



Universidad de las Ciencias Informáticas

Facultad 3

Título: Mercados de datos para los departamentos Civil y Familia, Administrativo y Atención a la Población de la Fiscalía General de la República.

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

Autor (es): Yaimari Rodríguez Torres

Leanet Solis Oliva

Tutor: Ing. Manuel Álvarez Alonso

Co- Tutores: Ing. José Carlos Pupo Acosta

Ing. Yeleny Almora Galvez

La Habana, 2014

“Año 56 de la Revolución.”

DECLARACIÓN DE AUTORÍA

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

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

Yaimari Rodríguez Torres

Leanet Solis Oliva

Ing. Manuel Álvarez Alonso



....Lo fundamental es que seamos capaces de hacer cada día algo que perfeccione lo que hicimos el día anterior....

Ernesto Che Guevara

Dedicatoria

A mi mamá Maribel por el apoyo y amor que siempre me ha brindado.

A mi papá Luis Alberto por ser un padre tan especial.

A mis abuelos por todo el amor y la dedicación que me han brindado siempre.

A mis tíos Noel e Ismarí que me han querido siempre como una hija.

A mis hermanos Reidel y Yelena por ser mi razón de ser.

A mi prima Iransy por ser mi niña querida.

Yaimari

Dedico esta tesis a mi mamá Yohanní Oliva Martínez por ser la mejor madre y al mismo tiempo padre, que pude tener y darme tanto amor.

A mi hermana Leyanet Hidalgo Oliva por ser la persona más importante en mi vida.

A mi abuelita María que no conocí pero que siempre está presente.

Leyanet

Agradecimientos

Quiero agradecerle a mi mamá por querer siempre lo mejor para mí, por ser tan buena madre y por el apoyo y amor que siempre me ha brindado. A mi papá por ser un padre ejemplar y por quererme tanto.

A mis abuelos por todo el amor y la dedicación que me han brindado siempre. A mis tíos que siempre han estado tan pendientes de mí. A mis hermanos y prima Iransy que son mi razón de ser.

A mi madrastra y toda su familia porque sé que me quieren como si fuese de la familia. A Ivan y toda su familia por el apoyo brindado durante los años que compartimos juntos.

A mis amistades Leodan, Yuselís, Alejandro González y Maridaisy por ser unos amigos especiales.

A mis hermanas del apartamento Kenia, Jenny, Leanet y Leidy. A las personas que durante estos meses me han ayudado en todo lo que han podido como son Polanco, Abel y mi tutor Manuel Álvarez. A mi compañera de tesis y a Yordanka que tanto hemos trabajado juntas en lograr nuestro objetivo.

A Denis y toda su familia por el apoyo y cariño brindado en estos últimos meses.

A todo aquel que de una forma u otra marco mi paso por esta universidad.

Yaimari

Agradecimientos

Quiero agradecerle a mi mamá por hacer posible todo lo que quiero, por apoyarme siempre, por ser comprensiva, respetar mis decisiones y principalmente ser madre y al mismo tiempo padre en todo momento.

A mi hermana que sabe que la adoro, por ser la razón de mi esfuerzo y sacrificio.

A mis tíos, tías y primos por darme todo su apoyo y ayuda. A mis abuelos y mi familia en general. A mi padrastro y su familia por hacer posible que formara parte de una familia tan unida

A mi hermanito Luis por ser ese hermano mayor que no tengo. A todos mis compañeros de aula, amigos y a mis hermanas en la UCI Leydis, Kenia, Jennifer y Yaimari por crear una nueva familia en la UCI.

A mi tutor Manuel, Abel, Polanco y mis co-tutores Pupo y Yeleny por su ayuda. A mi compañera de tesis y a Yordanka que tanto hemos trabajado juntas en lograr nuestro objetivo.

A mis profesores en la universidad por ser unos excelentes guías, principalmente a Dariela Leyva por confiar en mí.

A todo aquel que de una forma u otra marco mi paso por esta universidad.

Leannet

RESUMEN

Las empresas, instituciones y organismos han comenzado a ver la información como un elemento fundamental para la toma de decisiones, garantizando el éxito y el buen funcionamiento de los mismos. Un buen recurso para utilizar son los repositorios de información histórica. Entre las herramientas que facilitan el análisis de información histórica se encuentran los mercados de datos.

El presente trabajo permitió el desarrollo de diferentes mercados de datos para proporcionar el acceso a datos históricos de los departamentos Civil y Familia, Administrativo y Atención a la Población de la Fiscalía General de la República (FGR), permitiendo su análisis desde diversas perspectivas, contribuyendo a la toma de decisiones.

Estos mercados de datos surgen por la necesidad que presenta la FGR de poder almacenar datos de una manera más eficiente para su explotación y análisis. La implementación está basada en la metodología de Kimball, la estructura lógica propuesta y el diseño son consecuentes con ésta. Se realizan pruebas de volumen, carga y estrés para validar la capacidad y rendimiento de los mercados de datos en consonancia con los requerimientos definidos por el cliente.

Palabras Claves

Mercados de datos, Toma de decisiones

TABLA DE CONTENIDO

Introducción.....	1
1. Capítulo 1: Fundamentación teórica	4
1.1 Introducción	4
1.2 Almacenes de Datos	5
1.2.1 Características de los Almacenes de datos.....	5
1.2.2 Arquitectura de los Almacenes de Datos.....	6
1.3 Mercado de datos	7
1.3.1 Características de un mercado de datos	7
1.3.2 Ventajas de un mercado de datos.....	8
1.4 Metodologías de desarrollo para un Mercado de datos.	8
1.5 Metodología de Ralph Kimball.....	9
1.6 Proceso de Extracción Transformación y Carga	10
1.7 Sistemas OLAP.....	11
1.7.1 Cubos OLAP.....	12
1.7.2 Modos de almacenamiento de OLAP	13
1.8 Bases de datos multidimensional	14
1.8.1 Tablas de hechos.....	14
1.8.2 Tablas dimensiones.....	15
1.9 Tipos de modelado	15
1.10 Tendencias	17
1.10.1 Tendencias Actuales de Mercados de datos en el Mundo.....	18
1.10.2 Tendencias Actuales de los Almacenes de Datos en Cuba.....	18
1.11 Herramientas y tecnologías utilizadas para la solución	19
1.11.1 Herramientas de Inteligencia de Negocios	19
1.11.2 Sistema gestor de base de datos	21
1.11.3 Administrador de bases de datos.....	21
1.11.4 Herramienta de modelado	22
1.11.5 Contenedor Web.....	22
1.11.6 Herramientas para las pruebas.....	23
1.12 Conclusiones del Capítulo.....	23
2. Capítulo 2: Diseño e implementación de los mercados de datos.	24
2.1 Introducción.....	24
2.2 Planificación	24
2.3 Análisis de requerimientos	24

2.3.1	Requerimientos para el mercado del departamento Civil y Familia.....	24
2.3.2	Requerimientos para el mercado del departamento Atención a la Población.	26
2.3.3	Requerimientos para el mercado del departamento Administrativo.....	27
2.3.4	Tablas de Temas Analíticos	29
2.3.5	Tablas de matrices procesos/dimensiones	31
2.4	Modelado Dimensional.....	33
2.4.1	Elegir el proceso de negocio	33
2.4.2	Establecer el nivel de granularidad	34
2.4.3	Identificación de las dimensiones	36
2.4.4	Identificar las tablas de hechos y medidas.....	39
2.4.5	Identificación de atributos de dimensiones y tablas de hechos	41
2.5	Patrones de diseño de bases de datos para el MD	45
2.5.1	Patrón de llaves subrogadas	45
2.5.2	Patrón constelación de hechos	46
2.6	Estandarización del código.	46
2.7	Tipo de modelo lógico del los mercados de datos ¡Error! Marcador no definido.	
2.8	Arquitectura de los mercados de datos.....	47
2.9	Diseño físico de las bases de datos.....	48
2.10	Implementar el modelo dimensional detallado.....	51
2.11	Prueba y revisión del modelo.....	52
2.12	Diseño de los sistemas de extracción, transformación y carga de los datos.	53
2.13	Automatización del proceso ETL	55
2.14	Diseño de las dimensiones y cubos de información	55
2.15	Visualización de los resultados.....	56
3.	Capítulo 3 Pruebas de Validación.....	57
3.1	Introducción	57
3.2	Prueba de Carga y Estrés	58
3.3	Pruebas de Volumen y Carga.	61
3.4	Conclusiones del capítulo.	63
	Conclusiones	64
	Recomendaciones.....	65
	Bibliografía.....	66
	Anexos.....	68
	Glosario de términos.	80

