidimensional), donde los datos se almacenan de forma lógica en arreglos matriciales con estructura multidimensional para optimizar los tiempos de respuesta (18).

Las ventajas que permite **MOLAP** son (19):

- ✓ Mayor rapidez en el procesamiento de consultas.
- ✓ Poco tiempo de cálculo realizado en el momento.
- ✓ Puede escribir sobre la BD.
- ✓ Posibilita hacer cálculos más complicados.

Entre sus desventajas se encuentran (19):

- ✓ Tamaño limitado para la arquitectura del cubo.
- ✓ No puede acceder a datos que no están en el cubo.

- **ROLAP** (Procesamiento analítico en línea relacional), es una implementación OLAP que almacena los datos en un motor relacional. Típicamente, los datos son detallados evitando las agregaciones y las tablas se encuentran normalizadas. Usa BD relacionales para el manejo, acceso y obtención de los datos (18).

Las ventajas que permite **ROLAP** son (19):

- ✓ Permite el análisis de una enorme cantidad de datos.
- ✓ Los datos pueden ser compartidos con aplicaciones SQL.
- ✓ Datos y estructura más dinámicos.

Entre sus desventajas se encuentran (19):

- ✓ Las consultas son más lentas.
- ✓ Los cálculos están limitados a las funciones de las BD.

- **HOLAP** (Procesamiento analítico en línea híbrido), estos sistemas mantienen los registros detallados en la BD relacional, mientras que los datos resumidos o agregados se almacenan en una BD multidimensional separada. Son conocidos como híbridos debido a que contienen las características de los sistemas anteriores intentando rescatar lo mejor de cada uno (20).

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

Se selecciona el modo de almacenamiento ROLAP, debido a que el SGBD PostgreSQL no soporta el almacenamiento multidimensional, solo el relacional. PostgreSQL es un SGBD de código abierto y multiplataforma, a diferencia de los SGBD que dan soporte al almacenamiento multidimensional existentes en la actualidad.

1.5.4 Herramienta para el perfilado y limpieza de los datos

De las herramientas de código abierto que brindan soporte para la gestión de calidad de datos existentes en el mercado se decidió utilizar el **DataCleaner** ya que presenta las características que a continuación se describen.

DataCleaner 1.5.4

DataCleaner está orientada a preparar los datos para cualquier proyecto en el que se deban aplicar técnicas de calidad de datos, es multiplataforma e incluye múltiples funcionalidades como (21):

- ✓ Data Profiling, para determinar la calidad de los datos.
- ✓ Data Validator, para validar datos contra reglas que deben verificarse bajo la política de calidad establecida.
- ✓ Comparator, para comparar la información de diferentes fuentes de origen.
- ✓ Monitor, para establecer un seguimiento de la calidad de los datos.
- ✓ Dictionary, permite crear un repositorio de datos maestros y correctos contra los que validar nuestros datos.

1.5.5 Herramientas para los procesos de extracción, transformación y carga

En la actualidad han aparecido nuevas posibilidades de adquisición de soluciones para el proceso de Extracción, Transformación y Carga (ETL), pues existen diversas y numerosas herramientas. Para apoyar el proceso de ETL en la integración de los datos al almacén, se utilizará el Pentaho Data Integration (PDI) en su versión 4.2.1.

Pentaho Data Integration 4.2.1.

Pentaho Data Integration 4.2.1 o también conocido como Kettle es un motor de transformaciones de código abierto diseñado para cubrir las necesidades en la integración de datos. Es una herramienta muy

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

intuitiva y multiplataforma. Las soluciones de Pentaho (Pentaho Data Integration, Pentaho BI Server y Pentaho Report Designer), se encuentran implementadas en java, lo cual hace que sea una solución muy flexible para cubrir las necesidades empresariales (22).

Las características básicas de esta herramienta son (23):

- ✓ Librería de transformaciones completa con más de 100 objetos de mapeo.
- ✓ 100% Java, amplio soporte multiplataforma y soporte de una amplia cantidad de fuentes de datos incluyendo aplicaciones integradas sobre 30 plataformas propietarias y código abierto, archivos planos, documentos excel y más.
- ✓ Basado en repositorio lo cual facilita la utilización de componentes de transformación, colaboración y administración de modelos, conexiones y logs.

Incluye cuatro herramientas:

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

1.5.6 Herramientas para la inteligencia de negocio

La Inteligencia de negocios es una práctica que busca transformar los datos de las compañías en información y conocimiento, para que los funcionarios que toman decisiones, cuenten con un apoyo de información actual e histórica del estado de la compañía y así determinar en donde están con respecto a sus objetivos y cuál es el camino más fácil para llegar a ellos (24).

En el desarrollo de la solución se utilizará para la construcción de los cubos OLAP el Schema Workbench 3.2.1, para las vistas de análisis el Pentaho BI Server en su versión 3.10 y el Pentaho Report Designer 3.8.2 para la creación de los reportes operacionales.

Schema Workbench 3.2.1

El Schema Workbench 3.2.1 tiene como objetivo el diseño de cubos OLAP. Su sencilla interfaz permite modelar un XML con el diseño del cubo a través de opciones lógicas e intuitivas que no requieren de un manejo avanzado de este formato de archivo (25).

Dentro de sus características se destacan (26):

- ✓ Diseñador intuitivo de esquemas OLAP.

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

- ✓ Permite crear, editar, actualizar y publicar esquemas OLAP para ser desplegados por aplicaciones de visualización de Pentaho.

Pentaho BI Server 3.10

Pentaho BI Server en su versión 3.10 es un servidor de integración de aplicaciones el cual incluye un motor de solución que integra informes y análisis. Funciona como un sistema basado en la web y está diseñado para integrarse fácilmente en los procesos de negocio. Sobre esta plataforma se definen las áreas de interés que poseen los usuarios para la preparación de reportes, consultas dinámicas y la realización del análisis OLAP (27).

Dentro de las características que proporciona esta herramienta de administración se destacan (27):

- ✓ Configuraciones de seguridad: provee un directorio centralizado desde donde es posible manejar las configuraciones de seguridad e integración con directorios ya existentes.
- ✓ Utilidades de repositorios: provee un set de utilidades para simplificar backups, migraciones y para restaurar repositorios relevantes de información.
- ✓ Usuarios y roles: permite definir en detalle los usuarios que tendrán acceso a las aplicaciones, junto con el nivel de acceso que estos tendrán a la información. Esto se realiza a través de la configuración de usuarios y roles, en el que se define qué permisos tendrá cada uno de ellos en el sistema.

Pentaho Report Designer 3.8.2

El Pentaho Report Designer (PRD) es una aplicación que proporciona un entorno de diseño visual para crear informes que pueden ser ejecutados y se guardan localmente a través del PRD o se publican en un servidor de Pentaho BI. Permite además a muchas personas a acceder y programar la ejecución de los informes. El PRD está dirigido a usuarios con experiencia, que están familiarizados con los conceptos y las fuentes de datos utilizadas (28).

Dentro de sus características se destacan (29):

- ✓ Permite crear informes y análisis.
- ✓ Proporciona tipos de salida tales como: PDF, Excel, HTML y archivos XML.
- ✓ Multiplataforma.

Conclusiones del capítulo

Luego de finalizado este capítulo, se concluye que fueron abordados los principales conceptos relacionados con los MD, lo que permite deducir la necesidad de crear un MD para el área Civil-Administrativo. Se decide utilizar como metodología de desarrollo la propuesta de Metodología de almacenes de datos en DATEC. Se seleccionaron como herramientas a utilizar el Visual Paradigm para UML en su versión 8.0 como herramienta de modelado, PostgreSQL 9.1 como gestor de base de datos, PgAdmin III 1.14.0 como administrador de base de datos, DataCleaner 1.5.4 para el perfilado y limpieza de los datos, Pentaho Data Integration 4.2.1 para la implementación del proceso de ETL, para el diseño de los cubos OLAP el Schema Workbench 3.2.1 y para el análisis de la información el Pentaho BI Server 3.10 y el Pentaho Report Designer 3.8.2.

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo.

Introducción

En el presente capítulo se realiza el análisis del MD mediante el levantamiento de los requisitos y su agrupamiento en casos de uso. El diseño del subsistema de almacenamiento del mercado de datos a través de la definición de las dimensiones, hechos y medidas identificadas, la construcción de la matriz buz y posteriormente el modelo de datos. Se realiza además el diseño del subsistema de integración de datos mediante el perfilado de los datos y el diseño general de las transformaciones y se diseña además el subsistema de visualización donde queda estructurada la arquitectura de información.

2.1 Necesidades de información

Para el proceso y confección de una herramienta informática se realizan diferentes etapas; sin embargo una de las más importantes es el levantamiento de requisitos. Para el caso de los AD, la etapa de captura de requisitos se realiza luego de que el cliente provea las necesidades de información y con ellas los indicadores que le servirán de ayuda para la toma de decisiones.

Para dar cumplimiento a esta etapa en la presente investigación, se realizaron entrevistas con los especialistas del departamento de estadística en el área Civil-Administrativo, con el propósito de lograr un mejor entendimiento del negocio y definir las necesidades de información.

Luego de varias entrevistas, se decide clasificar la información en siete grupos principales, los cuales se corresponden con los temas de análisis que agrupan toda la información solicitada por el usuario final, ellos son:

- ✓ Asuntos de familia
- ✓ Asuntos civiles
- ✓ Sentencias civiles
- ✓ Procesos de divorcios y alimentos
- ✓ Civil y administrativo
- ✓ Sentencias administrativas
- ✓ Apelaciones

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

2.2.1 Requisitos de información

Después de realizar un análisis del negocio se definen los requisitos de información existentes en el área Civil-Administrativo del Tribunal Supremo Popular.

Los requisitos de información (RI), describen la información que debe almacenar el sistema para satisfacer las necesidades del cliente (30). Estos fueron clasificados según los temas de análisis definidos.

Asuntos de familia

RI 1. Obtener la cantidad de expedientes de asuntos de familia del informe estadístico civil asuntos de familia por expedientes, temporal mes, dpa¹⁰ municipio, asuntos de familia municipal y tribunal municipal.

RI 2. Obtener la cantidad de expedientes de asuntos de familia del informe estadístico civil asuntos de familia por expedientes, temporal mes, dpa provincia, asuntos de familia provincial, tribunal provincial y tipo de sala.

Asuntos civiles

RI 3. Obtener la cantidad de expedientes de asuntos civiles del informe estadístico civil por expedientes, temporal mes, dpa municipio, asuntos civiles y tribunal municipal.

Sentencias civiles municipales

RI 4. Obtener la cantidad de indicadores de sentencias civiles del modelo de ejecución de sentencias Civiles por sentencias civiles, temporal mes, dpa municipio, tipo de ejecución y tribunal municipal.

RI 5. Obtener la cantidad de indicadores sentencias civiles del modelo de ejecución de sentencias civiles por sentencias civiles, temporal mes, dpa provincia, tipo de ejecución, tribunal provincial y tipo de sala.

Procesos de divorcios y alimentos

RI 6. Obtener la cantidad de las partes del informe estadístico para las partes en los procesos de divorcios y alimentos por tipo de proceso, sexo, color de piel, edad, ciudadanía, parentesco, último nivel de escolaridad aprobado, actividad que realiza o cargo que ocupa, integración, dpa municipio, tipo de condición y temporal mes.

Civil y administrativo

¹⁰ dpa: División política-administrativa

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

RI 7. Obtener la cantidad de expedientes civil administrativo del informe estadístico civil y administrativo por expedientes, temporal mes, dpa provincia, asuntos civil administrativo, tribunal provincial y tipo de sala.

Sentencias administrativas

RI 8. Obtener la cantidad de indicadores de sentencias administrativas del modelo de ejecución de sentencias administrativas por sentencias administrativas, temporal mes, dpa provincia, tribunal provincial y tipo de sala.

Apelaciones

RI 9. Obtener la cantidad de expedientes de apelaciones del informe estadístico de apelaciones TPP por expedientes, temporal mes, dpa provincia, apelaciones, tribunal provincial y tipo de sala.

2.2.2 Requisitos Funcionales

Los RF representan las capacidades y condiciones que el sistema debe cumplir (30). Para la realización del MD se identificaron 20 RF, a continuación se muestran 3 de los RF identificados, para consultar el resto ver el artefacto 0113_Especificación de requisitos de software del expediente de proyecto del MD Civil-Administrativo.

Subsistema de almacenamiento

RF 1. Acceder a la información persistente de los últimos diez años.

Subsistema de integración de datos

RF 2. Cargar los ficheros a partir del año 2004 hasta el 2012.

Subsistema de visualización

RF 3. Organizar la información por libros de trabajo.

2.2.3 Requisitos No Funcionales

Los requisitos no funcionales (RNF) son las propiedades o cualidades que el sistema debe cumplir (30). Para la realización del MD se definieron 12 RNF, los cuales se encuentran detallados en el artefacto 0113_Especificación de requisitos de software del expediente de proyecto del MD Civil-Administrativo, a continuación se muestran dos de estos requisitos.

Requisitos de software

RNF1. Instalar en las estaciones de trabajo el software necesario para el correcto funcionamiento del sistema.

Las máquinas clientes deben contar con los siguientes elementos:

- ✓ Sistema operativo Windows XP, GNU/Linux. Ubuntu 10.04 o una versión superior respectivamente.
- ✓ Navegador web Mozilla Firefox 2.0 o una versión superior.

El servidor de base de datos debe contar con los siguientes elementos:

- ✓ Sistema operativo Windows server 2003, GNU/Linux. Ubuntu 10.04 o una versión superior respectivamente.
- ✓ PostgreSQL 9.1 como SGBD.
- ✓ PgAdmin III 1.14.0 como interfaz de administración del SGBD.

El servidor de aplicaciones debe contar con los siguientes elementos:

- ✓ Sistema operativo Windows XP, GNU/Linux. Ubuntu 10.04 o una versión superior respectivamente.
- ✓ Máquina virtual de Java 6.0 o una versión superior.
- ✓ Pentaho BI Server 3.10.
- ✓ Pentaho Report Designer 3.8.2.
- ✓ Schema Workbench 3.2.1.

Requisito de hardware

RNF2. Definir las interfaces de hardware que soportará el sistema.

El sistema solamente podrá interactuar con una interfaz de hardware: la impresora. Esta interacción se ocasionará cuando sea necesario imprimir en formato físico un reporte o vista de análisis. El acceso a la impresora se realizará a través del protocolo TCP/IP.

2.2 Reglas del negocio

Las reglas del negocio (RN) describen las políticas, normas, operaciones, definiciones y restricciones del negocio (30), las cuales son de vital importancia para alcanzar los objetivos.

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

Reglas de variables

RF 1. Los pendientes al final se calculan de la siguiente manera: (pendientes al inicio + radicados + procesos retrotraídos + nuevo curso) - total de resueltos.

Reglas de almacenamiento

RF 2. Los identificadores de los indicadores no pueden estar repetidos ni tomar valores nulos.

Reglas de transformación

RF 3. El pendiente al final de un determinado mes será el pendiente al inicio del mes siguiente.

Reglas de visualización:

RF 4. Los porcentos deben visualizarse solo con dos espacios después de la coma.

Las demás RN se pueden encontrar en el artefacto Reglas de negocio del expediente de proyecto del MD Civil-Administrativo.