ÍNDICE DE TABLAS

Figura # 1	Arquitectura de los almacenes de datos.....	7
Figura # 2	Proceso de extracción, transformación y carga.....	11
Figura # 3	Cubo multidimensional.....	13
Figura # 4	Esquema en estrella.....	16
Figura # 5	Esquema en copo de nieve.....	16
Figura # 6	Esquema de constelación de hechos.....	17
Figura # 7	Esquema representativo de constelación de hechos.....	46
Figura # 8	Arquitectura de los mercados de datos a desarrollar.....	48
Figura # 9	Diseño físico para Atención a la Población.....	48
Figura # 10	Diseño físico para Civil y Familia.....	49
Figura # 11	Diseño físico para Administrativo.....	51
Figura # 12	Mapeo de datos.....	52
Figura # 13	Parte del modelo físico de Administrativo.....	53
Figura # 14	Transformación para la dimensión instancia.....	54
Figura # 15	Transformación para el hecho revisiones administrativas.....	54
Figura # 16	Trabajo Revisiones Administrativas.....	55
Figura # 17	Esquema multidimensional para el mercado Administrativo.....	56
Figura # 18	Visualización de los datos del cubo Procesos Contenciosos para el mercado Administrativo.....	57
Figura # 19	Resultados del plan de pruebas simulando cantidad de usuarios de la Fiscalía Provincial.....	60
Figura # 20	Resultados del plan de pruebas simulando cantidad de usuarios de la Fiscalía General.....	60
Figura # 21	Prueba de volumen y carga para Administrativo.....	61
Figura # 22	Resultados del plan de pruebas simulando la cantidad de usuarios de la Fiscalía Provincial.....	62
Figura # 23	Resultados del plan de pruebas simulando la cantidad de usuarios de la Fiscalía General.....	62
Figura # 24	Transformación realizada para la dimensión actuación del fiscal para recurso casación.....	71
Figura # 25	Transformación realizada para la dimensión actuación del fiscal para proceso sucesorio.....	71
Figura # 26	Transformación realizada para la dimensión motivo.....	71
Figura # 27	Transformación realizada para la dimensión resolución.....	72

Figura # 28 Transformación realizada para la dimensión tipo de bien.	72
Figura # 29 Transformación realizada para la dimensión entidad involucrada....	72
Figura # 30 Transformación realizada para la dimensión materia.	72
Figura # 31 Transformación realizada para la dimensión persona.	72
Figura # 32 Transformación realizada para la dimensión pronunciamiento.	72
Figura # 33 Transformación realizada para la dimensión respuesta impugnación	73
Figura # 34 Transformación realizada para la dimensión asunto de proceso contencioso sumario.....	73
Figura # 35 Transformación realizada para la dimensión asunto de jurisdicción voluntaria.	73
Figura # 36 Transformación realizada para la dimensión asunto notarial.....	73
Figura # 37 Transformación realizada para la dimensión proceso contencioso ordinario.....	73
Figura # 38 Transformación realizada para la dimensión fiscal recurrente.....	73
Figura # 39 Transformación realizada para la dimensión motivo de patria potestad.....	74
Figura # 40 Transformación realizada para el hecho procedimientos contenciosos	74
Figura # 41 Transformación realizada para el hecho procesos contenciosos	74
Figura # 42 Transformación realizada para el hecho recursos de casación.	74
Figura # 43 Transformación realizada para el hecho atención población.	74
Figura # 44 Transformación realizada para el hecho impugnaciones.	74
Figura # 45 Transformación realizada para el hecho reclamaciones.....	75
Figura # 46 Transformación realizada para el hecho asuntos notariales	75
Figura # 47 Transformación realizada para el hecho expropiación forzosa.....	75
Figura # 48 Transformación realizada para el hecho expediente prejudicial.	75
Figura # 49 Transformación realizada para el hecho jurisdicción voluntaria.....	75
Figura # 50 Transformación realizada para el hecho procesos contenciosos ordinarios.....	75
Figura # 51 Transformación realizada para el hecho procesos contenciosos sumarios.....	76
Figura # 52 Transformación realizada para el hecho procesos sucesorios.....	76
Figura # 53 Transformación realizada para el hecho recursos de apelación.....	76
Figura # 54 Transformación realizada para el hecho recursos de casación.	76
Figura # 55 Transformación realizada para el hecho revisiones civiles.	76
Figura # 56 Trabajo realizado para el hecho procedimiento contencioso.	76

Figura # 57 Trabajo realizado para el hecho procesos contenciosos.....	76
Figura # 58 Trabajo realizado para el hecho recursos de casación.....	77
Figura # 59 Trabajo realizado para el hecho atención población.....	77
Figura # 60 Trabajo realizado para el hecho impugnaciones.....	77
Figura # 61 Trabajo realizado para el hecho asuntos notariales.....	77
Figura # 62 Trabajo realizado para el hecho expropiación forzosa.	77
Figura # 63 Trabajo realizado para el hecho jurisdicción voluntaria.	77
Figura # 64 Trabajo realizado para el hecho proceso contencioso ordinario.	77
Figura # 65 Trabajo realizado para el hecho proceso contencioso sumario.....	78
Figura # 66 Trabajo realizado para el hecho procesos sucesorios.....	78
Figura # 67 Trabajo realizado para el hecho recursos de apelación.....	78
Figura # 68 Trabajo realizado para el hecho recursos de casación.....	78
Figura # 69 Trabajo realizado para el hecho revisiones civiles.....	78
Figura # 70 Resultados del plan de pruebas simulando la cantidad de usuarios de la Fiscalía Provincial.	¡Error! Marcador no definido.
Figura # 71 Resultados del plan de pruebas simulando la cantidad de usuarios de la Fiscalía General.	¡Error! Marcador no definido.
Figura # 72 Resultados del plan de pruebas simulando la cantidad de usuarios de la Fiscalía Provincial	¡Error! Marcador no definido.
Figura # 73 Resultados del plan de pruebas simulando la cantidad de usuarios de la Fiscalía General.	¡Error! Marcador no definido.
Figura # 74 Resultados del plan de pruebas simulando la cantidad de usuarios de la Fiscalía Provincial	¡Error! Marcador no definido.
Figura # 75 Resultados del plan de pruebas simulando la cantidad de usuarios de la Fiscalía General	¡Error! Marcador no definido.

INTRODUCCIÓN

Los rápidos cambios que se viven en el mercado actual junto a las competencias que se generan cada día, hacen que las empresas no puedan postergar las decisiones relacionadas directamente con el negocio; una demora en este sentido puede llevar la gestión de la empresa al fracaso.

Las personas encargadas de la toma de decisiones en las empresas, han comenzado a ver la información no solo como un historial de conducción de la entidad sino como un factor que alimenta el negocio para poder determinar el éxito o el fracaso del mismo.

Es necesario entonces contar con un sistema que juegue el papel de soporte para la toma de decisiones, de respuesta ágil y rápida con información precisa para poder aprovechar las oportunidades. Bajo el nombre de Inteligencia de Negocios se cobijan diferentes acrónimos, herramientas y disciplinas que apuntan a dar soporte a la tarea de toma de decisiones.

En Cuba, se llevan a cabo diversas transformaciones económicas, entre las que se encuentra la introducción en muchas empresas y entidades de sistemas de apoyo a la toma de decisiones. Sin embargo existen muchas empresas que aún no poseen un sistema que satisfaga con éxito sus necesidades, ni se adecuen a las características del sistema económico del país. Es en este punto donde toman lugar relevante la creación de sistemas de apoyo a la toma de decisiones, que den solución a las necesidades existentes en el país.

La Universidad de las Ciencias Informáticas (UCI), es un centro que se caracteriza por formar ingenieros informáticos, los cuales también desarrollan proyectos reales que benefician al país. Entre los proyectos a desarrollar se encuentra Sistema de Gestión de la Fiscalía (SIGEF), el mismo se encarga de informatizar los procesos de la Fiscalía General de la República (FGR). La FGR no es más que el órgano encargado de vigilar por el estricto cumplimiento de la constitución, las leyes y las disposiciones legales.

Entre los módulos de SIGEF se encuentran Civil y Familia, Administrativo y Atención a la Población, que tienen como objetivo principal informatizar los procesos de los departamentos del mismo nombre en los distintos niveles de las FGR. La dirección de la FGR y la de estos módulos necesitan poseer un repositorio de datos histórico que permita un acceso rápido al mismo.

El sistema transaccional existente no facilita realizar la búsqueda de información histórica con agilidad, debido a la forma en que está organizada la misma, y además tampoco permite realizar comparaciones de la totalidad de los datos significativos en un determinado intervalo en el tiempo, elementos que afectan la efectividad del proceso de toma de decisiones de los directivos de los departamentos de la FGR.

Dada la situación anteriormente descrita, se plantea el siguiente **problema a resolver**: ¿cómo contribuir a la toma de decisiones en los departamentos Civil y Familia, Administrativo y Atención a la Población de la Fiscalía General de la República?.

Para analizar el problema a resolver, se toma como **objeto de estudio** los sistemas de apoyo a la toma de decisiones y como **campo de acción** la gestión de los mercados de datos.

El **objetivo General** es desarrollar mercados de datos para los departamentos Civil y Familia, Administrativo y Atención a la Población de la Fiscalía General de la República que contribuya a la toma de decisiones de los directivos de estas áreas.

Teniendo como referencia el objetivo general y el problema se plantea como **idea a defender**: el desarrollo de mercados de datos para los departamentos Civil y Familia, Administrativo y Atención a la Población, contribuirá al proceso de la toma de decisiones.

Del objetivo general del trabajo se desglosan los siguientes **objetivos específicos**:

1. Realizar el marco teórico referencial de la investigación.
2. Identificar las necesidades de información.
3. Diseñar los mercados de datos.
4. Realizar los procesos de extracción, transformación y carga.
5. Realizar pruebas a la solución.

Para lograr el cumplimiento a los objetivos específicos se plantean las siguientes **tareas de investigación**:

1. Estudio del estado del arte sobre los principales sistemas de soporte a la toma de decisiones y sistemas de software que realizan procesos similares.

- Estudiar los temas relacionados a las mejores prácticas en el desarrollo de almacén de datos.
 - Definir la metodología a utilizar en el desarrollo.
2. Obtención del análisis de los principales requerimientos de información en los módulos Civil y Familia, Administrativo y Atención a la Población.
 3. Definición de la granularidad del proceso del negocio.
 4. Definición de las dimensiones de los Mercados de datos.
 5. Definición de los hechos mensurables asociados a las dimensiones definidas.
 6. Estructuración del modelo dimensional.
 7. Transformación del modelo lógico dimensional al diseño físico.
 8. Definición de la estrategia de extracción, transformación y carga de los datos.
 9. Implementación de los mercados de datos.
 10. Validación de la solución implementada mediante pruebas de software.

Métodos de investigación científica a aplicar:

Los métodos científicos son los procedimientos que son usados para estudiar la realidad existente con el propósito de descubrir su esencia y sus relaciones, los métodos científicos que se utilizarán para darle solución a los objetivos propuestos son:

Métodos teóricos:

Analítico-Sintético: este método se utiliza en el estudio de las diferentes bibliografías para sintetizar el conocimiento aprendido.

Modelación: el modelo es una reproducción simplificada de la realidad que permite descubrir nuevas relaciones y cualidades del objeto de estudio. Este método se evidenciará en el diseño de los mercados de datos.

Métodos empíricos:

Entrevista: es una técnica de recopilación de información mediante una conversación profesional, con la que se adquiere información acerca de lo que se investiga. Según el fin que se persigue esta puede estar o no estructurada mediante un cuestionario previamente elaborado. Este método se evidenciará en las conversaciones con los analistas de los departamentos.

Para dar cumplimiento a los objetivos trazados, el presente documento estará estructurado en tres capítulos. A continuación se expone una breve descripción de cada uno de ellos.

Capítulo 1: Fundamentación teórica sobre el desarrollo de los mercados de datos. En este capítulo se realiza una breve explicación de las metodologías, técnicas y herramientas que se tuvieron en cuenta para dar solución al problema; así como los principales conceptos y características de las mismas.

Capítulo 2: Diseño e implementación de los mercados de datos. En este capítulo se muestran aspectos relacionados con el modelado y diseño de la solución propuesta. Además de la implementación de las reglas de transformación de la base de datos a los mercados de datos.

Capítulo 3: Pruebas de validación. En el capítulo se realizan un conjunto de pruebas de volumen, carga y estrés con el fin de garantizar la calidad de la solución propuesta.

1. CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

1.1 INTRODUCCIÓN

En el presente capítulo se abordan algunas de las definiciones fundamentales de los almacenes de datos (AD) y mercados de datos (MD), así como algunas de sus características principales. Se realiza un estudio de las principales metodologías para realizar su desarrollo, además de analizar las herramientas y tecnologías que pueden ser empleadas en su construcción.

1.2 ALMACENES DE DATOS

Desde su surgimiento, hace ya algunos años, muchos autores han estudiado a los almacenes de datos, y han planteado sus consideraciones acerca de los mismos, a continuación se referencian algunas de las definiciones más conocidas.

Entre las definiciones existentes está la de Inmon, el define un almacén de datos como una colección de datos orientados a temas, integrados, no-volátiles y variante en el tiempo, organizados para soportar necesidades empresariales. (1)

Según Ralph Kimball (considerado el principal promotor del enfoque dimensional para el diseño de almacenes de datos), un almacén de datos es una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis. (2)

De las definiciones anteriores se puede concluir que un almacén de datos no es más que un sistema integrado y multidimensional de almacenamiento histórico de datos estables, funcionalmente dirigido a consultas, reportes y análisis de información.

1.2.1 CARACTERÍSTICAS DE LOS ALMACENES DE DATOS.

Los almacenes de datos se caracterizan por los siguientes aspectos:

Integrados: integra datos recogidos de diferentes sistemas operacionales de la organización, los cuales deben integrarse en una estructura consistente.

Temáticos: los datos son organizados por temas para facilitar su entendimiento y acceso por parte de los usuarios.

Históricos: se tiene información histórica para poder comparar los datos en diferentes períodos del tiempo e identificar las tendencias de estos. Los datos no son actualizados si no que se tiene almacenado los valores que ha tenido a lo largo del tiempo.

No volátiles: luego que es incorporada la información al almacén de datos debe mantenerse invariable, cargándose una vez en el tiempo y no se permite actualizarla. Los datos solo pueden ser leídos, y no modificados.

1.2.2 ARQUITECTURA DE LOS ALMACENES DE DATOS.

El almacén de datos está formado por diversos elementos que interactúan entre sí y que cumplen una función específica dentro del sistema. A continuación se brinda información de los principales elementos que conforman la arquitectura de un almacén de datos.

OLTP representa toda aquella información transaccional que genera la empresa en su accionar diario, además, de las fuentes externas con las que puede llegar a disponer. Esta información generalmente reside en diferentes tipos de sistemas y arquitecturas y tienen formatos muy variados.

En el **gestor de carga** los datos son integrados, transformados y limpiados, para luego ser cargados en el AD, básicamente es el proceso de extracción, transformación y carga (ETL) por sus siglas en inglés. Se extraen datos de las diversas fuentes que se requieran, los transforman para resolver posibles problemas de inconsistencias entre los mismos y finalmente, después de haberlos depurado se procede a su carga en el depósito de datos.

En el **gestor del almacén de datos** la información del almacén de datos se estructura en cubos multidimensionales, los cuales preparan esta información para responder a consultas dinámicas. Realiza copias de resguardo incremental o total de los datos del almacén de datos.

El **gestor de consultas** realiza las operaciones necesarias para soportar los procesos de gestión y ejecución de consultas relacionales, tales como Join y agregaciones, y de consultas propias del análisis de datos, como drill-up y drill-down. Recibe las consultas del usuario, las aplica a las tablas correspondientes y devuelve los resultados obtenidos.

Las **herramientas de consulta y análisis** son sistemas que permiten al usuario realizar la exploración de datos del AD. Básicamente constituyen el nexo entre el depósito de datos y los usuarios. Los usuarios acceden al AD utilizando diversas herramientas de consulta, exploración, análisis, reportes, etc.

Los **usuarios** que posee el AD son aquellos que se encargan de tomar decisiones y de planificar las actividades del negocio, es por ello que se hace tanto énfasis en la integración y limpieza de datos, para poder conseguir que la información posea toda la

calidad posible. Es a través de las herramientas de consulta y análisis, que los usuarios exploran los datos en busca de respuestas para poder tomar decisiones proactivas. (3)



Figura # 1 Arquitectura de un Almacén de Datos.

1.3 MERCADO DE DATOS

Un mercado de datos es un repositorio de información, similar a un almacén de datos, pero orientado a un área o departamento específico de la organización, un mercado de datos se organiza por lo general alrededor de un único proceso de negocio. (2)

En síntesis, los mercados de datos son pequeños almacenes de datos centrados en un tema o un área de negocio específico dentro de una organización.

1.3.1 CARACTERÍSTICAS DE UN MERCADO DE DATOS

Los mercados de datos tienen las siguientes características:

Usuarios limitados debido a que solo utilizarán el sistema los usuarios con conocimientos del negocio en el área a la que se realiza el mercado de datos y dedicados a la toma de decisiones en la organización.

Área específica porque la información presentada es de un área o departamento en específico y no de toda la organización.

Propósito específico debido a que se tiene como objetivo analizar la información importante y relevante en un tiempo determinado del área.

Función de apoyo ya que ayuda a la consulta, análisis de información y a los reportes de la organización facilitando la toma de decisiones.

Fácilmente entendibles y navegables: porque emplea términos propios de la organización para facilitar su entendimiento además de presentar graficas interactivas.