2.3 Diagrama de Casos de uso del sistema

Los casos de uso representan la información de manera visual, ayudando a describir qué es lo que el sistema debe hacer relacionado con los usuarios que interactúan con él (31). Los RI y RF se agrupan en casos de uso de información (CUI) y casos de uso funcionales (CUF) respectivamente. Para la realización del MD Civil-Administrativo se identificaron siete CUI y seis CUF. En el diagrama de casos de uso del sistema se utilizaron los patrones: crud total y múltiples actores (ver figura 3).

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

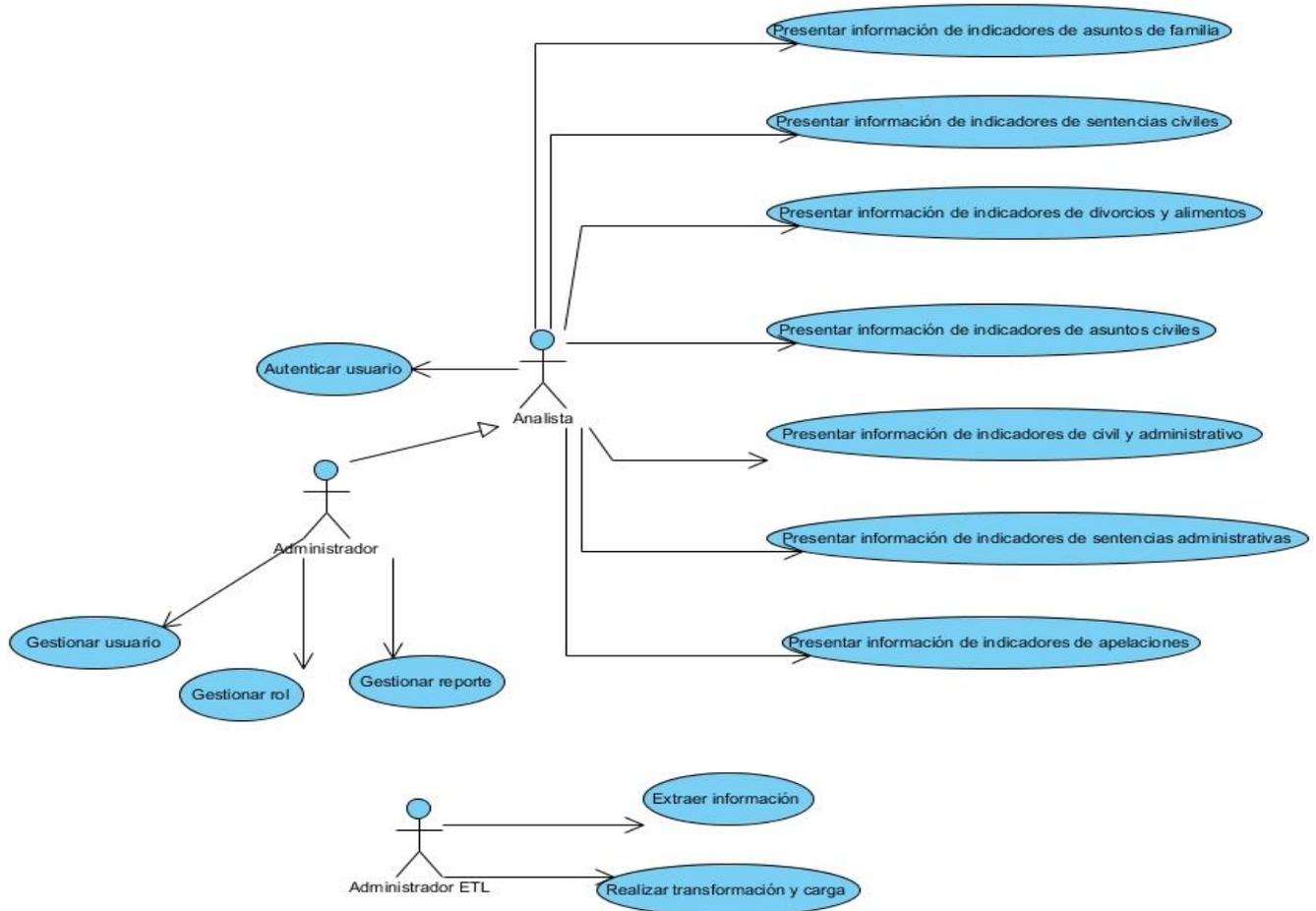


Figura 3: Diagrama de casos de uso del sistema.

2.3.1 Descripción de los actores

- ✓ **Administrador:** encargado de administrar los usuarios del sistema, asignarles sus respectivos roles, además de administrar dichos roles y los reportes.
- ✓ **Analista:** encargado de los casos de uso relacionados con la presentación de la información del MD.
- ✓ **Administrador ETL:** encargado de llevar a cabo los procesos de extracción, transformación y carga de los datos del sistema fuente.

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

2.4 Casos de uso de información

Tabla 1: Descripción de los casos de uso de información.

| Nombre | Descripción |
|---|---|
| Presentar información de indicadores de asuntos de familia. | Visualiza los reportes de los requisitos de información de los asuntos de familia municipales y provinciales del área Civil Administrativa. |
| Presentar información de indicadores de asuntos civiles. | Visualiza los reportes de los requisitos de información de los asuntos civiles del área Civil Administrativa. |
| Presentar información de indicadores de sentencias civiles. | Visualiza los reportes de los requisitos de información de las sentencias civiles municipales y provinciales del área Civil Administrativa. |
| Presentar información de indicadores de divorcios y alimentos. | Visualiza los reportes de los requisitos de información de los procesos de divorcios y alimentos del área Civil Administrativa. |
| Presentar información de indicadores de civil y administrativo. | Visualiza los reportes de los requisitos de información de civil y administrativo del área Civil Administrativa. |
| Presentar información de indicadores de sentencias administrativas. | Visualiza los reportes de los requisitos de información de las sentencias administrativas del área Civil Administrativa. |
| Presentar información de indicadores de apelaciones. | Visualiza los reportes de los requisitos de información de las apelaciones del área Civil Administrativa. |

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

A continuación se describe el CUI Presentar información de indicadores de asuntos de familia. Las restantes descripciones de los CU se encuentran en el artefacto 0114_Especificación de los casos de uso del expediente de proyecto del MD Civil-Administrativo.

Tabla 2: Especificación de CUI Presentar información de indicadores de asuntos de familia.

| | | |
|--|---|---|
| Objetivo | Mostrar toda la información referente a los indicadores de asuntos de familia. | |
| Actores | Analista: (Inicia) seleccionando el A.A.G. SIEJT. | |
| Resumen | El caso de uso se inicia cuando el actor desea hacer un análisis de la información relacionada con los asuntos de familia desde diferentes perspectivas. El actor selecciona el reporte que desea ver, el sistema muestra la información contenida en él, selecciona las opciones de los posibles cambios que le puede hacer al reporte. Finaliza el caso de uso. | |
| Complejidad | Alta | |
| Prioridad | Alta | |
| Precondiciones | El usuario se autenticó correctamente. Los datos correspondientes fueron cargados en el mercado de datos. Los reportes relacionados con los asuntos de familia fueron creados. | |
| Post-condiciones | Los reportes correspondientes al caso de uso fueron consultados. | |
| Flujo de eventos | | |
| Flujo básico Presentar información de indicadores de Asuntos de familia. | | |
| | Actor | Sistema |
| 1. | Selecciona el A.A.G. SIEJT para los Tribunales Populares. | |
| 2. | | Muestra las A.A. contenidas en el A.A.G. SIEJT para los Tribunales Populares. |
| 3. | Selecciona el A.A. Civil-Administrativo. | |

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

| | | |
|---|---|--|
| 4. | | Muestra las A.A. contenidas en el A.A. Civil-Administrativo. |
| 5. | Selecciona el A.A. instancia municipal o el A.A. instancia provincial. | |
| 6. | | Muestra las A.A. contenidas en el A.A. instancia municipal o el A.A. instancia provincial. |
| 7. | Selecciona el A.A. DPA (1976-2010) o el A.A. DPA (2011-actualidad). | |
| 8. | | Muestra los L.T. contenidos en el A.A. DPA (1976-2010) o el A.A. DPA (2011-actualidad). |
| 9. | Selecciona el L.T. 1321 Informe estadístico civil asuntos de familia o el L.T. 3271 Informe estadístico asuntos de familia. | |
| 10. | | Muestra los reportes contenidos en el L.T. 1321 Informe estadístico civil asuntos de familia o el L.T. 3271 Informe estadístico asuntos de familia. |
| 11. | Selecciona el reporte que desea analizar. | |
| 12. | | Muestra la información contenida en el reporte seleccionado y brinda opciones al actor para hacer cambios al reporte durante su análisis. Finaliza el CU. |
| Flujos alternos | | |
| | | |
| Opciones de reportes de Presentar información de indicadores de asuntos de familia | | |

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

| Perspectivas de análisis | Posibles resultados | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|--|--|--------------------|---|------------|---|-----------------|----------------------------------|-------------|--|---------|--|---------|--|---|-----------------|---|-----------------|---|------------|---|-----------------|----------------------------------|--|--|--|---|--|--|--|--|--|--|--|--|--|--|--|
| | Medidas | Periodicidad | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Variables de entrada relacionadas con el CUI Presentar información de indicadores de asuntos de familia: <ul style="list-style-type: none"> • temporal mes • dpa municipio • dpa provincia • expedientes • asuntos de familia municipal • asuntos de familia provincial • tribunal municipal • tribunal provincial • tipo sala | Variables de salida disponibles en el hecho indicadores de asuntos de familia: <ul style="list-style-type: none"> • Cantidad de expedientes de asuntos de familia | Rango de tiempo en que se solicitan las variables de salida: <ul style="list-style-type: none"> • Mensual | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Prototipo de interfaz de usuario: | <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th colspan="4" style="background-color: #e0e0e0;">Asuntos de familia</th> <th colspan="2" style="background-color: #e0e0e0;">Provincias</th> <th colspan="2" style="background-color: #e0e0e0;">Expedientes</th> <th colspan="2" style="background-color: #e0e0e0;">Tiempo</th> <th colspan="2" style="background-color: #e0e0e0;">Medidas</th> </tr> </thead> <tbody> <tr> <td style="background-color: #e0e0e0;">+</td><td style="background-color: #e0e0e0;">Acumulado total</td> <td style="background-color: #e0e0e0;">+</td><td style="background-color: #e0e0e0;">Acumulado total</td> <td style="background-color: #e0e0e0;">+</td><td style="background-color: #e0e0e0;">Expediente</td> <td style="background-color: #e0e0e0;">+</td><td style="background-color: #e0e0e0;">Acumulado total</td> <td colspan="4" style="background-color: #e0e0e0;">• Cantidad de asuntos de familia</td> </tr> <tr> <td colspan="12" style="background-color: #e0e0e0;"> <div style="display: flex; justify-content: space-between;"> Asuntos de familia Provincias Expedientes Tiempo Medidas </div> </td> </tr> </tbody> </table> | | Asuntos de familia | | | | Provincias | | Expedientes | | Tiempo | | Medidas | | + | Acumulado total | + | Acumulado total | + | Expediente | + | Acumulado total | • Cantidad de asuntos de familia | | | | <div style="display: flex; justify-content: space-between;"> Asuntos de familia Provincias Expedientes Tiempo Medidas </div> | | | | | | | | | | | |
| | Asuntos de familia | | | | Provincias | | Expedientes | | Tiempo | | Medidas | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| + | Acumulado total | + | Acumulado total | + | Expediente | + | Acumulado total | • Cantidad de asuntos de familia | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <div style="display: flex; justify-content: space-between;"> Asuntos de familia Provincias Expedientes Tiempo Medidas </div> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

2.5 Casos de uso funcionales

Los RF se agruparon en seis CUF, los cuales se presentan a continuación:

CUF 1. Autenticar usuario.

CUF 2. Gestionar rol.

CUF 3. Gestionar usuario.

CUF 4. Gestionar reporte.

CUF 5. Extraer información.

CUF 6. Realizar transformación y carga.

A continuación se describe el CUF Autenticar usuario. Las restantes descripciones de los CU se encuentran en el artefacto 0114_Especificación de los casos de uso del expediente de proyecto del MD Civil-Administrativo.

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

Tabla 3: Especificación de CUF autenticar usuario.

| | | |
|--|---|--|
| Objetivo: | Permitir que la persona se autentique y acceda al sistema si es un usuario del mismo. | |
| Actores: | Analista, Administrador. | |
| Resumen: | Inicia cuando el actor se conecta al sistema para acceder a las funcionalidades del mismo. Para ello debe proporcionar un usuario y contraseña. El caso de uso termina una vez validado el usuario y contraseña cuando el usuario logra acceder al sistema. | |
| Complejidad: | Media | |
| Prioridad: | Media | |
| Precondiciones: | El sistema debe estar disponible. El usuario debe existir en la BD. | |
| Post-condiciones: | El usuario ha sido autenticado en el sistema. | |
| Flujo Normal de Eventos | | |
| 1. El usuario interactúa con el sistema. | 2. El sistema muestra un formulario para introducir los datos. | |
| 3. El usuario introduce los datos. | 4. El sistema valida los datos. | |
| | 5. El sistema le da los permisos para entrar a la aplicación. Finaliza el C.U. | |
| | | |
| Flujo Alternativo | | |
| Acción del Actor | Respuesta del Sistema | |
| | 5.1 Si los datos no son válidos, muestra un mensaje de error y regresa al punto 2 del Flujo Normal de Eventos . | |

| | |
|--|--------------------------|
| | Finaliza el caso de uso. |
|--|--------------------------|

2.6 Arquitectura

El MD Civil-Administrativo tiene una arquitectura compuesta por las fuentes de datos, el subsistema de integración, el subsistema de almacenamiento y el subsistema de visualización (ver figura 4).

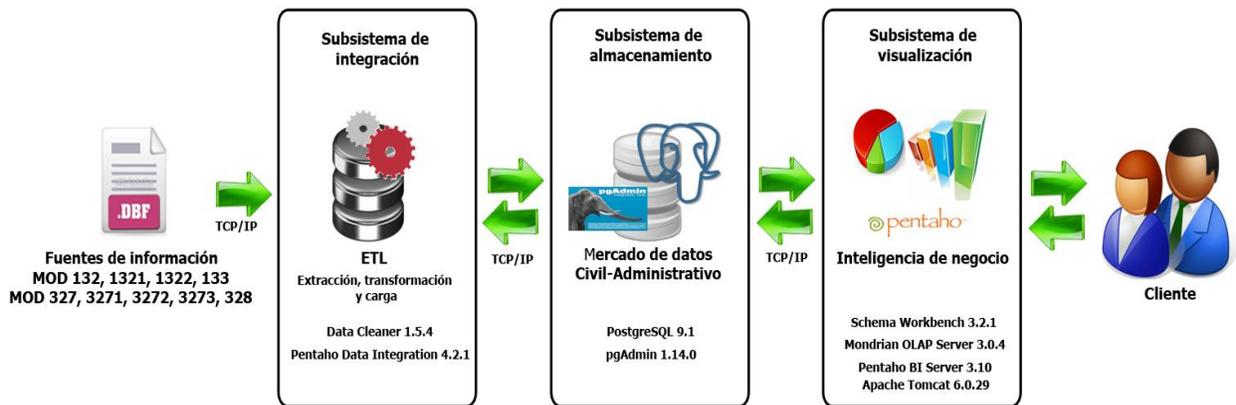


Figura 4: Arquitectura de información del MD Civil-Administrativo.

Subsistema de integración: la información correspondiente al área Civil-Administrativas está contenida en diferentes modelos en formato dbf, la cual es extraída para realizarle procesos de limpieza, transformación e integración, quedando los datos listos para la carga en la BD del área Civil-Administrativo.

Subsistema de almacenamiento: en este subsistema se crea la BD donde se almacenarán los datos arrojados por el subsistema de integración luego de realizados los procesos de ETL, mediante el protocolo TCP/IP.

Subsistema de visualización: una vez poblada la BD se procede a la creación de los cubos OLAP, donde se definen las dimensiones, medidas y niveles de jerarquía de cada dimensión. Posteriormente se generan las vistas de análisis y los reportes correspondientes al área Civil-Administrativa.

2.7 Diseño del subsistema de almacenamiento

Para la realización y correcto funcionamiento del MD se confecciona el modelo dimensional de los datos, el cual contiene las tablas de hechos que se identificaron en el negocio, las dimensiones y las relaciones entre estas.

2.7.1 Dimensiones

Las tablas de dimensiones contienen información descriptiva, las cuales están descrita por un conjunto de atributos organizados en jerarquías y niveles (32).

Seguidamente se describen algunas dimensiones que fueron identificadas en el MD Civil-Administrativo y su nivel jerárquico, las restantes dimensiones se encuentran en el artefacto Especificación del modelo de datos del expediente de proyecto del MD Civil-Administrativo.

D 1. Dimensión temporal mes (dim_temporal_mes): esta dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo a los meses y años en lo que se almacena.

Jerarquía: temporal_anno , temporal_mes.

D 2. Dimensión dpa municipio (dim_dpa_municipio): esta dimensión almacena los valores pertenecientes a la división política administrativa que presenta el país, atendiendo a las provincias y los municipios.

Jerarquía: dpa_municipio_2011, dpa_municipio_1976.

D 3. Dimensión asuntos de familia municipal (dim_asuntos_familia_municipal): esta dimensión describe los valores bajo los cuales puede clasificarse la información atendiendo a los asuntos de familia.

Jerarquía: asuntos_familia_municipal.

2.7.2 Hechos

Los hechos son aquellos datos que proporcionan una información cuantitativa sobre las características del negocio (32), los cuales están compuestos por medidas y las llaves foráneas de las dimensiones con las que se encuentran relacionados. A continuación se describen los hechos que fueron identificados en el MD Civil-Administrativo.

H 1. Hecho indicadores asuntos familia municipal

(hech_indicadores_asuntos_familia_municipal): en este hecho se almacena toda la información relacionada con indicadores definidos para los asuntos de familia municipales.

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

H 2. Hecho indicadores asuntos familia provincial (hech_indicadores_asuntos_familia_provincial): en este hecho se almacena toda la información relacionada con indicadores definidos para los asuntos de familia provinciales.

H 3. Hecho indicadores asuntos civiles (hech_indicadores_asuntos_civiles): en este hecho se almacena toda la información relacionada con indicadores definidos para los asuntos civiles.

Para más detalles consultar el artefacto Especificación del modelo de datos del expediente de proyecto del MD Civil-Administrativo.

2.7.3 Medidas

Una medida es un atributo numérico o de tipo texto, la cual representa el análisis del hecho. Las medidas se clasifican de acuerdo a su aditividad (6):

- ✓ **Aditivas:** pueden ser agregadas a través de cualquier dimensión.
- ✓ **Semiaditivas:** pueden no ser adicionadas a través de una o más dimensiones.
- ✓ **No aditivas:** no pueden ser añadidas a través de ninguna dimensión.

Además dependiendo del tipo de la variable que puede ser:

- ✓ **Física:** cuando la variable es algo físico que debe estar en el almacén.
- ✓ **Calculable:** es cuando la variable se calcula de alguna manera teniendo en cuenta medidas físicas del almacén, y se calculan a través de BI.

En la presente investigación se identificaron nueve medidas, las cuales se encuentran reflejadas en el artefacto Especificación del modelo de datos del expediente de proyecto del MD Civil-Administrativo.

2.7.4 Matriz buz

En la tabla 4 se observa una porción de la matriz bus que representa la relación que existe entre las dimensiones y los hechos del MD Civil-Administrativo. Constituye una validación del análisis realizado evitando que exista solapamiento entre los hechos. La matriz bus permite además tener una perspectiva de cómo queda estructurado el modelo de datos e identificar las dimensiones que son comunes para otros MD. A continuación se muestra un fragmento de la misma:

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

Tabla 4: Matriz bus.

| Dimensiones/Hechos | H1 | H2 | H3 | H4 | H5 | H6 | H7 | H8 | H9 |
|-------------------------------|----|----|----|----|----|----|----|----|----|
| temporal_mes | X | X | X | X | X | X | X | X | X |
| dpa_municipio | X | | X | X | | | | | X |
| dpa_provincia | | X | | | X | X | X | X | |
| tribunal_provincial | | X | | | X | X | X | X | |
| tribunal_municipal | X | | X | X | | | | | X |
| tipo_sala | | X | | | X | X | X | X | |
| expedientes | X | X | X | | | | X | X | |
| asuntos_de_familia_municipal | X | | | | | | | | |
| asuntos_de_familia_provincial | | X | | | | | | | |
| asuntos_civiles | | | X | | | | | | |
| sentencias_c | | | | X | X | | | | |
| sentencias_a | | | | | | X | | | |
| tipo_ejecucion | | | | X | X | | | | |
| apelaciones | | | | | | | X | | |
| asuntos_civil_administrativo | | | | | | | | X | |
| raza | | | | | | | | | X |
| edad | | | | | | | | | X |
| ciudadania | | | | | | | | | X |
| parentesco | | | | | | | | | X |
| sexo | | | | | | | | | X |
| actividad_que_realiza | | | | | | | | | X |
| tipo_proceso | | | | | | | | | X |
| Condición | | | | | | | | | X |
| nivel_escolaridad | | | | | | | | | X |
| integracion | | | | | | | | | X |
| divorcios_alimentos | | | | | | | | | X |

Para más detalles consultar el artefacto Especificación del modelo de datos del expediente de proyecto del MD Civil-Administrativo.

2.7.5 Topología de los almacenes de datos

Para el diseño del modelo de datos se debe tener en cuenta la topología que se empleará. Es necesario analizar cuál de las tres topologías existentes es la que se adapta a los requisitos y necesidades del cliente.

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

Esquema en estrella: está formado por una o varias tablas de dimensiones relacionadas con una tabla principal de hechos. La llave primaria en la tabla de hechos se compone de la relación de las llaves primarias de las tablas de dimensiones (ver figura 5) (33).

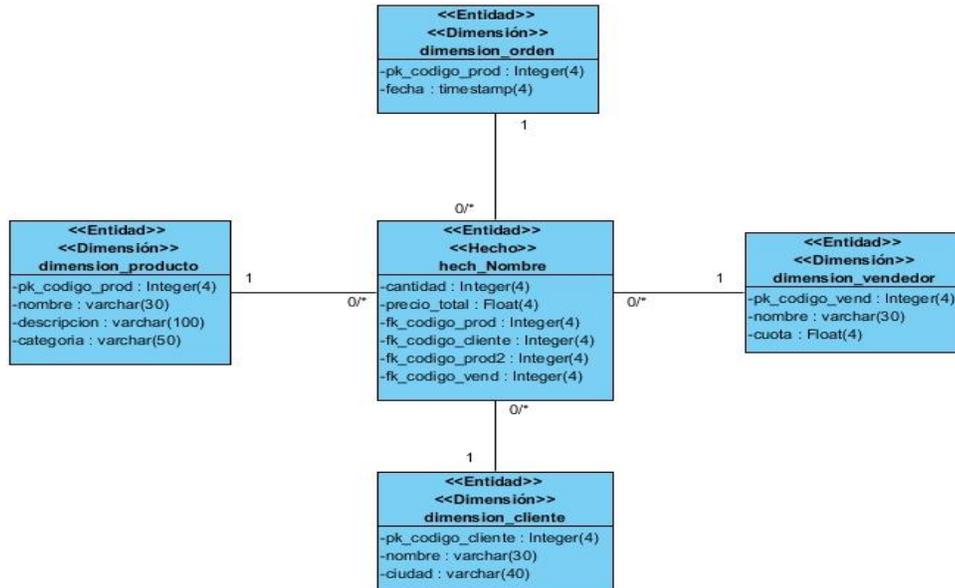


Figura 5: Representación del esquema en estrella.

Esquema copo de nieve: constituye una extensión del modelo en estrella, cuando las dimensiones se ordenan teniendo en cuenta sus jerarquías. La estructura de las tablas de las dimensiones se encuentran normalizadas (ver figura 6) (33).

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

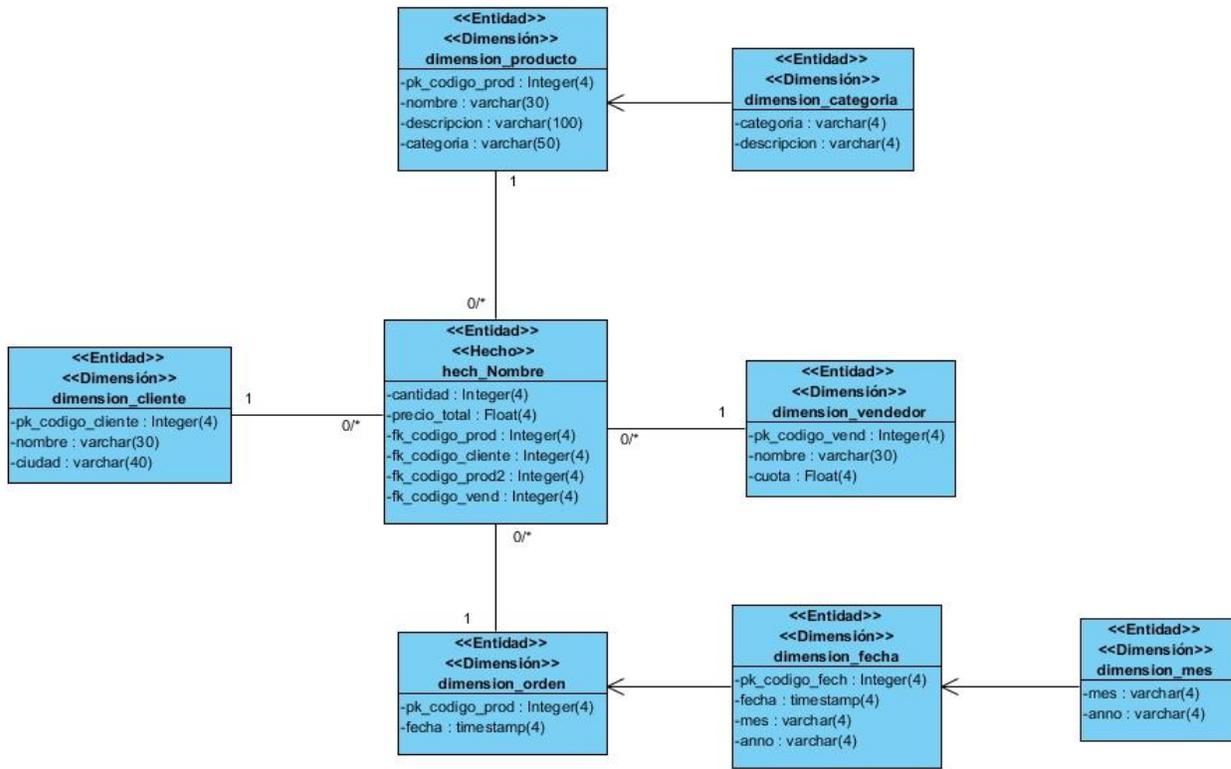


Figura 6: Representación del esquema copo de nieve.

Constelaciones de hechos: este esquema es la generalización de los esquemas en estrella y copo de nieve con la inclusión de distintas tablas de hechos que comparten todas o algunas tablas de dimensiones. Este esquema es más complejo que las otras topologías debido a que contiene múltiples tablas de hechos (33).

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

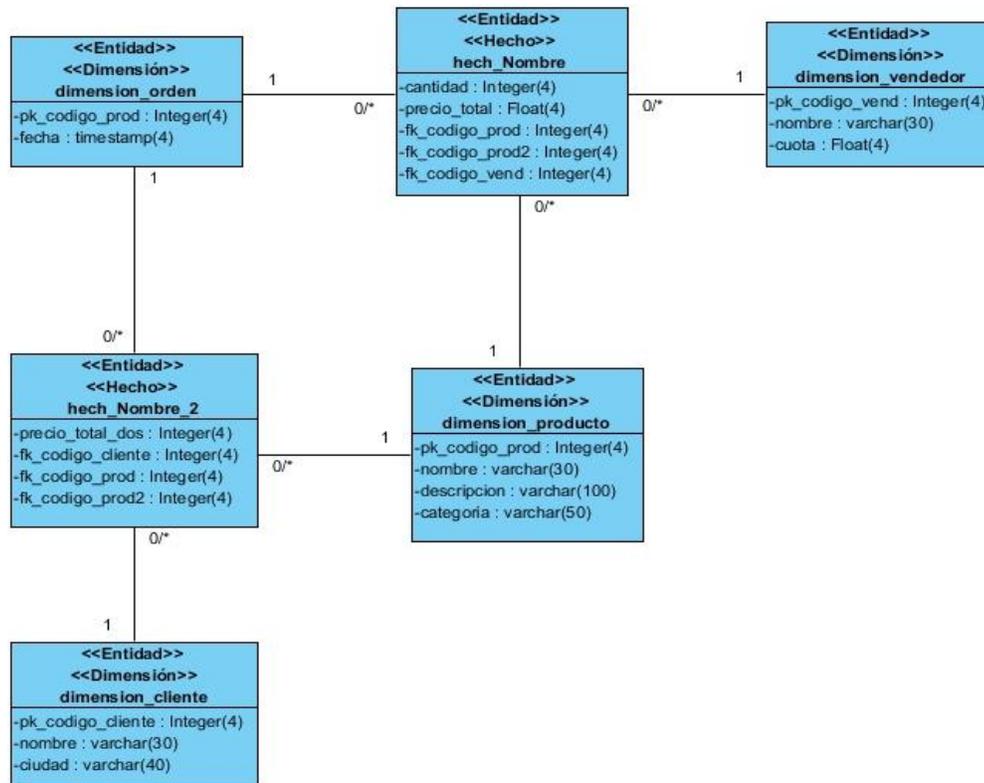


Figura 7: Representación del esquema constelación de hechos.

Para la presente investigación se selecciona el esquema constelación de hechos, debido a que existen varias tablas de hechos que comparten varias dimensiones.

2.7.6 Modelo de datos

Una vez definidos los elementos descritos anteriormente, se da paso al diseño del modelo dimensional. El principal objetivo es que los datos del negocio queden representados en una estructura lógica, evidenciando las relaciones que existen entre las tablas de hechos y las dimensiones.

En la figura 8 se observan un fragmento del modelo de datos, en el Anexo 1 se encuentra el modelo dimensional, donde se identificaron nueve tablas de hechos y 26 tablas de dimensiones, evidenciando una topología de constelación de hechos.