1.3.2 VENTAJAS DE UN MERCADO DE DATOS

- Transforma datos orientados a las aplicaciones en información orientada a la toma de decisiones.
- Integra y consolida diferentes fuentes de datos en una única plataforma sólida y centralizada.
- Provee la capacidad de analizar y explotar las diferentes áreas de trabajo y de realizar un análisis inmediato de las mismas.
- Elimina la producción y el procesamiento de datos que no son utilizados ni necesarios.
- Mejora la entrega de información, es decir información completa, consistente y que los usuarios necesitan en el momento y formato apropiado.
- Los usuarios pueden acceder directamente a la información en línea, contribuyendo a su capacidad para operar en las tareas rutinarias.
- Permite la toma de decisiones estratégicas y tácticas.

1.4 METODOLOGÍAS DE DESARROLLO PARA UN MERCADO DE DATOS.

Existen muchas metodologías de diseño y construcción de un MD. Cada fabricante de software de inteligencia de negocios busca imponer una metodología con sus productos. Sin embargo, se imponen entre la mayoría tres metodologías, la de Kimball, la de Inmon y Hefestos.

La metodología Kimball es la más aceptada en todo el mundo como la metodología más efectiva para desarrollar una solución de Inteligencia de Negocio. Propone una arquitectura ascendente, partiendo de la idea de que un almacén de datos es la unión de todos los mercados de datos de una entidad. Define un diseño dirigido por procesos simples.

Por su parte **Inmon** define una metodología descendente a la hora de diseñar un almacén de datos, ya que de esta forma se consideran mejor todos los datos

corporativos. En esta metodología los mercados de datos se crean después de haber terminado el almacén de datos completo de la organización.

Hefestos: está fundamentada en una amplia investigación, comparación de metodologías existentes y experiencias propias en proceso de confección de almacenes de datos. Puede adaptarse muy bien a cualquier ciclo de vida de desarrollo de software, aunque para algunas fases en particular, las acciones que se han de realizar serán muy diferentes. Lo que se busca es entregar una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas del almacén de datos y motivar a los usuarios.

Una vez estudiadas las metodologías antes mencionadas, se decide utilizar la metodología Kimball como guía para la construcción de los mercados de datos. Esta metodología es de fácil comprensión, rápida, resistente y adaptable ante los cambios. Es acorde al proyecto SIGEF, porque se pueden implementar pequeños mercados de datos en áreas o módulos específicos del proyecto y después irlos integrando en un almacén de datos, elemento fundamental este, ya que SIGEF no ha terminado su implementación.

1.5 METODOLOGÍA DE RALPH KIMBALL.

La metodología se basa en lo que Kimball denomina Ciclo de Vida Dimensional del Negocio (Kimball et al 98, 08, Mundy & Thornthwaite 06). Este ciclo de vida del proyecto está basado en cuatro principios básicos:

- **Centrarse en el negocio:** Hay que concentrarse en la identificación de los requerimientos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio, agudizando el análisis del mismo y la competencia consultiva de los implementadores.
- **Construir una infraestructura de información adecuada:** Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa.
- **Realizar entregas en incrementos significativos:** Crear el almacén de datos en incrementos entregables. Hay que usar el valor de negocio de cada elemento identificado para determinar el orden de aplicación de los incrementos.

- Ofrecer la solución completa: Proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Para comenzar, esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible.

La construcción de una solución de un almacén de datos es sumamente compleja, y Kimball nos propone una metodología que nos ayuda a simplificar esa complejidad. A continuación se mencionan las tareas de esta metodología:

- ✓ Planificación.
- ✓ Análisis de requerimientos.
- ✓ Modelado Dimensional.
 - Elegir el proceso de negocio.
 - Establecer el nivel de granularidad.
 - Elegir las dimensiones.
 - Identificar las tablas de hechos y medidas.
 - Identificación de atributos de dimensiones y tablas de hechos.
 - Implementar el modelo dimensional detallado.
 - Prueba del modelo.
- ✓ Diseño Físico.
- ✓ Diseño del sistema de Extracción, Transformación y Carga (ETL).
- ✓ Especificación y desarrollo de aplicaciones de inteligencia de negocio. (4)

1.6 PROCESO DE EXTRACCIÓN TRANSFORMACIÓN Y CARGA

El sistema de Extracción, Transformación y Carga (ETL), por sus siglas en inglés, es la base sobre la cual se alimenta el almacén de datos. Si el sistema ETL se diseña adecuadamente, puede extraer los datos de los sistemas de origen de datos, aplicar diferentes reglas para aumentar la calidad y consistencia de los mismos, consolidar la información proveniente de distintos sistemas, y finalmente cargar la información en el

almacén de datos en un formato acorde para la utilización por parte de las herramientas de análisis.

Extracción: Es aquí, donde, basándose en las necesidades y requisitos de los usuarios, se exploran las diversas fuentes OLTP que se tengan a disposición, y se extrae la información. Si los datos operacionales residen en un Sistema Gestor de Base de Datos Relacional (SGBDR), el proceso de extracción se puede reducir a, por ejemplo, consultas en SQL o rutinas programadas.

Transformación: Esta función es la encargada de convertir aquellos datos inconsistentes en un conjunto de datos compatibles y congruentes, para que puedan ser cargados en el almacén de datos. Además de lo antes mencionado, esta función se encarga de realizar, entre otros, los procesos de limpieza de datos y calidad de datos.

Carga: Consiste en almacenar los datos transformados y validados en la base de datos final.



Figura # 2 Proceso de extracción, transformación y carga.

1.7 SISTEMAS OLAP

Las herramientas OLAP (Procesamiento Analítico en Línea) por sus siglas en inglés, son una tecnología de software para el análisis en línea, administración y ejecución de consultas, que permiten inferir información del comportamiento del negocio. Su principal objetivo es brindar rápidas respuestas a complejas preguntas encaminadas generalmente a interpretar la situación del negocio y tomar decisiones.

Este tipo de herramientas, puede analizar el negocio desde diferentes escenarios históricos, y proyectar como se ha venido comportando y evolucionando en un

ambiente multidimensional, o sea, mediante la combinación de diferentes perspectivas, temas de interés o dimensiones. Esto permite deducir tendencias, por medio del descubrimiento de relaciones entre las perspectivas que a simple vista no se podrán encontrar sencillamente.

1.7.1 CUBOS OLAP

En la base de cualquier sistema OLAP se encuentra el concepto de cubo OLAP (también llamado cubo multidimensional o hipercubo). Un cubo OLAP no es más que un vector de varias dimensiones. Desde un punto de vista relacional, puede verse como una tabla de hechos que tiene dos elementos fundamentales:

Indicadores: también denominados métricas o ratios, son los valores numéricos con los que se opera.

Dimensiones: son las características por las que se pueden filtrar y cruzar los indicadores.

En general suele resultar necesario dimensionar la volumetría de los cubos para conseguir que generen los informes deseados, y monitorizar su tamaño para garantizar que los resultados se obtienen en el tiempo esperado.

En este sentido suele resultar fundamental tener en cuenta dos características básicas: la **cardinalidad del cubo**, posibles combinaciones de todos los valores de todas las dimensiones, y la **granularidad del cubo**, nivel de detalle máximo de los datos, o lo que es lo mismo, nivel de agregación mínimo de la información.

De esta forma, en un cubo multidimensional los atributos existen a lo largo de varios ejes o dimensiones, y la intersección de las mismas representa el valor que tomará el indicador que se está evaluando. (5)

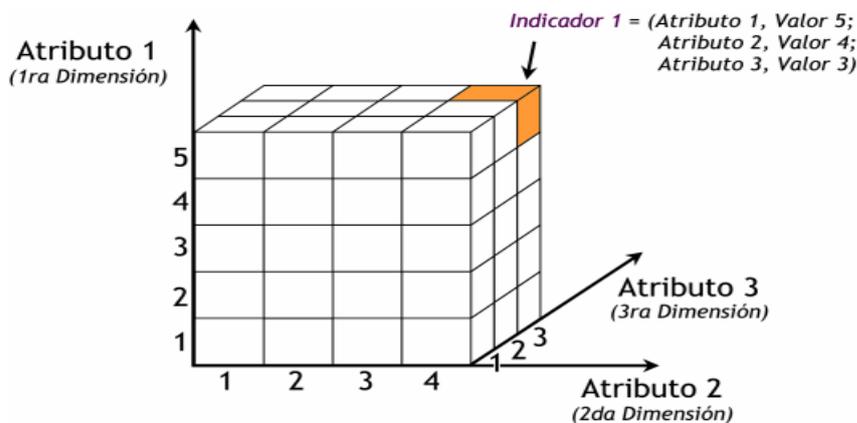


Figura # 3 Cubo multidimensional

1.7.2 MODOS DE ALMACENAMIENTO DE OLAP

Hay diversos tipos de implementaciones de la tecnología OLAP, las que varían según el tipo de motor en el que se almacenan los datos. Los cuales se pueden clasificar de la siguiente manera:

OLAP multidimensional (MOLAP): El modo de almacenamiento MOLAP da lugar a que las agregaciones de la partición y una copia de sus datos de origen se almacenen en una estructura multidimensional de alto rendimiento, por tanto las consultas pueden resolverse sin obtener acceso a los datos de origen de la partición, incluso cuando no se puedan obtener los resultados de las agregaciones de la partición. El modo de almacenamiento MOLAP proporciona los tiempos de respuesta de consulta más rápidos posibles, en función del diseño y porcentaje de las agregaciones de la partición, así como un rendimiento y una compresión de los datos excelente.