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

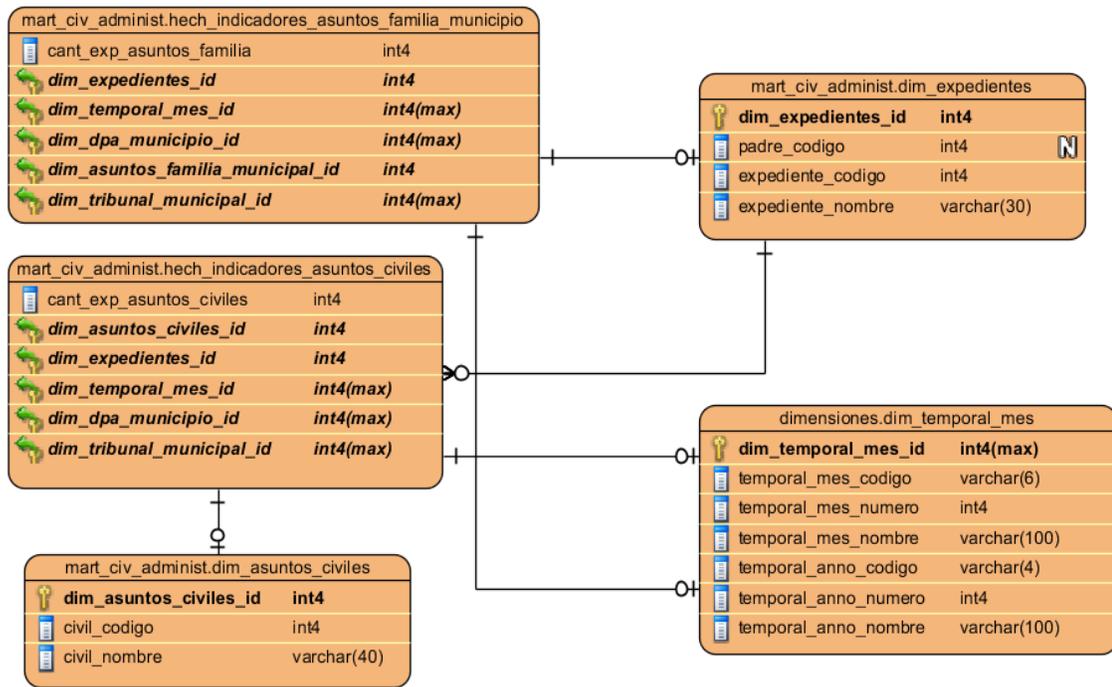


Figura 8: Fragmento del modelo de datos dimensional con topología de constelación de hechos.

2.8 Diseño del subsistema de integración

Para proceder con el diseño del subsistema de integración se hace necesario realizar el perfilado de los datos, con el fin de desarrollar una comprensión profunda del contenido, estructura y calidad de estos. Un sistema de perfilado de datos puede procesar grandes cantidades de información, y con las habilidades del analista se identifican las principales dificultades que presenta la fuente; con el objetivo de establecer nuevas reglas de negocio y transformación (34).

Una vez realizado el perfilado de datos a la fuente de información, se pudo observar la cantidad de campos nulos y no nulos correspondiente a los años 2010, 2011 y 2012 (ver figura 9).

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

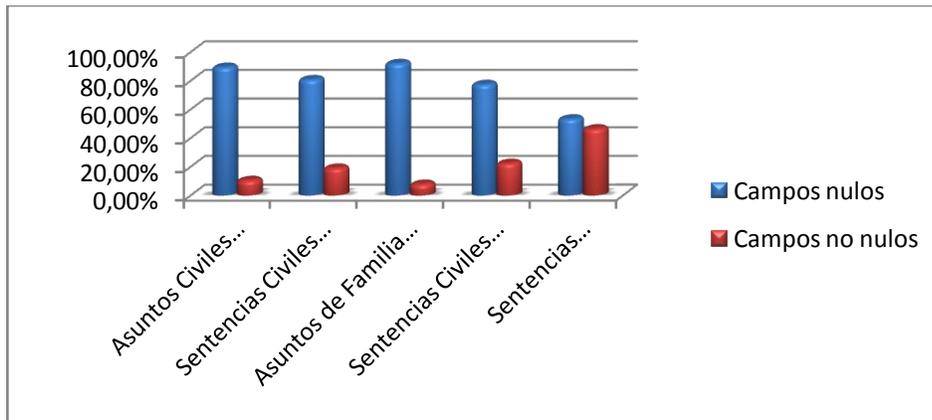


Figura 9: Distribución en porcentaje de los campos nulos y no nulos.

El perfilado de datos a la fuente permitió conocer la distribución de los tipos de datos (ver figura 10).

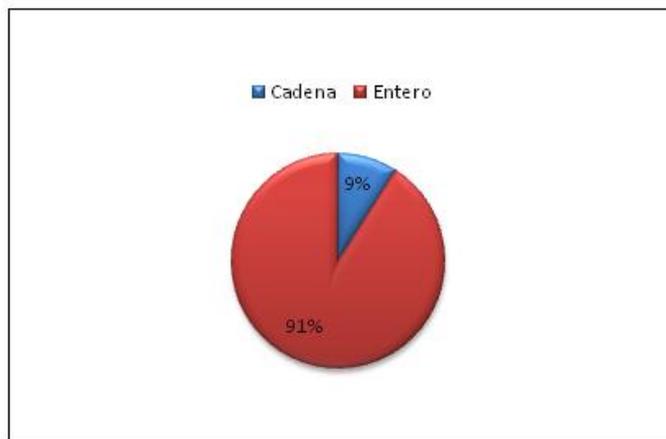


Figura 10: Distribución de los tipos de datos.

Al realizar el análisis de los resultados arrojados por el perfilado de los datos de los ficheros fuentes se concluye que:

- ✓ Los datos encontrados son de tipo: cadenas y enteros.
- ✓ La cantidad máxima de caracteres encontrados es 39.
- ✓ Los campos que aparezcan nulos serán sustituidos por ceros.

Para más información del perfilado de los datos consultar el artefacto Perfil de los datos del expediente de proyecto del MD Civil-Administrativo.

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

2.8.1 Diseño de los procesos de integración

Una vez diseñado el subsistema de almacenamiento y conocida la estructura, contenido y fiabilidad de los datos, se procede a realizar el diseño de las transformaciones. Estos diseños pueden variar a la hora de la implementación de las transformaciones, porque durante el proceso de desarrollo suelen surgir situaciones con los datos y se llevan a cabo disímiles estrategias para resolverlas. El diseño queda estructurado de la siguiente manera (ver figura 12).

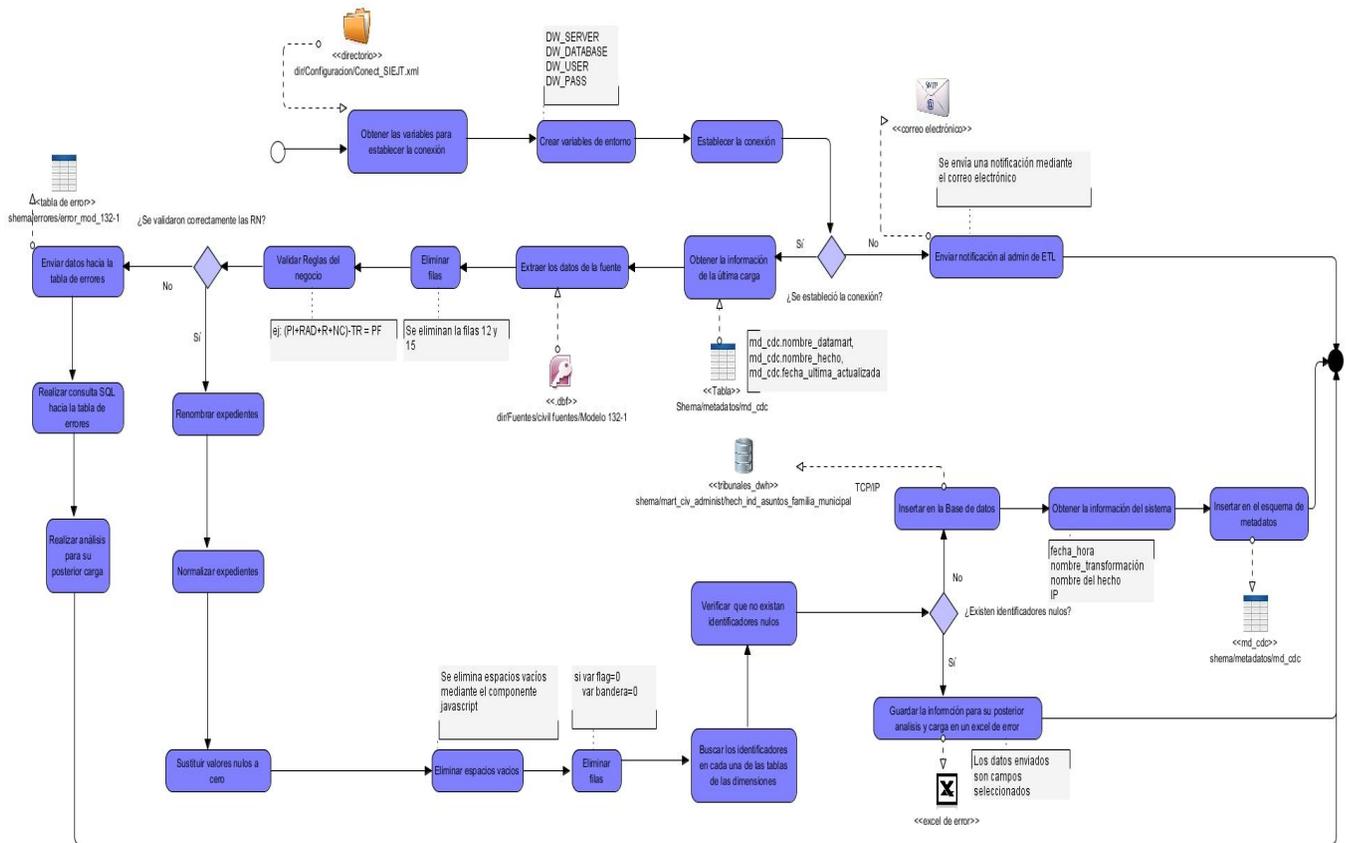


Figura 11: Diseño general de las transformaciones para la carga del hecho asuntos de familia.

2.9 Diseño del subsistema de visualización

El subsistema de visualización constituye un elemento fundamental para el MD, debido a que el usuario final interactúa con él. Por ello realizar un correcto y detallado diseño del mismo brinda muchas facilidades en la implementación.

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

2.9.1 Arquitectura de la información

Según las necesidades del cliente en el Área de Análisis General (A.A.G.) SIEJT se propone el Área de Análisis (A.A.) Civil-Administrativo, la cual contendrá el A.A. Instancia municipal para obtener los informes correspondientes a las instancias municipales y el A.A. Instancia provincial para obtener los informes correspondientes a las instancias provinciales. Estas áreas contienen el A.A. DPA (1976-2010) para obtener las vistas de análisis correspondientes a la división político administrativa antes del 2010 y el A.A. DPA (2011-Actualidad) para obtener los datos correspondientes a la nueva división político administrativa. Estas A.A. agrupan la información referente a los indicadores en los Libros de Trabajo (L.T.).

En los L.T. se encuentran las vistas de análisis referentes a los indicadores: asuntos de familia, sentencias civiles, divorcios y alimentos, asuntos civiles, asuntos civil y administrativo, sentencias administrativas y apelaciones. A continuación se presenta el mapa de navegación, el cual constituye una representación del diseño definido para la arquitectura de información del MD (ver figura 14).

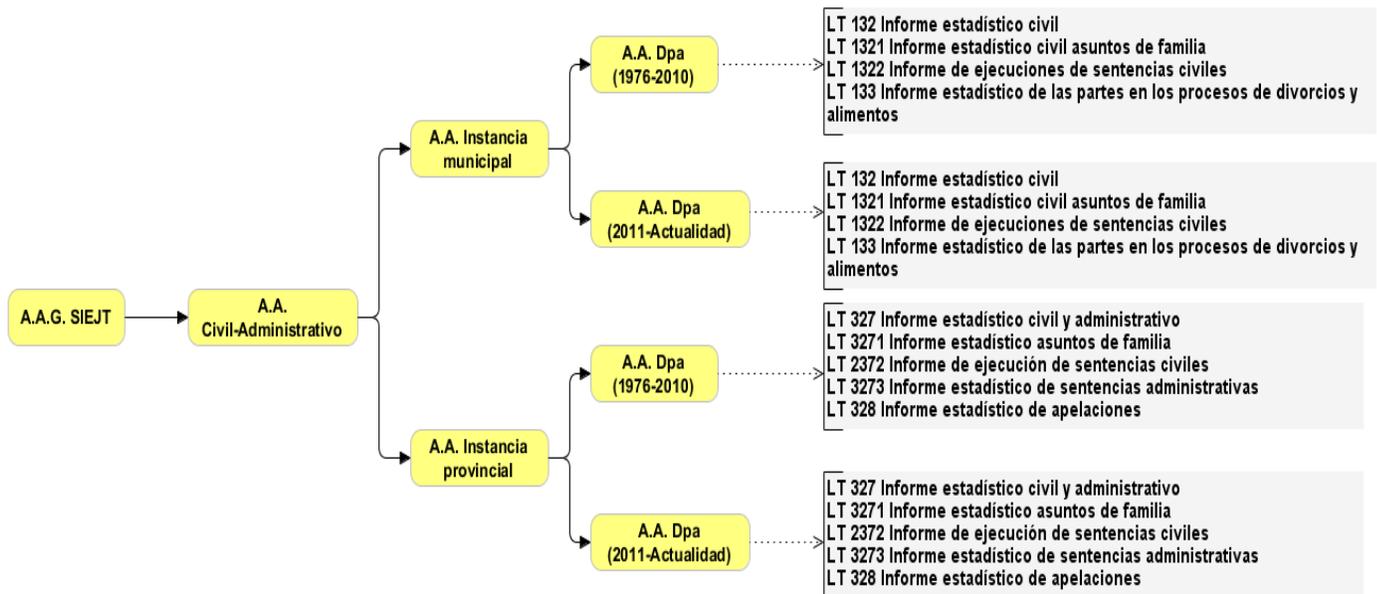


Figura 12: Mapa de navegación del MD Civil-Administrativo.

A continuación se describen los libros de trabajo:

A.A. DPA (1976-2010) y A.A. DPA (2011-Actualidad)

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

L.T. 132 Informe estadístico civil: contiene tres reportes con los que se logra realizar un análisis general de los datos en función de los indicadores cantidad de expedientes de asuntos civiles.

L.T. 133 Informe estadístico para las partes en los procesos de divorcios y alimentos: contiene un reporte con el que se logra realizar un análisis general de los datos en función del indicador cantidad de las partes.

L.T. 327 Informe estadístico civil y administrativo: contiene dos reportes con los que se logra realizar un análisis general de los datos en función del indicador cantidad de expedientes de civil y de administrativo.

L.T. 328 Informe estadístico de apelaciones TPP: contiene un reporte con el que se logra realizar un análisis general de los datos en función del indicador cantidad de expedientes de apelaciones.

L.T. 1321 Informe estadístico civil asuntos de familia: contiene dos reportes con los que se logra realizar un análisis general de los datos en función del indicador la cantidad de expedientes de asuntos de familia municipal.

L.T. 1322 Informe de ejecución de sentencias civiles: contiene cuatro reportes con los que se logra realizar un análisis general de los datos en función del indicador cantidad de indicadores de sentencias civiles municipales.

L.T. 3271 Informe estadístico asuntos de familia: contiene dos reportes con los que se logra realizar un análisis general de los datos en función del indicador cantidad de expedientes de asuntos de familia provincial.

L.T. 3272 Informe de ejecución de sentencias civiles: contiene cuatro reportes con los que se logra realizar un análisis general de los datos en función del indicador cantidad de indicadores de sentencias civiles provinciales.

L.T. 3273 Informe de ejecución de sentencias administrativas: contiene un reporte con el que se logra realizar un análisis general de los datos en función del indicador cantidad de indicadores de sentencias administrativas.

Diseño de las vistas de análisis

En el diseño del subsistema de visualización es importante detallar las vistas de análisis, estos dan respuesta a las necesidades de los usuarios, permitiendo identificar las salidas que debe producir el sistema. A continuación se muestra un ejemplo de una vista de análisis "Comportamiento de los asuntos de familia en los TPP".

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

Tabla 5: Comportamiento de los asuntos de familia en los TPP.

| | |
|--|--|
| Área de Análisis (A.A) | Civil-Administrativo |
| Área de Análisis (A.A) | Instancia provincial |
| Área de Análisis (A.A) | DPA (2011-Actualidad) |
| Libro de trabajo (L.T) | 3271 Informe estadístico asuntos de familia |
| Reporte (Tabla de Salida – TS) | TS1- Cantidad de expedientes de asuntos de familia del informe estadístico civil asuntos de familia por expedientes, temporal mes, dpa provincia, asuntos de familia provincial, tribunal provincial y tipo de sala. |
| Descripción | La vista de análisis muestra el número de expedientes de asuntos de familia del informe estadístico civil asuntos de familia provincial. |
| Elementos de la vista de análisis | <p>Perspectivas de análisis</p> <ul style="list-style-type: none"> • expedientes • temporal_mes • dpa_provincia • asuntos_familia_provincial • tribunal_provincial • tipo_sala <p>Indicador a medir</p> <ul style="list-style-type: none"> • cant_exp_asuntos_familia |

2.10 Esquema de seguridad

En el mercado de datos Civil-Administrativo es fundamental la seguridad de la información, para esto se crearon roles con el objetivo de darle los permisos debidos a los usuarios. Para la seguridad en la BD queda definido el rol Administrador de ETL, el cual posee acceso total a la BD, además es el encargado de realizar los procesos de ETL. Para la seguridad en la aplicación se define el rol Administrador, el cual posee acceso total al A.A.G. SIEJT y además es el encargado de crear nuevos usuarios, así como asignarle los roles y los permisos a cada usuario; y el rol Analista que tiene acceso al A.A. Civil-

Capítulo 2: Análisis y diseño del mercado de datos Civil-Administrativo

Administrativo y de este modo puede consultar los LT correspondientes a esta área, así como sus reportes.

2.10.1 Políticas de respaldo y recuperación

En el mercado de datos Civil-Administrativo se lleva a cabo una política de respaldo y recuperación basada en la periodicidad de las salvas, esto se evidencia en la realización mensual de las salvas de la información contenida en la BD. Las tablas involucradas en este proceso son las tablas de hechos y las tablas de dimensiones identificadas en la fase de análisis. Aunque no existen salvas en esta área hasta el momento, se prevé que estas sean reemplazadas mensualmente, así como chequear su estado mediante pruebas de rendimiento.

Conclusiones del capítulo

Una vez concluido el capítulo se arrojan las siguientes conclusiones:

Se identificaron nueve RI y 20 RF agrupados en siete CUI y seis CUF respectivamente, sirviendo de base para elaborar el diagrama de casos de uso del sistema. De igual forma se identificaron las RN que servirán de apoyo para realizar las transformaciones en los procesos de integración. El modelo dimensional diseñado representa las relaciones que se establecen entre las nueve tablas de hechos, las 26 dimensiones identificadas para el MD. Con el fin de poblar la BD se realizó el diseño general de las transformaciones. En correspondencia con las necesidades del cliente se diseñaron 36 reportes candidatos. Para la seguridad de la aplicación se definieron los roles y permisos para el MD.

Capítulo 3: Implementación y prueba del MD Civil-Administrativo

Introducción

En este capítulo se hace referencia a la implementación de la solución, abordando específicamente cómo se realiza la implementación de los subsistemas de almacenamiento, de integración y de visualización del área Civil-Administrativo, teniendo en cuenta los requisitos y necesidades del negocio. De igual forma, hace referencia a las pruebas mediante la utilización de las listas de chequeo y los casos de pruebas, basados en casos de usos y reglas de transformación, para validar los reportes del MD y el cumplimiento de las reglas del negocio respectivamente.

3.1 Implementación del subsistema de almacenamiento

En la implementación del subsistema de almacenamiento se definen los estándares de codificación para lograr una mejor organización de las estructuras que componen el AD y además se desarrolla la estructura física del modelo de datos.

3.1.1 Implementación del modelo de datos físico

Para el desarrollo del MD se definieron tres esquemas: el esquema dimensiones, el cual contiene las dimensiones compartidas con los demás MD pertenecientes al Sistema de Información Estadística Judicial para los Tribunales Populares, el esquema mart_civ_administ el cual está compuesto por las tablas de dimensiones y las tablas de hechos correspondientes al MD y el esquema metadatos el cual contiene las tablas que permiten el control de los cambios en el MD y los errores encontrados.

Tabla 6: Relación de las tablas por esquemas del MD Civil-Administrativo.

| dimensiones | mart_civ_administ | metadatos |
|-----------------|------------------------------|------------|
| •13 dimensiones | •26 dimensiones •9 hechos | •13 tablas |

3.1.2 Estándares de codificación

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

Tabla 7: Estándares de codificación del mercado de datos Civil-Administrativo.

| Tipo de objeto | Función | Nomenclatura | Descripción |
|----------------|-------------------------|-----------------|---|
| Esquemas | Dimensiones compartidas | dimensiones | Esquema donde se organizan las dimensiones compartidas por varios MD. |
| | Esquemas de datos | mart_[temática] | Esquema donde se almacenan las tablas de dimensiones y hechos. |
| Tablas | Dimensiones | dim_[nombre] | Tablas dimensionales utilizadas como perspectivas de análisis. |
| | Hechos | hech_[nombre] | Tablas de hechos que definen las principales medidas. |
| | Metadatos | md_[nombre] | Tablas de metadatos donde se guarda la información correspondiente a los datos del mercado. |

3.2 Implementación del subsistema de integración

La implementación del subsistema de integración implica que se lleven a cabo los procesos de extracción, transformación y carga. No es recomendable iniciar los procesos de integración de datos sin haber realizado un análisis previo de las fuentes de datos, las cuales consisten en sistemas fuentes dispersos que contienen la información histórica. Dicha información, en el caso del sistema en cuestión, se encuentra en archivos ".dbf".

Luego de realizado el análisis de los sistemas fuentes, se extraen los datos seleccionando los registros y campos relevantes para el MD Civil-Administrativo, teniendo en cuenta el modelo de datos realizado. La etapa de transformación y limpieza provee una información lista para ser cargada en la BD. Con la

limpieza se detectan los datos erróneos, además de las entradas duplicadas y con las transformaciones se combinan los datos.

Durante la carga de las dimensiones y hechos los datos son transmitidos del modelo temporal hacia el MD Civil-Administrativo. La información cargada ha sido depurada y será utilizada para la generación de los reportes que serán analizados en función de la toma de decisiones.

Seguidamente se describen algunos de los subsistemas de integración identificados por Kimball y que son utilizados en el desarrollo de la solución, los cuales se encuentran en cuatro grupos, que se describen a continuación (35):

Extracción:

- ✓ **Perfilado de datos:** permite explorar los datos para verificar su calidad y el cumplimiento de los estándares conforme a los requisitos especificados por el cliente. Mediante el uso de este subsistema fueron definidas nuevas reglas de transformación.
- ✓ **Sistema de extracción:** permite la extracción de los datos desde la fuente de origen para su transformación y posterior carga. Para ello se tuvo en cuenta la información relacionada con cada uno de los hechos y dimensiones.
- ✓ **Captura de cambio en los datos:** permite detectar los cambios asociados a las fechas en que se realizará la carga de la información.

Limpieza y conformación:

- ✓ **Rastreo de eventos de errores:** posibilita la captura de los errores que proporcionan información valiosa sobre la calidad de los datos y permiten la mejora de los mismos.

Entrega:

- ✓ **Cambio lento de las dimensiones (SCD):** implementa la lógica para crear atributos de variabilidad lenta a lo largo del tiempo.
- ✓ **Llave subrogada:** permite crear claves subrogadas independientes para cada tabla.
- ✓ **Creador de cubos OLAP:** permite alimentar con datos a esquemas OLAP desde esquemas dimensionales relacionales.

Gestión:

- ✓ **Programador de trabajos:** permite gestionar los trabajos, los cuales se encargan de la ejecución de las transformaciones en un orden específico y atendiendo a la periodicidad definida para la carga de la información.
- ✓ **Control de versiones:** permite hacer control de versiones del proyecto ETL y de los metadatos asociados.
- ✓ **Repositorio de metadatos:** captura los metadatos de los procesos de ETL, de los datos de negocio y de los aspectos técnicos.

3.2.1 Implementación de las transformaciones

Una transformación es el elemento básico de la implementación de los procesos de ETL. Esta se compone de pasos entrelazados entre sí a través de los saltos, mediante los cuales va fluyendo la información.

A continuación se muestra el flujo de transformación para la carga del hecho: Indicadores asuntos de familia provincial, pertenecientes al esquema mart_civ_administ.

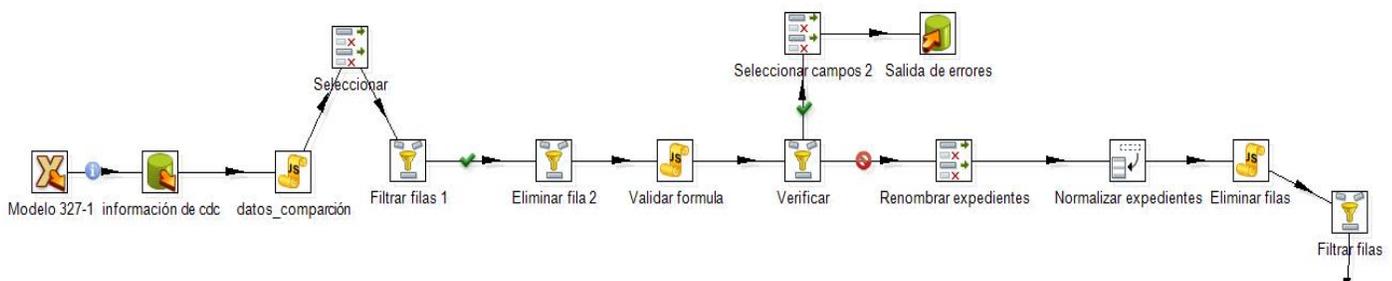


Figura 13: Representación de la carga del hecho indicadores asuntos de familia provincial, parte 1.

Como se puede apreciar en la figura 15, el primer paso de la transformación es una entrada .dbf del modelo 3271. Después de realizar la extracción de los datos se procede a las transformaciones. Primeramente se busca en la tabla de metadatos.md_cdc el nombre del MD, el nombre del hecho y la última fecha de actualización. Posteriormente se seleccionan todos los campos y seguidamente se verifica que la información que se desee transformar no se encuentre ya insertada en el MD. Solo se realizarán las transformaciones a los datos a partir del último año y mes insertado.

Se elimina del campo FILA el número 4, pues este no posee información y no es necesario tenerlo en cuenta. Se validan RN que fueron detectadas en la entrevista con el cliente, en caso de que no se

cumplan toda la información será insertada en una tabla de errores para su posterior tratamiento. Además se seleccionan los campos y se renombran los expedientes de acuerdo al código que posee la dimensión `dim_expedientes` en la BD. Se realiza la normalización de los expedientes con el objetivo de que los datos se encuentren en una única columna. Se realiza la eliminación de filas y columnas que no contienen información.

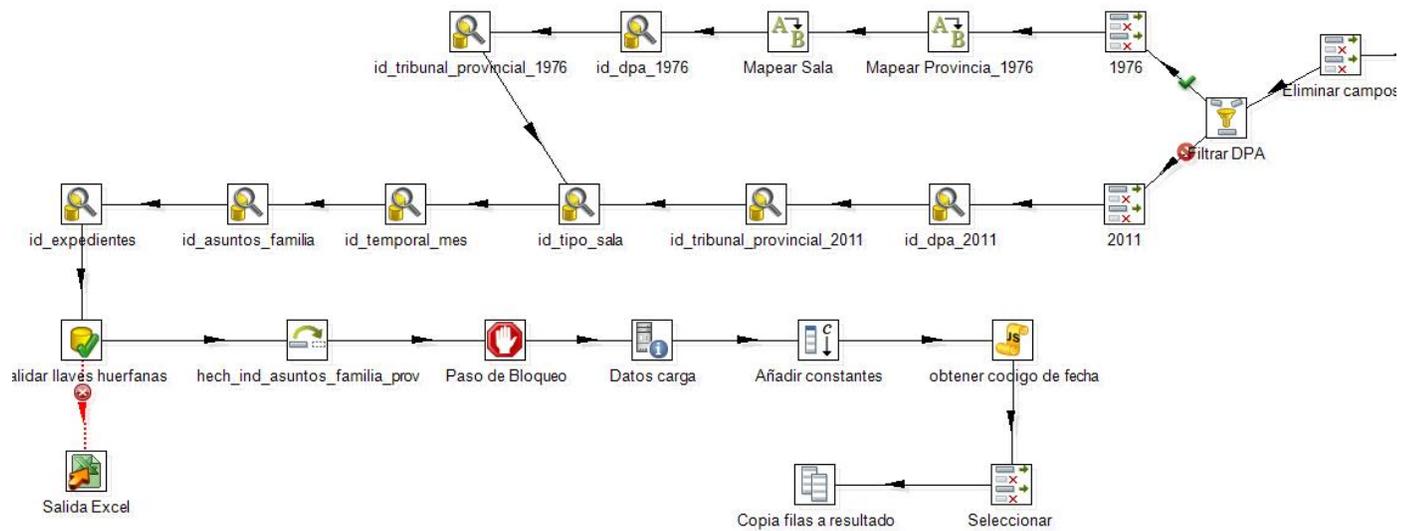


Figura 14: Representación de la carga del hecho indicadores asuntos de familia provincial, parte 2.

Posteriormente se divide el flujo en dos partes, si el año es menor que 2011 se mapean las provincias y las salas de acuerdo al código que tenían antes del año 2011. Seguidamente se buscan los identificadores en cada una de las tablas de dimensiones, en caso que el año sea mayor o igual que 2011 se buscan los identificadores como se realizó anteriormente. Después se validan los tipos de datos y se verifica que los identificadores no se encuentren nulos, en caso de que no se cumpla con estos requisitos se enviarán los datos hacia un excel de error. Se insertan los datos con el componente Insertar/Actualizar y por último se guarda la información referente a la transformación ejecutada (`fecha_hora`, `transformación`, `nombre del hecho` e `IP`), se añade el modelo insertado y la frecuencia de la carga.

3.2.2 Implementación de los trabajos

Un trabajo es un conjunto de transformaciones que llevan a cabo la extracción física, transformación, y las rutinas de carga (36).

En la presente investigación se realizaron tres trabajos, uno destinado a ejecutar las transformaciones para la carga de las 26 dimensiones (ver figura 17), otro para ejecutar las transformaciones para la carga de los nueve hechos (ver figura 18) y el general que ejecuta los dos trabajos anteriores.