OLAP relacional (ROLAP): Este modo de almacenamiento hace que las agregaciones de la partición se almacenen en tablas de la base de datos relacional especificada en el origen de datos de la partición, permitiendo sacarle las ventajas a la tecnología relacional. A diferencia del modo de almacenamiento MOLAP, no da lugar a que se almacene una copia de los datos de origen; cuando los resultados no pueden derivarse de las agregaciones o de la caché de cliente, se obtiene acceso a la tabla de hechos de la partición para responder a las consultas. Con este modo de

almacenamiento, la respuesta de consultas suele ser más lenta que la de los otros dos modos.

OLAP híbrido (HOLAP): HOLAP combina atributos de los modos MOLAP y ROLAP. Al igual que MOLAP, almacena los datos de las agregaciones de la partición en una estructura multidimensional de alto rendimiento, pero no almacena una copia de los datos de origen. HOLAP es el equivalente de MOLAP para las consultas que sólo obtienen acceso a los datos de resumen contenidos en las agregaciones de una partición. Las consultas que obtienen acceso a los datos de origen deben recuperar datos de la base de datos relacional y no serán tan rápidas como lo serían si los datos de origen se hubieran almacenado en la estructura MOLAP. Las particiones almacenadas como HOLAP son más pequeñas que sus equivalentes MOLAP y responden más rápidamente que las particiones ROLAP a las consultas que implican datos de resumen.

Se decide que el tipo de almacenamiento será MOLAP porque la estructura de datos almacenada es una estructura multidimensional. Permite consultas rápidas debido a la optimización del rendimiento de almacenamiento. Ocupa menor tamaño en disco en comparación con los datos almacenados en base de datos relacional debido a técnicas de compresión. Muy compacto para conjuntos de datos de pocas dimensiones.

1.8 BASES DE DATOS MULTIDIMENSIONAL

Las bases de datos multidimensionales se utilizan principalmente para crear aplicaciones OLAP. Las mismas están formadas por dos tipos de tablas, estas son las tablas de hechos y las tablas de dimensiones. Las principales ventajas de este tipo de bases de datos son la versatilidad para cruzar información y la alta velocidad de respuesta. Esto las convierte en herramientas básicas para soluciones de Inteligencia de Negocio donde el análisis de los datos resulta crucial. (6)

1.8.1 TABLAS DE HECHOS

La tabla de hechos es donde las mediciones numéricas del negocio son almacenadas. Cada una de las mediciones es tomada como la intersección de todas las dimensiones. Esta tabla permite la evaluación principal de los indicadores del modelo de Inteligencia de Negocios. Las tablas de hechos pueden contener un gran número

de filas, a veces cientos de millones de registros cuando contienen uno o más años de la historia de una gran organización, esta cardinalidad estará acotada superiormente por la cardinalidad de las tablas dimensionales. (6)

1.8.2 TABLAS DIMENSIONES

Las tablas dimensionales son aquellas donde las descripciones textuales de las dimensiones del negocio son almacenadas. Cada una de las descripciones textuales ayuda a describir un miembro de la dimensión respectiva. Representan los ejes del cubo, y los aspectos de interés, mediante los cuales el usuario podrá filtrar y manipular la información almacenada en la tabla de hechos. La mayoría de los almacenes de datos necesitan una tabla dimensional de tiempo explícita, aun cuando la clave primaria tiempo pueda ser un objeto SQL evaluado como fecha. (6)

1.9 TIPOS DE MODELADO

Las bases de datos multidimensionales implican tres variantes posibles de modelado que permiten realizar consultas de soporte de decisión: esquema en estrella, esquema copo de nieve y esquema constelación de hechos.

Esquema en Estrella: es la arquitectura de almacén de datos más simple. En este diseño del almacén de datos, la tabla de hechos está rodeada por dimensiones y juntos forman una estructura que permite implementar mecanismos básicos para poder utilizarla con una herramienta de consultas OLAP.

En este modelo, para obtener información solicitada no hay que construir una sentencia SQL muy compleja que léase muchas tablas. Una herramienta de consultas sólo tiene que acceder a una tabla. La clave primaria de una tabla de hechos está formada por todas las columnas que corresponden a las dimensiones. Las columnas que contienen los datos numéricos no forman parte de la clave primaria, porque están agregadas en los informes.

Se puede encontrar casi cada información de una tabla de hechos en una tabla de dimensiones. Lo característico de la arquitectura de estrella es que sólo existe una tabla de dimensiones para cada dimensión. (6)

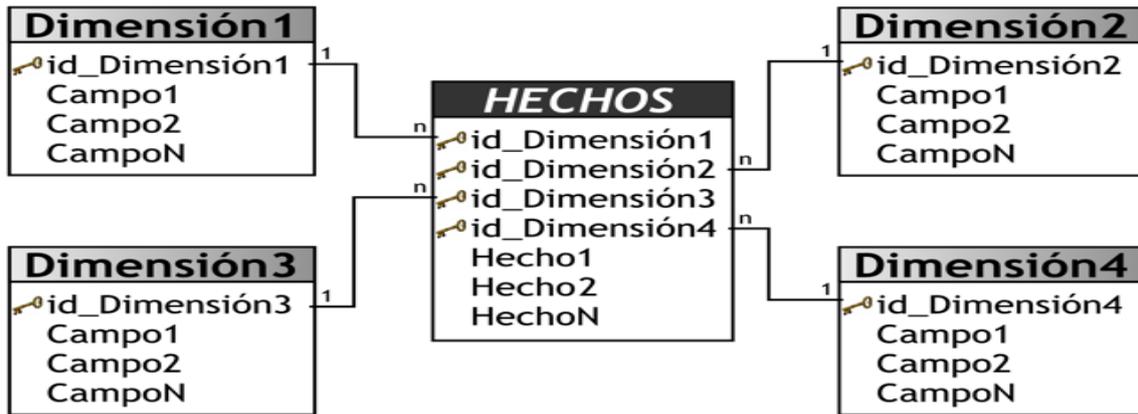


Figura # 4 Esquema en estrella

Esquema en copo de nieve: es una variedad más compleja del esquema estrella. El afinamiento está orientado a facilitar mantenimiento de dimensiones.

Con varios usos del esquema en copo de nieve, el más común es cuando las tablas de dimensiones están muy grandes o complejas y es muy difícil representar los datos en esquema estrella.

El problema es que para extraer datos de las tablas en esquema de copo de nieve, a veces hay que vincular muchas tablas en las sentencias SQL que puede llegar a ser muy complejo y difícil para mantener. (6)

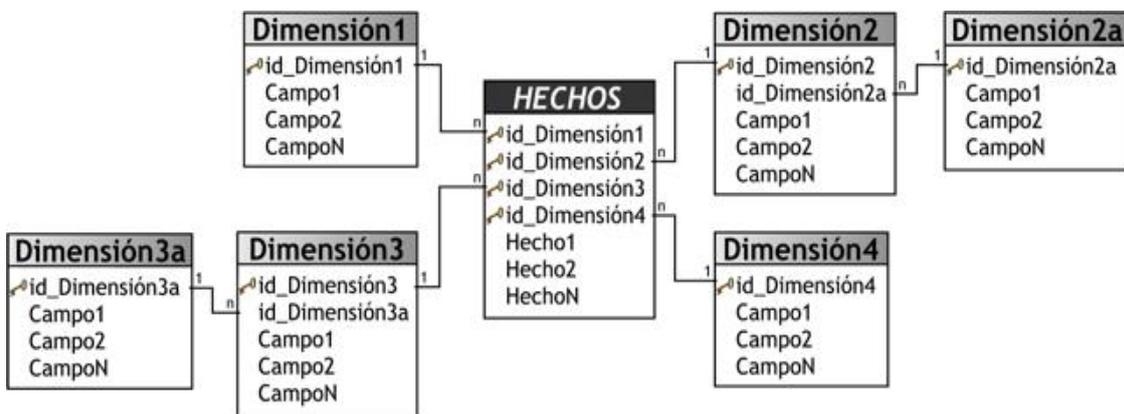


Figura # 5 Esquema en copo de nieve

Esquema de constelación de hechos: para cada esquema estrella o esquema del copo de nieve en almacén de datos es posible construir un esquema de constelación de hechos.

Este esquema es más complejo que los otros tipos de esquemas debido al hecho de que contiene múltiples tablas de hechos. Con esta solución las tablas de dimensiones pueden estar compartidas entre más de una tabla de hechos.