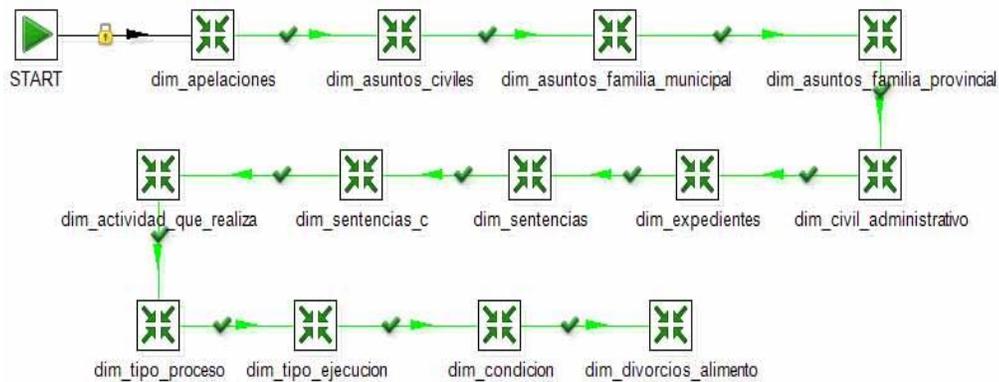


Figura 15: Trabajo que permite la ejecución de las transformaciones para la carga de las dimensiones.

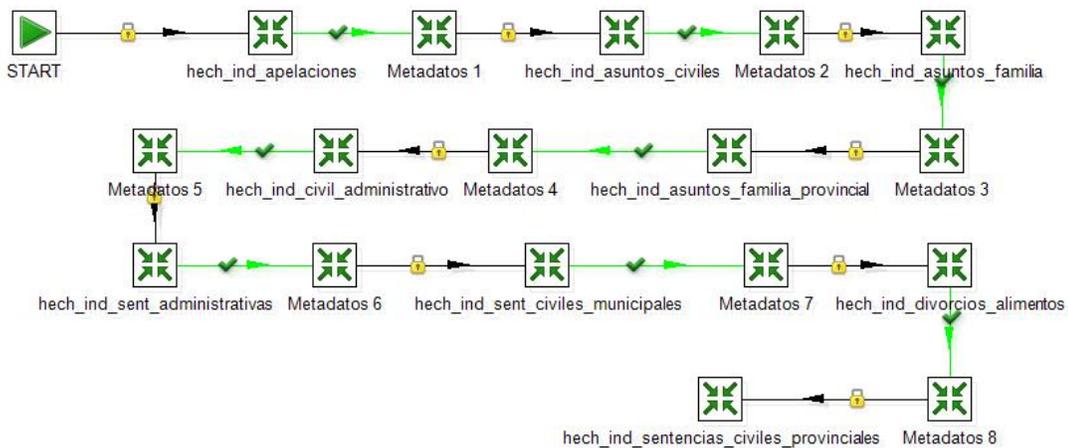


Figura 16: Trabajo que permite la ejecución de las transformaciones para la carga de los hechos.

3.2.3 Gestión del cambio lento en las dimensiones

Las dimensiones lentamente cambiantes o SCD¹¹ son dimensiones en las cuales los datos tienden a modificarse a través del tiempo, ya sea de forma ocasional o constante, que implique a un solo registro o a la tabla completa. Cuando ocurren estos cambios se puede optar por seguir alguna de estas dos opciones:

- ✓ Registrar el historial de cambios.
- ✓ Reemplazar los valores que sean necesarios.

Inicialmente Ralph Kimball plantea tres estrategias a seguir para el tratamiento de las SCD: tipo 1, tipo 2 y tipo 3.

Es importante señalar que si bien hay diferentes maneras de implementar cada técnica, es indispensable contar con llaves subrogadas en las tablas de dimensiones para poder aplicarlas (37).

- ✓ **Tipo 1** (sobrescribir): este tipo es el más básico y sencillo de implementar ya que si bien no guarda los cambios históricos tampoco requiere ningún modelado especial y no necesita que se añadan nuevos registros a la tabla. En este caso cuando un registro presente un cambio en alguno de los valores de sus campos, se debe proceder simplemente a actualizar el dato en cuestión, sobrescribiendo el antiguo. Usualmente este tipo es utilizado en casos en donde la información histórica no sea importante de mantener.
- ✓ **Tipo 2** (añadir fila): esta estrategia requiere que se agreguen algunas columnas adicionales a la tabla de dimensión, para que almacenen el historial de cambios. Las columnas que suelen agregarse son: fecha de inicio, donde se almacena la fecha en que entró en vigencia el registro actual, fecha fin en la cual se guarda la fecha en que dejó de estar en vigencia el registro actual, versión en la cual se almacena un número secuencial que se incrementa cada nuevo cambio y la versión actual en la que se especifica si el campo actual es vigente (0 si es falso y 1 si es verdadero). Entonces cuando ocurra algún cambio en los valores de los registros se añadirá una nueva fila y se deberán completar los datos referidos al historial de cambios.
- ✓ **Tipo 3** (añadir columna): esta estrategia requiere que se agregue a la tabla de dimensión una columna adicional por cada columna cuyos valores se desea mantener. De este modo en la nueva

¹¹ Slowly Changing Dimension

columna se coloca el valor antiguo antes de sobrescribir con el nuevo valor. Esta técnica permite guardar un historial limitado de los datos, dependiendo del número de columnas que se creen.

Debido a la ocurrencia de cambios en la `dpa_provincia`, `dpa_municipio`, `dim_tribunal_provincial` y `dim_tribunal_municipal` en los años anteriores y posteriores al 2011, se decide implementar el SCD de tipo 3 para poder realizar análisis entre los datos correspondientes a la etapa anterior y la actual. Luego del cambio de la `dim_dpa_municipio`, la dimensión `dim_tipo_sala` se ve afectada debido a que el código de la misma no se corresponde con el código de los municipios, pues anteriormente los códigos de las salas se relacionaban con los de los municipios. En este caso se implementa el SCD de tipo 2 para mantener la información histórica.

3.2.4 Gestión de los metadatos del proceso de integración

La palabra metadato, (del griego *meta* "después de", y latín *datum* "dato") significa "sobre datos" es decir, datos acerca de los datos. Cuando hablamos de metadatos en realidad hablamos sobre conocimiento del sistema, conocimiento del negocio y conocimiento del MD. Principalmente hay tres tipos de metadatos (38).

- ✓ **Metadatos de negocio:** posibilita obtener los datos y la información referente a los aspectos del negocio, como son los datos provenientes de la fuente.
- ✓ **Metadatos técnicos:** están relacionados con la función de un sistema o el modo en que interrelacionan sus componentes.
- ✓ **Metadatos de proceso:** permiten obtener información de los procesos que se ejecutan.

En la investigación se utilizaron los metadatos de proceso para obtener la información correspondiente a los procesos de las transformaciones y los trabajos de los subprocesos de ETL. Además se utilizaron los metadatos técnicos para el control de cambio de la fuente (CDC).

3.3 Implementación del subsistema de visualización

Una vez efectuada la carga de los datos se procede a realizar la implementación del subsistema de visualización, donde se implementan los cubos OLAP, las vistas de análisis y los reportes operacionales, además de configurarse el control de acceso al sistema.

3.3.1 Implementación de los cubos OLAP

Los cubos proveen un mecanismo para buscar datos con rapidez independientemente de la cantidad de datos en el cubo o la complejidad del procedimiento de búsqueda. La información contenida son las dimensiones, jerarquías, niveles y medidas en una estructura multidimensional (39).

En correspondencia con las tablas de hechos definidas para el área Civil-Administrativa se crearon nueve cubos (ver figura 19). El esquema cuenta con 26 dimensiones y nueve medidas que dan respuesta a los requisitos del cliente.

A continuación se describen los cubos con sus dimensiones y medidas:

- ✓ Cubo asuntos de familia municipal: cuenta con cinco dimensiones y una medida.
- ✓ Cubo asuntos de familia provincial: cuenta con seis dimensiones y una medida.
- ✓ Cubo civil y administrativo: cuenta con seis dimensiones y una medida.
- ✓ Cubo apelaciones: cuenta con seis dimensiones y una medida.
- ✓ Cubo sentencias administrativas: cuenta con cinco dimensiones y una medida.
- ✓ Cubo sentencias civiles municipales: cuenta con cinco dimensiones y una medida.
- ✓ Cubo sentencias civiles provinciales: cuenta con seis dimensiones y una medida.
- ✓ Cubo asuntos civiles: cuenta con cinco dimensiones y una medida.
- ✓ Cubo divorcios y alimentos: cuenta con seis dimensiones y una medida.



Figura 17: Representación de los hechos del MD Civil-Administrativo. Relación del hecho indicadores asuntos de familia provincial con sus dimensiones y su medida.

3.3.2 Arquitectura de la información

La capa de visualización quedó estructurada acorde a la arquitectura propuesta en el mapa de navegación. El A.A. Civil-Administrativo contenida en el A.A.G. Sistema de Información Estadística Judicial para los Tribunales Populares (SIEJT), está compuesta por el A.A. Instancia municipal y el A.A. Instancia provincial, dichas áreas contienen el A.A. DPA (1976-2010) y el A.A. DPA (2011-Actualidad); las cuales están formadas por varios L.T. en los que se encuentran los 36 reportes o tablas de salida implementados (ver figura 20).

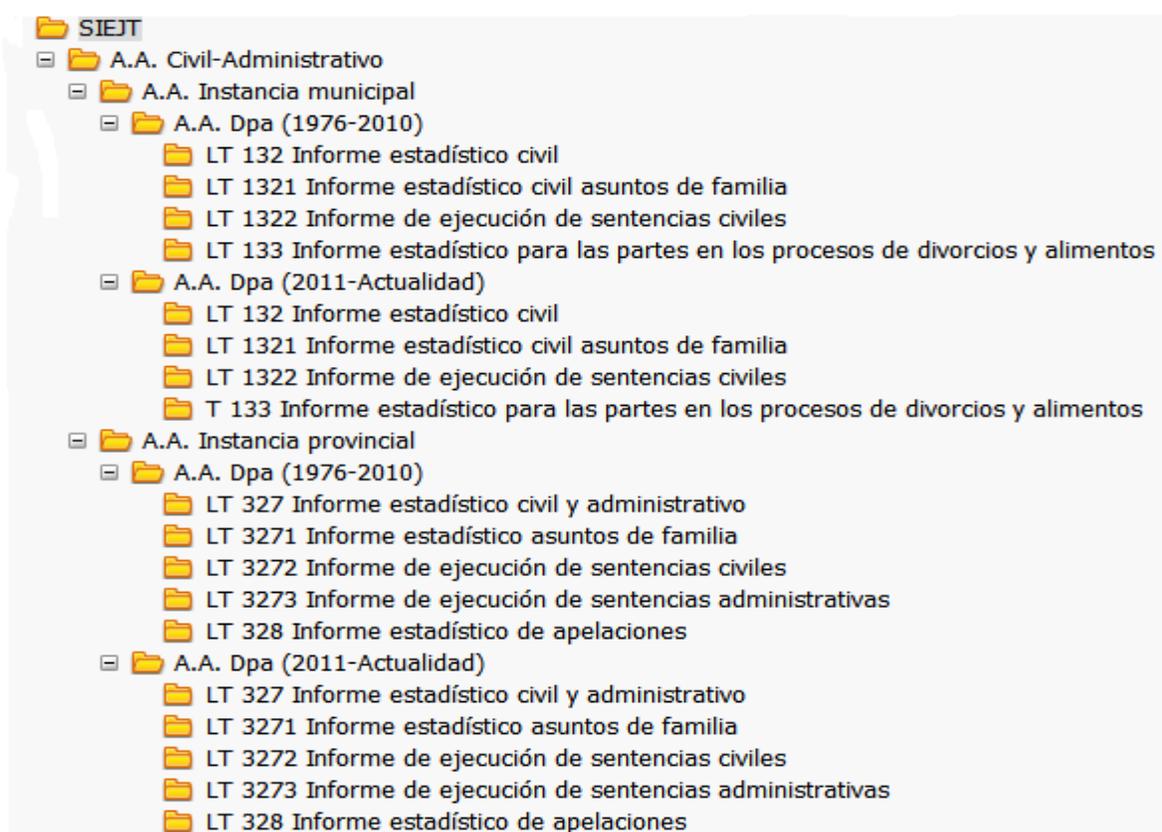


Figura 18: Arquitectura de información del MD Civil-Administrativo.

A continuación se presenta la vista de análisis del resumen nacional por provincias de asuntos de familia provincial que se encuentra en el L.T. 3271 Informe estadístico asuntos de familia. Donde se identifican las dimensiones temporal_mes, expedientes y dpa_municipio.

| Provincias | Tiempo | | | | | | | | | |
|---------------------|----------------------|-----------|-----------------------|-------------|------------------|-----------------|---------------|---------|-------|-------------|
| | 2011 | | | | | | | | | |
| | Expedientes | | | | | | | | | |
| | Pendientes al inicio | Radicados | Procesos retrotraídos | Nuevo curso | Total a resolver | Total resueltos | Desistimiento | Archivo | Otros | Total Autos |
| Pinar del Río | 19 | 51 | 0 | 0 | 70 | 51 | 2 | 0 | 11 | 13 |
| Artemisa | 32 | 31 | 0 | 0 | 63 | 29 | 1 | 3 | 2 | 6 |
| La Habana | 428 | 231 | 3 | 0 | 662 | 215 | 0 | 3 | 43 | 46 |
| Mayabeque | 94 | 21 | 0 | 15 | 130 | 30 | 0 | 2 | 7 | 9 |
| Matanzas | 66 | 87 | 0 | 0 | 153 | 89 | 0 | 0 | 8 | 8 |
| Villa Clara | 36 | 92 | 0 | 0 | 128 | 92 | 0 | 0 | 21 | 21 |
| Cienfuegos | 45 | 46 | 3 | 0 | 94 | 50 | 2 | 0 | 7 | 9 |
| Sancti Spiritus | 26 | 57 | 0 | 0 | 83 | 55 | 1 | 0 | 5 | 6 |
| Ciego de Ávila | 30 | 12 | 0 | 0 | 42 | 10 | 0 | 2 | 0 | 2 |
| Camagüey | 45 | 103 | 0 | 0 | 148 | 103 | 0 | 0 | 13 | 13 |
| Las Tunas | 19 | 24 | 0 | 0 | 43 | 24 | 0 | 1 | 0 | 1 |
| Holguín | 55 | 103 | 0 | 0 | 158 | 107 | 0 | 0 | 17 | 17 |
| Granma | 46 | 91 | 0 | 0 | 137 | 91 | 1 | 7 | 3 | 11 |
| Santiago de Cuba | 86 | 182 | 3 | 0 | 271 | 187 | 0 | 5 | 35 | 40 |
| Guantánamo | 21 | 30 | 0 | 0 | 51 | 24 | 0 | 0 | 3 | 3 |
| Isla de la Juventud | 26 | 15 | 0 | 0 | 41 | 13 | 2 | 0 | 2 | 4 |

Figura 19: Representación de una vista de análisis del resumen nacional por provincias de los asuntos de familia provinciales.

3.4 Implementación de la seguridad de los usuarios

La implementación del subsistema de visualización culmina con la configuración de los roles a los usuarios para determinar los permisos de acceso a la información, por lo que se definieron dos roles y dos usuarios, a continuación se describen cada uno de ellos:

- ✓ Rol administrador: tiene control total sobre la aplicación, mediante el usuario asignado el cual es Administrador.
- ✓ Rol analista: solo tiene permiso de lectura sobre la aplicación, mediante el usuario asignado el cual es el Analista.

3.5 Pruebas aplicadas al MD Civil-Administrativo

Implementado el MD Civil-Administrativo se pasa a la etapa de pruebas. Las pruebas de software son procesos que permiten verificar y revelar la calidad de un determinado producto. Son utilizadas para identificar posibles fallos de implementación y calidad en la solución (40). Los resultados obtenidos se captan para ser observados y registrados realizando evaluaciones de algún aspecto en específico. Las pruebas de software son un elemento con carácter crítico en cuanto a la calidad del producto y figura como la revisión final de las especificaciones del diseño y la implementación.

La prueba es un proceso de ejecución de un programa con la intención de descubrir un error. Un buen caso de prueba es aquel que tiene una alta probabilidad de mostrar un error no descubierto hasta el momento (41).

3.6 Pruebas de calidad de software

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

- ✓ **Pruebas unitarias:** permiten probar el correcto funcionamiento de un componente o subsistema específico y son desarrollados por los propios desarrolladores durante la implementación (10).
- ✓ **Pruebas de integración:** permite verificar la correcta integración de los componentes y subsistemas que conforman la solución. Pone a prueba la vista arquitectónica del sistema definida en una infraestructura de desarrollo. Estas pruebas son ejecutadas por los arquitectos de software (10).
- ✓ **Pruebas de sistema (validación):** permiten comprobar el cumplimiento de los requisitos de información y funcionales definidos por los clientes. Son las pruebas más cercanas a la realidad del cliente debido a que los probadores utilizan el sistema de la misma manera que será usado por los clientes (10).

En las pruebas de sistema es donde generalmente se encuentra la mayor cantidad de errores, por ello se realizan a diferentes instancias con el objetivo de entregar al cliente un producto con mayor calidad. Seguidamente se presentan las pruebas que se realizan.

- ✓ **Pruebas internas:** son aplicadas por el grupo de calidad interno del proyecto y por el equipo de desarrollo (10).
- ✓ **Pruebas de liberación:** son realizadas primeramente el grupo de calidad interno del proyecto y luego por los especialistas de CALISOFT (10).
- ✓ **Pruebas de aceptación:** estas pruebas son realizadas por el cliente para verificar que se cumple con los requisitos planteados por el mismo y validar su conformidad con el producto (10).

3.7 Herramientas de pruebas

Para la aplicación de las pruebas fueron utilizadas las siguientes herramientas:

Casos de prueba

Mediante los casos de prueba, el probador podrá determinar si el requisito de una aplicación es parcial o completamente satisfactorio. En la presente investigación para probar la aplicación se utilizaron los casos de pruebas basados en CUI, en el Anexo 2 se encuentra el caso de prueba Indicadores asuntos de familia. A continuación se mencionan los casos de prueba realizados:

- ✓ Caso de prueba Indicadores asuntos de familia.
- ✓ Caso de prueba Indicadores asuntos de civiles.
- ✓ Caso de prueba Indicadores divorcios y alimentos.
- ✓ Caso de prueba Indicadores sentencias civiles.
- ✓ Caso de prueba Indicadores sentencias administrativas.
- ✓ Caso de prueba Indicadores civil y administrativo.
- ✓ Caso de prueba Indicadores de apelaciones.

En el expediente de proyecto del MD Civil-Administrativo se pueden encontrar los siete casos de prueba diseñados, en el artefacto 0111_Diseño de caso de prueba basado en CU.

Lista de chequeo

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

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

Evaluación del resultado de la lista de chequeo:

Se aborta el proceso de aplicación de la lista en caso de:

- ✓ Existan al menos dos indicadores críticos evaluados de mal
- ✓ Más del 50% de los indicadores a evaluar están evaluados de mal.

Se mantienen las no conformidades de una revisión a otra.

- ✓ Se evalúa de regular la calidad de los artefactos de ETL en caso de:
- ✓ Incumple con los indicadores críticos a evaluar de las secciones Estructura del documento y Semántica del documento de la lista de chequeo.
- ✓ Existe al menos un indicador crítico evaluado de mal.
- ✓ Existen al menos cinco indicadores no críticos evaluados de mal.
- ✓ Los artefactos de los procesos ETL son evaluados de bien cuando no cumple con ningún criterio de los dos puntos anteriores.

Luego de establecida la estructura de la lista de chequeo, esta es aplicada a los artefactos de los procesos de ETL:

- ✓ Lista de chequeo del mapa lógico de datos.
- ✓ Lista de chequeo de registro de sistema fuente.
- ✓ Lista de chequeo del perfilado de datos.
- ✓ Lista de chequeo del diccionario de datos.

3.8 Resultados de las pruebas

Las pruebas unitarias y de integración se aplicaron a los subsistemas que componen el MD identificando las siguientes No Conformidades (NC):

Pruebas unitarias y de integración

En el subsistema de almacenamiento se identificó una NC con complejidad media, la cual fue resuelta:

NC1. La dimensión dim_expedientes contiene tuplas con faltas de ortografía.

En el subsistema de integración se identificó una NC con complejidad alta, la cual fue resuelta y se muestra a continuación:

NC2. Los datos correspondientes a los reportes candidatos están duplicados.

En el subsistema de visualización se identificaron cinco NC. De ellas, dos NC con complejidad alta y tres NC con complejidad media, las cuales fueron resueltas en su totalidad. A continuación se exponen las NC detectadas:

NC1. Falta por calcular los totales de determinados campos.

NC2. Existen columnas en los reportes que no están en el orden especificado.

NC3. Falta por calcular el porcentaje de determinados campos en aquellos reportes que lo necesiten.

NC4. Corregir el orden de los LT según la materia a analizar.

NC5. Separar por A.A los LT pertenecientes a los informes de instancia municipal e instancia provincial.

Pruebas de sistema

Las pruebas se aplicaron por parte del grupo de calidad interno del proyecto y se admite una carta de liberación interna. Los siete casos de prueba basados en CU permitieron identificar un total de tres NC de complejidad media, las cuales fueron resueltas en su totalidad. A continuación se muestran las NC detectadas en las pruebas de sistema:

NC1. Uso excesivo de mayúscula en todos LT.

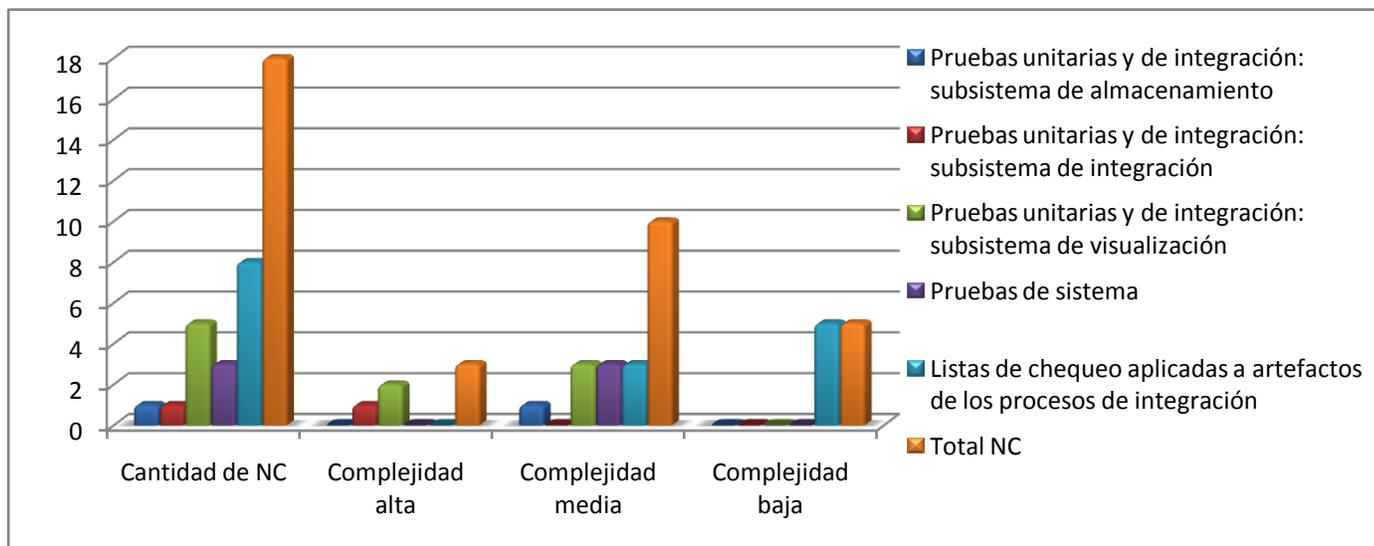
NC2. El flujo central de la descripción del caso de prueba está incorrecto.

NC3. En todas las descripciones de los casos de prueba las variables de entrada no coinciden con lo que muestra el sistema.

Evaluación de las pruebas de aceptación

El MD fue presentado por el equipo al especialista del área, que pudo comprobar que las funcionalidades implementadas satisfacen sus necesidades, emitiendo la carta de aceptación la cual valida que el sistema está listo para ser desplegado (Anexo 2). La siguiente gráfica muestra el resultado general de las pruebas aplicadas.

Tabla 8: Resultado general de las pruebas aplicadas al MD Civil-Administrativo según su complejidad.



Conclusiones del capítulo

Después de la implementación y prueba del MD Civil-Administrativo se arribaron a las siguientes conclusiones:

Quedó implementado el subsistema de almacenamiento el cual cuenta con tres esquemas: *dimensiones*, compuesto por 13 tablas; *mart_civ_administ* que contiene 22 tablas, correspondientes a los hechos y dimensiones propios del MD y el esquema *metadatos* con seis tablas. Se implementó el subsistema de integración mediante la realización de los trabajos, uno para las dimensiones, otro para los hechos, y por último el trabajo principal. Se desarrolló el subsistema de visualización, implementando los cubos OLAP, así como la realización de 36 vistas de análisis y un reporte operacional, para dejar implementado el subsistema de visualización.

Mediante las pruebas unitarias y de integración se detectaron siete NC las cuales fueron resueltas en su totalidad. Se aplicaron listas de chequeo al MD Civil-Administrativo detectándose ocho NC, todas se resolvieron satisfactoriamente.

Se realizaron las pruebas de aceptación, lo que permitió validar que la solución propuesta cumple con los requerimientos solicitados por el cliente.

Conclusiones Generales

Con la realización de la investigación se logró cumplir con los objetivos planteados, desarrollándose una solución que ofrece soporte a la toma de decisiones mediante un mercado de datos para el área Civil-Administrativo para el Tribunal Supremo Popular. Por lo que se concluye que:

- 1 El estudio de los fundamentos teóricos de la investigación, permitió seleccionar como metodología, la Propuesta de metodología para el desarrollo de los almacenes de datos en DATEC, la que permitió organizar de manera estructurada el proceso de desarrollo de software, así como las herramientas que cuentan con las características necesarias para desarrollar la solución propuesta.
- 2 El análisis y diseño del mercado de datos, posibilitó un mejor entendimiento del proceso de negocio del área Civil-Administrativo, permitiendo la implementación de una solución que responde a las necesidades del cliente.
- 3 La implementación del mercado de datos contribuye al proceso de apoyo a la toma de decisiones mediante el desarrollo de los diferentes subsistemas, garantizado la correcta organización, carga y visualización de los datos.
- 4 La aplicación de los casos de pruebas, el perfilado de los datos y las listas de chequeo permitió obtener un mercado funcional que cuenta con una correcta calidad de los datos y que cumple con los requisitos especificados por el cliente.

Recomendaciones

- Incorporar a los reportes y a las vistas de análisis del sistema gráficos, lo cual permitirá al usuario una mejor percepción de los datos solicitados para el proceso de toma de decisiones.

Referencias bibliográficas

1. **INC.** Management Information Systems (MIS). [En línea] [Citado el: 20 de abril de 2013.] <http://www.inc.com/encyclopedia/management-information-systems-mis.html>.
2. **SlideShare Inc.** Slideshare. [En línea] [Citado el: 20 de abril de 2013.] <http://www.slideshare.net/heyman/sistemas-de-apoyo-a-la-toma-de-decisiones-12905870>.
3. **Slideshare Inc.** Slideshare. [En línea] [Citado el: 20 de abril de 2013.] <http://www.slideshare.net/kdcsb/componentes-de-un-dss-presentacion-2>.
4. **Velasco, Roberto Hernando.** [En línea] [Citado el: mayo de 12 de 2012.] <http://www2.rhernando.net/modules/tutorials/doc/bd/dw.html>.
5. **Inmon, Willian H.** *¿What is a Datawarehouse?*
6. **Sánchez, Leopoldo Zenaido Cepeda.** Metodología para el Diseño Conceptual de los Almacenes de Datos. [En línea] 2008. <http://riunet.upv.es/bitstream/handle/10251/2506/tesisUPV2841.pdf>.
7. **ZRox Technology.** [En línea] 23 de septiembre de 2011. [Citado el: 2012 de diciembre de 8.] <http://compuroo.blogspot.com/2011/11/data-mart-mercado-de-datos.html>.
8. **Arley, Ricardo Chinchilla.** Mercado de datos: conceptos y metodologías de desarrollo. [En línea] 2010. <http://www.kerwa.ucr.ac.cr/bitstream/handle/10669/619/TM%2024-3%20art%206.pdf?sequence=1>.
9. **Romero, Hermenegildo.** [En línea] 2009. [Citado el: 5 de diciembre de 2012.] <http://www.slideshare.net/MeneRomero/metodologias-de-desarrollo>.
10. **Hernández, Yanisbel González.** *PROPUESTA DE METODOLOGIA PARA EL DASARROLLO DE ALMACENES DE DATOS.* Centro de Tecnologías de Gestión de Datos (DATEC), facultad 6, UCI, La Habana, Cuba : s.n., Octubre 2012.
11. Guión Visual Paradigm for UML. [En línea] 2011. <http://www.ie.inf.uc3m.es/grupo/docencia/reglada/ls1y2/PracticaVP.pdf>.
12. **Altamirano, Ing. Alfonso Valdez.** *Comparativo de Entornos de Desarrollo Integrados (IDE's).*
13. **Instituto Tecnológico de Colima.** Estructura general de un sistema de base de datos. [En línea] [Citado el: 5 de diciembre de 2012.] http://labredes.itcolima.edu.mx/fundamentosbd/sd_u1_5.htm.
14. **Antonio Aliaga Ibarra, Marcos Agustin Miani Flores.** PostgreSQL. [En línea] febrero de 2008. <https://iessanvicente.com/colaboraciones/postgreSQL.pdf>.
15. **PostgreSQL.** PostgreSQL. [En línea] [Citado el: 5 de diciembre de 2012.] <http://postgresql-dbms.blogspot.com/p/limitaciones-puntos-de-recuperacion.html>.