El esquema de constelación de hechos tiene mucha flexibilidad y este hecho es su grande virtud. Sin embargo, el problema es que cuando el número de las tablas vinculadas aumenta, la arquitectura puede llegar a ser muy compleja y difícil para mantener.

En un esquema de constelación de hechos las distintas tablas de los hechos están asignadas a las dimensiones relevantes para cada hecho. Esto puede ser útil cuando los hechos están asignados a un nivel de una dimensión y los otros hechos a otro nivel de detalle de una dimensión. (6)

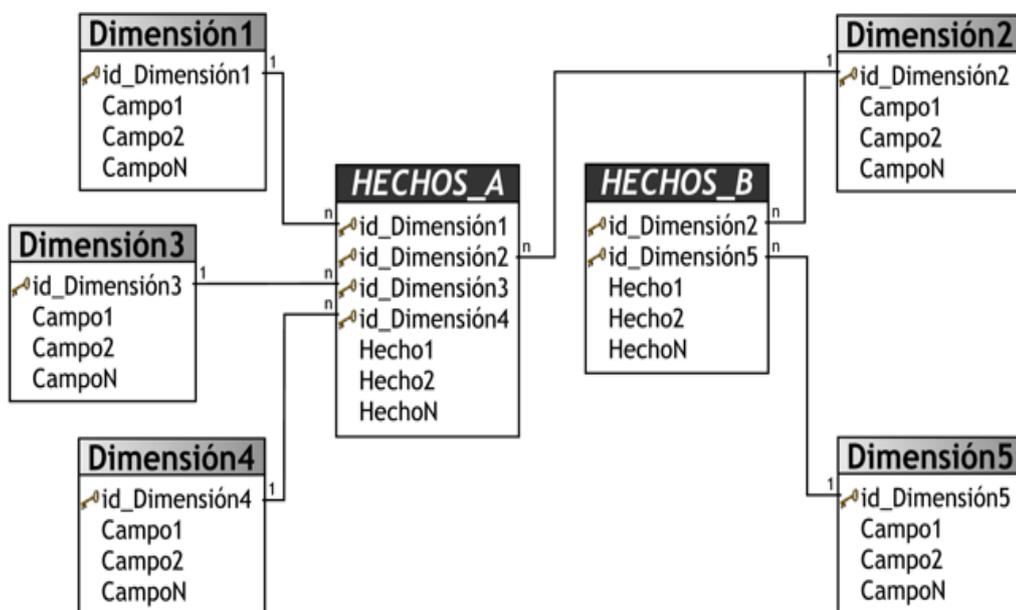


Figura # 6 Esquema de constelación de hechos.

1.10 TENDENCIAS

En todo el mundo empresas que manejan grandes volúmenes de datos relativos a clientes, compras, marketing, transacciones y operaciones usan almacenes de datos para llevar el control de su información. Los sectores donde principalmente se ha implantado almacenes de datos son las empresas de telecomunicaciones, transporte financieras, aseguradoras, de comercio minorista y de fabricación de bienes de consumo masivo. Entre ellas se encuentran Vodafone, France Telecom, Union Pacific, Coca-Cola, Adidas, Nike, 3M, Bosh Siemens.

1.10.1 TENDENCIAS ACTUALES DE MERCADOS DE DATOS EN EL MUNDO.

Existen muchas empresas y organizaciones que han creado almacenes de datos y mercados de datos para mejorar la gestión de la información, permitiendo analizar los datos disponibles y ayudar a la toma de decisiones.

Diseño de un Mercado de datos para la Unidad de Programación y Abastecimiento de Hidrocarburos de Petrocomercial, Filial de Petroecuador. Ofreciendo la base para técnicas eficaces de análisis y toma de decisiones en Sistemas Operacionales de Petrocomercial IBM S/390, Base de datos DB2 Sistema de Movimiento de Productos (MOPRO), Archivos VSAM Sistema de Comercialización (Ventas) IBM AS/400, Base de Datos DB2/400, Sistemas Financieros, Sistemas Administrativos, Sistema de Inventarios y Compras, Sistema de Mantenimiento y Sistema de Contratos. Con la construcción de un Mercado de datos los ejecutivos descubren que requieren información más precisa y rápida que antes. (7)

La Empresa Chilena de Inteligencia de Negocios DataMart S.A. dedicada a proyectos creadores de Cubix Olap Analyzer 3.1 + DataMining, con una vasta experiencia en sistemas de control de gestión y modelamiento de almacenes de datos y bases de datos multidimensionales.

Nacida en el 2000 están convertidos en el referente Latinoamericano para proyectos de Control de Gestión, Sistemas OLAP, Estudios de Minería de Datos y desarrollo de almacenes de datos relacional y multidimensional. Es la primera empresa Chilena en desarrollar un cliente OLAP de primera categoría, compatible con los estándares y definiciones de "The Microsoft Datawarehouse Framework". (8)

En agosto del 2007 la Universidad de San Carlos de Guatemala en la Facultad de Ingeniería en Ciencias y Sistemas desarrolló un Mercado de datos de información académica de estudiantes de la Escuela de Ciencias y Sistemas, por dos de sus estudiantes, para optar por el título de Ingeniero en Ciencias y Sistemas. Siendo este sistema de gran ayuda para la toma de decisiones en la universidad. (9)

1.10.2 TENDENCIAS ACTUALES DE LOS ALMACENES DE DATOS EN CUBA.

El desarrollo tecnológico en Cuba no se ha quedado atrás y algunas empresas utilizan el almacén de datos para ayudar a mantener un historial de su información y apoyar a la toma de decisiones de las mismas, entre estos se encuentran:

En la Empresa de Proyectos de Arquitectura e Ingeniería (EMPAI) de Matanzas se ha creado un almacén de datos que garantizará mayor disponibilidad y calidad de la información y la realización de análisis dinámicos de la misma, facilitando así la toma de decisiones estratégicas de sus directivos, el ahorro de tiempo de los especialistas del área de Contabilidad y de materiales de oficina. La novedad del mismo se manifiesta en dos sentidos, el primero relacionado con la solución seleccionada lo cual constituye una tecnología de avanzada y el segundo por ser la primera vez que se aplica una solución de este tipo en una Empresa de Proyectos de Arquitectura e Ingeniería de la Construcción en el país. El trabajo consistió en el diseño e implementación de una herramienta que permita gestionar y organizar homogéneamente la información relevante actual e histórica sobre los indicadores de eficiencia de la gestión contable.

En el Departamento de Informática de la Facultad de Ingeniería Eléctrica en la Universidad de Oriente se implantó un almacén de datos para la ayuda a la toma de decisiones en la Planeación Estratégica Clásica Empresarial efectuándose una investigación y análisis de los mercados de datos, almacenes de datos, los procesos de extracción, transformación, carga y de inteligencia de negocio; también se detallan las herramientas y metodología para el desarrollo de este tipo de soluciones. (10)

En el 2013 se realizó un Mercado de datos para una dirección de salud en la Universidad de Ciencias Informáticas. Enmarcando las investigaciones en temas de almacenes de datos, los mercados de datos y su utilización para los análisis estadísticos de la información en una dirección de salud. Basándose en las ventajas de las metodologías de Kimball y Inmon. Otro que se realizó fue para la dirección de cuadros de la Administración Provincial de Artemisa al igual que un almacén de datos aplicado a la seguridad ciudadana. (11)

1.11 HERRAMIENTAS Y TECNOLOGÍAS UTILIZADAS PARA LA SOLUCIÓN

1.11.1 HERRAMIENTAS DE INTELIGENCIA DE NEGOCIOS

Las herramientas de Inteligencia de Negocios han sido creadas para apoyar la toma de decisiones de las empresas o instituciones del estado. Estas muestran una visión general de los procesos que se llevan a cabo en las empresas, facilitan el análisis y la presentación de los datos.

Plataforma Pentaho Open Source Business Intelligence: La plataforma cubre las más amplias necesidades de análisis de los datos y de los informes empresariales. Las soluciones de Pentaho están escritas en Java y tienen un ambiente de implementación también basado en Java. Esto hace a Pentaho una solución muy flexible para cubrir una amplia gama de necesidades empresariales tanto las típicas como las sofisticadas y específicas al negocio. (Portada sobre la plataforma Pentaho Open Source Business Intelligence, 2006-2011).