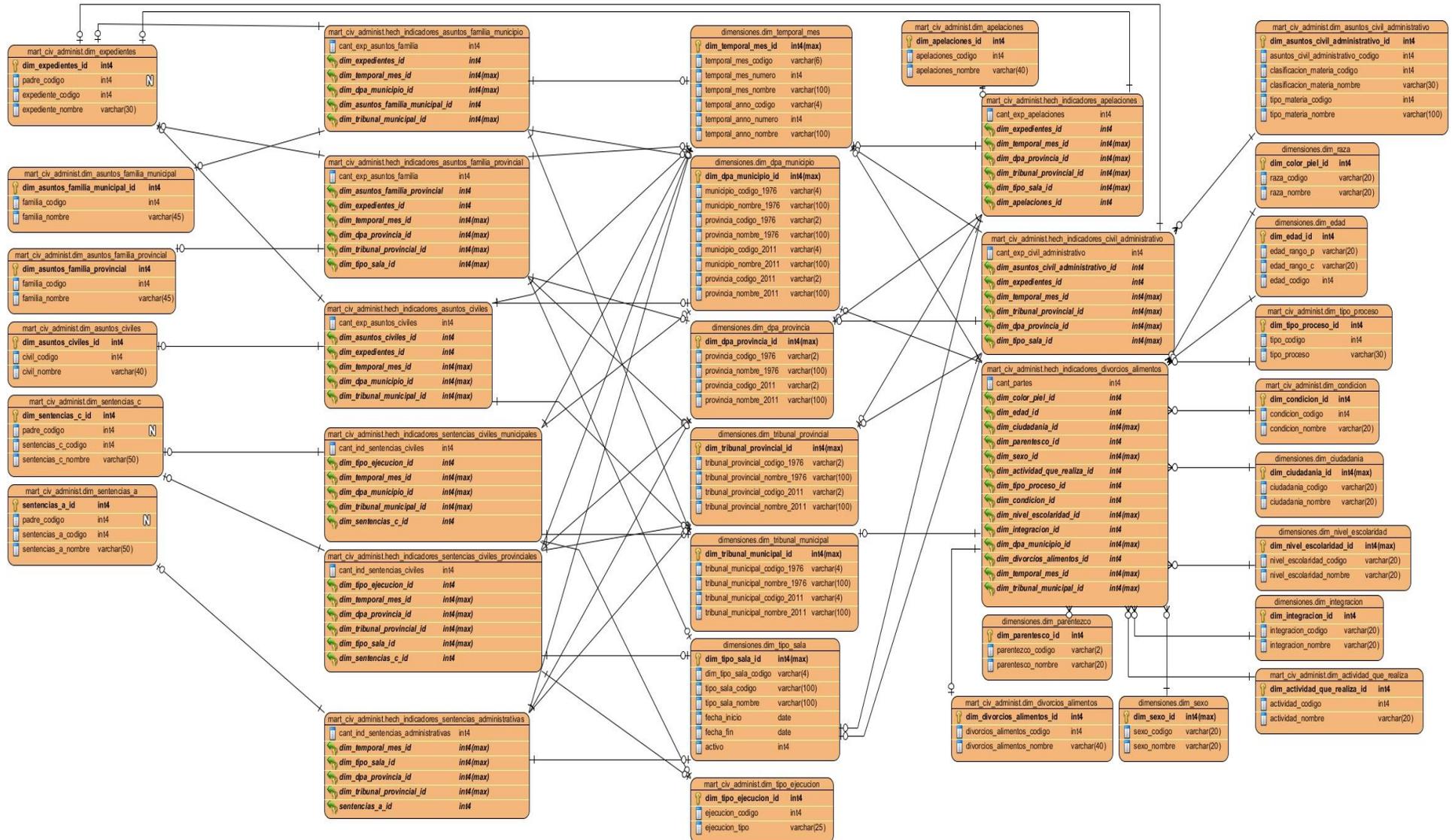
16. **MASTERMAGAZINE.** [En línea] [Citado el: 5 de diciembre de 2012.] <http://www.mastermagazine.info/termino/6841.php#ixzz2UJyOOF00>.
17. **Carlos Fernández.** [En línea] [Citado el: 5 de diciembre de 2012.] <http://www.dataprix.com/olap-rolap-molap>.
18. **Arley, Ricardo Chinchilla.** *Mercado de datos: conceptos y metodologías de desarrollo.* 2010.
19. **Ibarra, María de los Ángeles.** Procesamiento Análítico en Línea (OLAP). [En línea] 2010. <http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/OLAPMonog.pdf>.
20. —. *Procesamiento Análítico en Línea (OLAP).* Argentina : s.n., 2010.
21. **DataCleaner.** [En línea] 2012. [Citado el: 8 de diciembre de 2012.] <http://datacleaner.org/>.
22. **Pentaho.** Pentaho Data Integration (Kettle). [En línea] [Citado el: 6 de diciembre de 2012.] <http://kettle.pentaho.com/>.
23. **Bldv. Diaz Ordaz.** Gravatar. [En línea] 2012. [Citado el: 8 de diciembre de 2012.] <http://www.gravatar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.
24. **Procibernetica.** [En línea] [Citado el: 6 de diciembre de 2012.] <http://www.procibernetica.com/inteligencia-negocios.html>.
25. **Jaspersoft.** OLAP Workbench Guide. [En línea] [Citado el: 6 de diciembre de 2012.] <http://community.jaspersoft.com/sites/default/files/docs/olap-workbench-user-guide.pdf>.
26. **Sherman Wood, JasperSoft.** Mondrian Documentation. [En línea] 2009. [Citado el: 8 de diciembre de 2012.] <http://mondrian.pentaho.com/documentation/workbench.php>.
27. **Pentaho.** Pentaho BI Pplatform. [En línea] [Citado el: 6 de diciembre de 2012.] http://community.pentaho.com/projects/bi_platform/.
28. —. Pentaho Report Designer. [En línea] [Citado el: 25 de abril de 2013.] http://reporting.pentaho.com/report_designer.php.
29. —. Pentaho Reporting. [En línea] [Citado el: 25 de abril de 2013.] <http://reporting.pentaho.com/>.
30. **Pressman, Roger S.** *Ingeniería del software: un enfoque práctico.* MacGraw-Hill : s.n., 2005.
31. **Ceria, Santiago.** Ingeniería de Software 1: Casos de uso, un método práctico para explorar requerimientos. [En línea] Universidad de Buenos Aires. http://www-2.dc.uba.ar/materias/isoft1/2001_2/apuntes/CasosDeUso.pdf.

32. **Gerardo, Clemente García.** Tesis doctoral: Un Sistema para el Mantenimiento de Almacenes de Datos. [En línea] UNIVERSIDAD POLITÉCNICA DE VALENCIA, Valencia, España, Junio de 2008. <http://riunet.upv.es/bitstream/handle/10251/2505/tesisUPV2842.pdf>.
33. **Chuck Ballard, Amit Gupta, Vijaya Krishnan, Nelson Pessoa, Olaf Stephan.** *Data Mart Consolidation: Getting Control of Your Enterprise Information.* July 2005. SG24-6653-00.
34. **Kimball, Ralph.** *The Data Warehouse ETL Toolkit.* pag 526.
35. **Díaz, Josep Curto.** Introducción al Business Intelligence. [En línea] 6 de diciembre de 2012. <http://books.google.com/cu/books?id=iU3RAXYQXMkC&pg=PA88&lpg=PA88&dq=34+subsistemas+ETL+Kimball&source=bl&ots=Nmeahp49gn&sig=8V-ZaL5aO3Q60QA5ZCa4WpblhWo&hl=es&sa=X&ei=b7yoUY-IA57ltQa36IHwDA&ved=0CC4Q6AEwAQ>. pag 88.
36. **Ralph Kimball, Joe Caserta.** *The Data Warehouse ETL Toolkit: practical techniques for extracting, cleaning, conforming, and .* 2004.
37. **Kimball, R. y Ross, M.** *The Data Warehouse Toolkit: the Complete Guide to Dimensional Modelling.* New York, EE.UU: John Wiley & Sons : s.n., 2002.
38. **Albuerne, Juan Luis Martínez.** *Diseño del catálogo de metadatos para la automatización del proceso de carga de un DWH.* UNIVERSIDAD POLITÉCNICA DE MADRID : s.n., abril 2008.
39. **Juan David Peña Rivera, Jesús Armando Suárez Daza.** *Utilización de información histórica para decisiones empresariales.* Pontificia Universidad Javeriana : s.n., 2008.
40. PRUEBAS DE SOFTWARE. [En línea] 2009. [Citado el: 6 de mayo de 2013.] <http://www.slideshare.net/aracelij/pruebas-de-software>.
41. **EVA.** *Conferencia 7. Ingeniería de Software II. Sobre la disciplina de Prueba.* La Habana : Universidad de las Ciencias Informáticas. : s.n., 2010-2011.

Bibliografía

1. **Pressam, Roger S.** *Ingeniería del Software un enfoque práctico*. Madrid : Mc Graw-Hill, 2005.
2. **Ralph Kimball, Magy Ross.** *The Data Warehouse Toolkit*. s.l. : second edition , 2002.
3. **Ralph Kimball, Joe Caserta.** *Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. 2004.
4. **Ballard, Chuck.** *Data Mart Consolidation: Getting Control of Your Enterprise Information*. 2005 .
5. **Czernicki, Bart.** *Business Intelligence Software with Silverlight 3*. 2009.
6. **González, Rolando Alfredo Hernández León y Sayda Coello.** *EL PROCESO DE INVESTIGACIÓN CIENTÍFICA*. s.l. : Editorial Universitaria del Ministerio de Educación Superior, 2011.
7. **Inmon, Willian H.** *¿What is a Datawarehouse?*
8. **Sánchez, Leopoldo Zenaido Cepeda.** Metodología para el Diseño Conceptual de los Almacenes de Datos. [En línea] 2008. <http://riunet.upv.es/bitstream/handle/10251/2506/tesisUPV2841.pdf>.
9. **ZRox Technology.** [En línea] 23 de septiembre de 2011. [Citado el: 2012 de diciembre de 8.] <http://compuroo.blogspot.com/2011/11/data-mart-mercado-de-datos.html>.
10. **Arley, Ricardo Chinchilla.** Mercado de datos: conceptos y metodologías de desarrollo. [En línea] 2010. <http://www.kerwa.ucr.ac.cr/bitstream/handle/10669/619/TM%2024-3%20art%206.pdf?sequence=1>.
11. **Romero, Hermenegildo.** [En línea] 2009. [Citado el: 5 de diciembre de 2012.] <http://www.slideshare.net/MeneRomero/metodologias-de-desarrollo>.
12. **Hernández, Yanisbel González.** *PROPUESTA DE METODOLOGIA PARA EL DASARROLLO DE ALMACENES DE DATOS*. Centro de Tecnologías de Gestión de Datos (DATEC), facultad 6, UCI, La Habana, Cuba : s.n., Octubre 2012.

Anexos



Anexo 1: Modelo de datos dimensional con topología de constelación de hechos.

| Escenario | Descripción | Variables de Entrada | Variables de Salida | Respuesta del sistema | Flujo central | | | | | | |
|---|--|---|---------------------|-----------------------|---------------------|--------------------|--------------------|---|---|---|--|
| EC 1: Presentar información de indicadores de asuntos de familia. | Muestra la información de los asuntos de familia municipales. | <table border="1"> <tr><td>Provincias</td></tr> <tr><td>Tiempo</td></tr> <tr><td>Tribunal municipal</td></tr> <tr><td>Asuntos de familia</td></tr> <tr><td>Expedientes</td></tr> </table> | Provincias | Tiempo | Tribunal municipal | Asuntos de familia | Expedientes | Cantidad de expedientes de asuntos de familia | El sistema muestra todas las variables disponibles para los análisis, ubicados en las filas y las columnas que pueden ser visualizadas para cada reporte. | <p>Se autentica.</p> <p>Se despliega hacia la derecha el componente ubicado en el lateral izquierdo que contiene el navegador. Se selecciona el A.A.G SIEJT.</p> <p>Se selecciona el A.A. Civil-Administrativo.</p> <p>Se selecciona el A.A. Instancia municipal.</p> <p>Se selecciona el A.A. DPA (1976-2010) o el A.A. DPA (2011-Actualidad)</p> <p>Se selecciona el L.T correspondiente a los asuntos de familia.</p> <p>En la parte inferior izquierda se selecciona el reporte deseado.</p> <p>En el área de trabajo se visualiza la tabla correspondiente al reporte.</p> | |
| Provincias | | | | | | | | | | | |
| Tiempo | | | | | | | | | | | |
| Tribunal municipal | | | | | | | | | | | |
| Asuntos de familia | | | | | | | | | | | |
| Expedientes | | | | | | | | | | | |
| EC 2: Presentar información de indicadores de asuntos de familia. | Muestra la información de los asuntos de familia provinciales. | <table border="1"> <tr><td>Provincias</td></tr> <tr><td>Tiempo</td></tr> <tr><td>Tribunal provincial</td></tr> <tr><td>Expedientes</td></tr> <tr><td>Asuntos de familia</td></tr> <tr><td>Tipo de sala</td></tr> </table> | Provincias | Tiempo | Tribunal provincial | Expedientes | Asuntos de familia | Tipo de sala | Cantidad de expedientes de asuntos de familia | El sistema muestra todas las variables disponibles para los análisis, ubicados en las filas y las columnas que pueden ser visualizadas para cada reporte. | <p>Se autentica.</p> <p>Se despliega hacia la derecha el componente ubicado en el lateral izquierdo que contiene el navegador. Se selecciona el A.A.G SIEJT.</p> <p>Se selecciona el A.A. Civil-Administrativo.</p> <p>Se selecciona el A.A. Instancia provincial.</p> <p>Se selecciona el A.A. DPA (1976-2010) o el A.A. DPA (2011-Actualidad)</p> <p>Se selecciona el L.T correspondiente a los asuntos de familia.</p> <p>En la parte inferior izquierda se selecciona el reporte deseado.</p> <p>En el área de trabajo se visualiza la tabla correspondiente al reporte.</p> |
| Provincias | | | | | | | | | | | |
| Tiempo | | | | | | | | | | | |
| Tribunal provincial | | | | | | | | | | | |
| Expedientes | | | | | | | | | | | |
| Asuntos de familia | | | | | | | | | | | |
| Tipo de sala | | | | | | | | | | | |

Anexo 2: Caso de prueba Indicadores de asuntos de familia.

Glosario de términos

Área de Análisis: agrupación de información según su propósito, aunque el criterio depende de las necesidades de la institución o empresa donde se aplica el sistema. Permite restringir el número de usuarios que acceden a los datos.

Base de datos relacional: es una base de datos que cumple con el modelo relacional, el cual es el modelo más utilizado en la actualidad para implementar bases de datos ya planificadas. Permiten establecer relaciones entre los datos (que están guardados en tablas), y a través de ellas relacionar los datos de ambas tablas, de ahí proviene el nombre de: "Modelo Relacional".

Cubo: colección de dimensiones y medidas en un área temática particular.

Jerarquía: implica una organización de niveles dentro de una dimensión, donde cada nivel representa el total agregado de los datos del nivel inferior.

Libro de Trabajo: estructura organizativa que agrupa los reportes generados dentro de las Áreas de Análisis. Puede ser creado teniendo en cuenta criterios que permitan organizar la información: emisor de los reportes, receptor del reporte, contenido, entre otros.

TCP/IP: son las siglas de Protocolo de Control de Transmisión/Protocolo de Internet (por el inglés Transmission Control Protocol/Internet Protocol), un sistema de protocolos que hacen posibles servicios Telnet, FTP, E-mail, y otros, entre ordenadores que no pertenecen a la misma red. TCP garantiza la entrega de datos y que los paquetes sean entregados en el mismo orden en el cual fueron enviados. IP utiliza direcciones que son series de cuatro números octetos (byte) con un formato de punto decimal.