Las herramientas de la suite de Pentaho que se utilizan en la solución del trabajo son:

- **Pentaho Bi Server-ce en su versión 4.8.0:** es una aplicación cien por ciento Java que nos permite gestionar todos nuestros recursos de inteligencia de negocio. Cuenta con una interfaz de usuario de inteligencia de negocio donde encontramos disponibles todos nuestros informes, vistas OLAP y cuadros de mando. Así mismo como accesos a una consola de administración que nos permitirá gestionar y supervisar tanto nuestra aplicación como los usuarios, que informen consulta cada usuario, cuando se han consultado, el rendimiento de la aplicación, etc. Proporciona mucha libertad al usuario y los desarrolladores para crear contenidos nuevos. (12)
- **Pentaho Data Integration en su versión 4.2.1:** también conocido como Kettle o PDI, es una poderosa, intuitiva y eficiente herramienta, para la realización de procesos de Extracción, Transformación y Carga. Es multiplataforma, así que trabaja tanto en Windows como en Linux. Es muy gráfica, así que los procesos quedan plasmados en forma de flujos. Lo difícil es conocer el funcionamiento de cada uno de estos componentes de Kettle denominados pasos.

Incluye otras herramientas como son:

Spoon: Programa que permite construir procesos con una interfaz muy gráfica e intuitiva.

Kitchen: Componente que permite ejecutar trabajos por línea de comandos.

Pan: Componente que permite ejecutar transformaciones por línea de comandos. (13)

- **Mondrian Schema Workbench** en su versión 3.2.0: El Mondrian esquema Workbench es una interfaz de diseño que te permite crear y probar esquemas

de cubos OLAP Mondrian visualmente. El motor Mondrian procesa las solicitudes de MDX con la ROLAP (Relational OLAP) esquemas. Estos archivos de esquema son modelos de metadatos XML que se crean en una estructura específica que utiliza el motor de Mondrian. Estos modelos XML pueden ser consideradas estructuras de forma de cubo que utilizan tablas de hechos y de dimensiones. Proporciona las siguientes funcionalidades: editor de esquema integrado con el origen de datos subyacente para su validación y consultas MDX contra el esquema y la base de datos. (14)

La suite Pentaho Open Source Business Intelligence es multiplataforma y de código abierto.

1.11.2 SISTEMA GESTOR DE BASE DE DATOS

Los Sistemas Gestores de Base de Datos (SGBD) están diseñados para gestionar grandes bloques de información. Un SGBD es una aplicación que permite a los usuarios definir, crear, mantener la base de datos y proporcionar un acceso controlado.

PostgreSQL en su versión 9.1: es un sistema de gestión de bases de datos objeto-relacional, distribuido bajo licencia BSD y con su código fuente disponible libremente. Es el sistema de gestión de bases de datos de código abierto más potente del mercado y en sus últimas versiones no tiene nada que envidiarle a otras bases de datos comerciales. PostgreSQL utiliza un modelo cliente/servidor y usa multiprocesos en vez de multihilos para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto y el sistema continuará funcionando.

PostgreSQL 9.1 tiene características técnicas que lo hacen una de las bases de datos más potentes y robustas del mercado. Las características que más se han tenido en cuenta durante su desarrollo son la estabilidad, potencia, robustez, facilidad de administración e implementación de estándares. PostgreSQL funciona muy bien con grandes cantidades de datos y una alta concurrencia de usuarios accediendo a la vez al sistema. (15)

1.11.3 ADMINISTRADOR DE BASES DE DATOS

PgAdmin III en su versión 1.14.1: es una aplicación gráfica para gestionar el gestor de bases de datos PostgreSQL, siendo la más completa y popular con licencia Open Source. Está escrita en C++ usando la librería gráfica multiplataforma wxWidgets, lo que permite que se pueda usar en Linux, FreeBSD, Solaris, Mac OS X y Windows. Es capaz de gestionar versiones a partir de la PostgreSQL 7.3 ejecutándose en cualquier plataforma, así como versiones comerciales de PostgreSQL como Pervasive Postgres, EnterpriseDB, Mammoth Replicator y SRA PowerGres.

PgAdmin III está diseñado para responder a las necesidades de todos los usuarios, desde escribir consultas SQL simples hasta desarrollar bases de datos complejas. El interfaz gráfico soporta todas las características de PostgreSQL y facilita enormemente la administración. La aplicación también incluye un editor SQL con resaltado de sintaxis, un editor de código de la parte del servidor, un agente para lanzar scripts programados, y mucho más. La conexión al servidor puede hacerse mediante conexión TCP/IP o Unix Domain Sockets (en plataformas *nix), y puede encriptarse mediante SSL para mayor seguridad. (16)

1.11.4 HERRAMIENTA DE MODELADO

Visual Paradigm en su versión 8.0: es una herramienta que sirve para realizar modelado UML. Esta herramienta tiene unas características gráficas muy cómodas que facilitan la realización de los diagramas de modelado. Es factible a la hora de dibujar diagramas de clases, y generar script para diferentes Sistemas Gestores de Bases de Datos. Ha sido concebida para soportar el ciclo de vida completo del proceso de desarrollo del software a través de la representación de todo tipo de diagramas. Constituye una herramienta privada disponible en varias ediciones, cada una destinada a satisfacer diferentes necesidades: Enterprise, Professional, Community, Standard, Modeler y Personal. Existe una alternativa libre y gratuita de este software. (17)

1.11.5 CONTENEDOR WEB

Apache Tomcat en su versión 5.5.12: (también llamado Jakarta Tomcat o simplemente Tomcat) funciona como un contenedor de servlets desarrollado bajo el proyecto Jakarta en la Apache Software Foundation. Tomcat implementa las especificaciones de los servlets y de JavaServer Pages (JSP) de Sun Microsystems.

Tomcat puede funcionar como servidor web por sí mismo, hoy en día Tomcat es usado como servidor web autónomo en entornos con alto nivel de tráfico y alta disponibilidad. Dado que Tomcat fue escrito en Java, funciona en cualquier sistema operativo que disponga de la máquina virtual Java. Apache Tomcat se desarrolla en un entorno abierto y participativo y publicado bajo la licencia Apache versión 2. (18)

1.11.6 HERRAMIENTAS PARA LAS PRUEBAS.

Datanamic Data Generator for Postgre SQL en su versión 5.0: es un generador de datos que incluye características útiles para la ejecución de pruebas a bases de datos PostgreSQL entre las que podemos mencionar:

- Acceso directo a bases de datos PostgreSQL (con controlador nativo).
- Rellenar la base de datos directamente o generar scripts de inserción.
- Soporte para PostgreSQL versión 7,8 y 9.
- Incorpora en los generadores más de 40 idiomas.
- Carga datos aleatorios de fuentes tanto externas como internas.
- Previsualización en tiempo real de los datos que se generan.
- Alto rendimiento con miles de inserciones por segundo.
- Configuración para definir el porcentaje de valores nulos. (19)

Apache JMeter 2.9: Apache JMeter puede utilizarse para probar el rendimiento sobre los recursos estáticos y dinámicos (archivos, Servlets, secuencias de comandos Perl, objetos Java, bases de datos y consultas, servidores FTP y mucho más). Puede utilizarse para simular una carga pesada en un servidor, red u objeto a fin de probar su fuerza o analizar el rendimiento general en tipos diferentes de carga. Puedes utilizarlo para realizar un análisis gráfico del rendimiento o probar el comportamiento de tu objeto/secuencia de comandos/servidor bajo carga simultánea. Apache JMeter es una aplicación de escritorio práctica y fiable, diseñada para probar el comportamiento funcional en la carga y medir el rendimiento. (20)

1.12 CONCLUSIONES DEL CAPÍTULO.

En el presente capítulo se hizo referencia a los principales conceptos y características de un almacén de datos y de los mercados de datos. Después de un estudio de las metodologías existentes se decide que la más adecuada a las necesidades del proyecto para el desarrollo de los mercados de datos es la metodología propuesta por

Kimball. Se realizó un estudio de las tendencias tanto nacionales como internacionales de los almacenes de datos. Se abordó sobre las diferentes herramientas y tecnologías a utilizar en el desarrollo del trabajo de diploma.

2. CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DE LOS MERCADOS DE DATOS.

2.1 INTRODUCCIÓN

En el presente capítulo se realizará un estudio preliminar del negocio para identificar los requerimientos de los usuarios. Luego se efectuará un estudio de las características y el estado actual de las fuentes de datos, además se recogen todos los aspectos relacionados con el modelado y diseño del AD. En el mismo se identificarán las tablas dimensiones y las tablas hechos; y se realizarán las uniones entre las mismas para conformar el modelado del AD. De forma general se le dará cumplimiento a las tareas planteadas en la metodología de Kimball.

2.2 PLANIFICACIÓN

Alcance: Se realizaran tres mercados de datos uno para cada uno de los siguientes departamentos: Civil y Familia, Atención a la Población y Administrativo.

2.3 ANÁLISIS DE REQUERIMIENTOS

La primera etapa del desarrollo de la solución se comienza obteniendo los requerimientos de información, ya que es vital conocer cuál es la información que se necesita y como, con el objetivo de identificar la información relevante para los usuarios. Después de un análisis de los reportes perteneciente a los departamentos Civil y Familia, Administrativo y Atención a la Población se identificaron las siguientes necesidades de información:

2.3.1 REQUERIMIENTOS PARA EL MERCADO DEL DEPARTAMENTO CIVIL Y FAMILIA.

- Cantidad de asuntos promovidos según asunto según instancia.
- Cantidad de asuntos dictaminados según asunto según instancia.
- Cantidad de tutelas promovidas ante el tribunal por involucrados según instancia.

- Cantidad de escrituras firmadas por el fiscal en representación por involucrados según instancia.
- Cantidad de recursos de casación en los que participó el fiscal por resolución según instancia.
- Cantidad de tutelas recibidas por involucrados según instancia.
- Cantidad de tutelas denegadas al solicitante por involucrados según instancia.
- Cantidad de privación de la Patria Potestad según interesados.
- Cantidad de suspensión de la Patria Potestad según interesados.
- Cantidad de solicitudes de privación de la Patria Potestad denegadas según interesados.
- Cantidad de solicitudes de suspensión de la Patria Potestad denegadas según interesados.
- Cantidad de demandas de Revisión Civil presentadas en el tribunal.
- Cantidad de demandas de Revisión Civil recibidas en la Fiscalía.
- Cantidad de expropiación forzosa por involucrados según instancia.
- Cantidad de expropiación forzosa por resolución según instancia.
- Cantidad de procesos sucesorios según instancia.
- Cantidad de procesos en los que interviene el fiscal como actor según instancia.
- Cantidad de procesos en los que interviene el Fiscal como actor por asunto según instancia.
- Cantidad de otros procesos en los que interviene el Fiscal en virtud de la Instrucción 216 del TSP según instancia.
- Cantidad de otros procesos en los que interviene el Fiscal en virtud de la Instrucción 216 del TSP en defensa del interés social por asunto según instancia.
- Cantidad de otros procedimientos en los que interviene en defensa del interés social según instancia.
- Cantidad de otros procedimientos en los que interviene en defensa del interés social por asunto según instancia.
- Cantidad de demandas sobre Patria Potestad presentadas al tribunal por el fiscal según resolución.
- Cantidad de demandas sobre suspensión de la Patria Potestad notificadas al fiscal según resolución.

- Cantidad de demandas sobre privación de la Patria Potestad notificadas al fiscal según resolución.

2.3.2 REQUERIMIENTOS PARA EL MERCADO DEL DEPARTAMENTO ATENCIÓN A LA POBLACIÓN.

- Total de personas atendidas.
- Cantidad personas atendidas por color de la piel.
- Cantidad personas atendidas por sexo.
- Cantidad personas atendidas por edad.
- Cantidad de personas atendidas según materia.
- Cantidad de reclamaciones recibidas.
- Cantidad de reclamaciones recibidas por sexo.
- Cantidad de reclamaciones tramitadas.
- Cantidad de reclamaciones tramitadas por sexo.
- Cantidad de reclamaciones tramitadas por solución operativa.
- Cantidad de reclamaciones tramitadas por solución operativa según sexo.
- Cantidad de reclamaciones tramitadas por solución operativa según menores de 18 años.
- Cantidad de reclamaciones tramitadas por solución operativa según materia.
- Cantidad de Reclamaciones tramitadas mediante resolución.
- Cantidad de reclamaciones tramitadas mediante resolución según sexo.
- Cantidad de reclamaciones tramitadas mediante resolución según materia.
- Cantidad de reclamaciones tramitadas mediante resolución cumplidas en término.
- Cantidad de reclamaciones tramitadas mediante resolución pendientes fuera de término.
- Cantidad de reclamaciones tramitadas mediante resolución pendientes fuera de término de mujeres.
- Cantidad de reclamaciones tramitadas mediante resolución pendientes fuera de término según materia.
- Cantidad de reclamaciones tramitadas por Pronunciamientos no Violatorios.
- Cantidad de reclamaciones tramitadas por Orientación Jurídica.
- Cantidad de reclamaciones pendientes.
- Cantidad de reclamaciones pendientes según materia.

- Cantidad de reclamaciones archivadas.
- Cantidad de reclamaciones archivadas según materia.
- Cantidad de reclamaciones trasladadas.
- Cantidad de reclamaciones trasladadas según materia.
- Cantidad de impugnaciones recibidas.
- Cantidad de impugnaciones recibida por involucrados.
- Cantidad de impugnaciones recibida por involucrados según materia.
- Cantidad de impugnaciones resueltas.
- Cantidad de impugnaciones resueltas según involucrados.
- Cantidad de impugnaciones resueltas en el término.
- Cantidad de impugnaciones resueltas en el término según involucrados.
- Cantidad de impugnaciones resueltas fuera del término.
- Cantidad de impugnaciones resueltas fuera del término por involucrados.
- Cantidad de impugnaciones resueltas fuera del término por involucrados según materia.
- Cantidad de impugnaciones pendientes fuera de término.
- Cantidad de impugnaciones ratificadas.
- Cantidad de impugnaciones ratificadas según involucrados.
- Cantidad de impugnaciones revocadas.
- Cantidad de impugnaciones revocadas por involucrados.
- Cantidad de impugnaciones revocadas por involucrados según materia.

2.3.3 REQUERIMIENTOS PARA EL MERCADO DEL DEPARTAMENTO ADMINISTRATIVO.

- Total de procedimientos contenciosos.
- Cantidad de procedimientos contenciosos de vivienda.
- Cantidad de procedimientos contenciosos de vivienda según resolución.
- Cantidad de procedimientos contenciosos de otros.
- Cantidad de procedimientos contenciosos de otros según resolución.
- Cantidad de procedimientos contenciosos según motivo.
- Cantidad de procedimientos administrativos que impugnan ante el TPP.
- Total de procesos contenciosos en los que interviene el fiscal.
- Total de procesos en los que interviene el fiscal como actor.

- Cantidad de procesos contenciosos en los que interviene el fiscal como actor de vivienda.
- Cantidad de procesos contenciosos en los que interviene el fiscal como actor de vivienda según resolución.
- Cantidad de procesos contenciosos en los que interviene el fiscal como actor de otros.
- Cantidad de procesos contenciosos en los que interviene el fiscal como actor de otros según resolución.
- Cantidad de procesos contenciosos en los que interviene el fiscal como actor en según motivo.
- Cantidad por inconformidad con resoluciones administrativas en los que interviene el fiscal como demandado.
- Total de procesos en los que interviene el fiscal como demandado.
- Cantidad de procesos contenciosos en los que interviene el fiscal como demandado de vivienda.
- Cantidad de procesos contenciosos en los que interviene el fiscal como demandado de vivienda según resolución.
- Cantidad de procesos contenciosos en los que interviene el fiscal como demandado de otros.
- Cantidad de procesos contenciosos en los que interviene el fiscal como demandado de otros según resolución.
- Cantidad de procesos contenciosos en los que interviene el fiscal como demandado según motivo.
- Cantidad por inconformidad con resoluciones administrativas en los que interviene el fiscal como demandado.
- Total de recursos de casación en los que participó el fiscal.
- Cantidad de recursos de casación en los que participó el fiscal como recurrente.
- Cantidad de recursos de casación en los que participo el fiscal como recurrente según resolución.
- Cantidad de recursos de casación en los que participó el Fiscal como no recurrente.
- Cantidad de recursos de casación en los que participo el fiscal como no recurrente según resolución.

- Cantidad de recursos de casación en los que participo el fiscal como recurrente según instancia.
- Cantidad de recursos de casación en los que participo el fiscal como no recurrente según instancia.
- Cantidad de revisiones administrativas recibidas en la fiscalía según resolución.

Para el desarrollo de esta fase se realiza una reunión con los analistas del sistema SIGEF, para así determinar la información clave de las Fiscalías en los departamentos Civil y Familia, Administrativo y Atención a la Población. Como resultado se seleccionaron los procesos de negocio y los temas analíticos con sus requerimientos que se muestran en la siguiente tabla.

2.3.4 TABLAS DE TEMAS ANALÍTICOS

Departamento Civil y Familia

Tema Analítico	Proceso de Negocio	Requerimiento pedido o inferido	Comentario
Civil y Familia	Jurisdicción Voluntaria	Análisis de procesos sin litis.	Por asunto e instancia.
	Asuntos Notariales	Análisis de las escrituras y documentos notariales.	Por asunto e instancia.
	Recursos Casación	Análisis de los recursos de casación.	Dada la participación del fiscal por resolución e instancia.
	Recursos Apelación	Análisis de los recursos de apelación.	Dada la participación del fiscal por resolución e instancia.
	Expropiación	Análisis de los procesos	Por resolución, motivos e

	Forzosa	de bienes.	instancia.
	Procesos Sucesorios	Análisis de los procesos sucesorios.	Por los involucrados en el proceso e instancia.
	Procesos Contenciosos	Análisis de los procesos contenciosos.	Dada la participación del fiscal por la resolución, asuntos e instancia.
	Proceso Revisión Civil	Análisis de las revisiones civiles.	Dada las presentadas y recibidas.
	Expediente Prejudicial	Análisis de los expedientes que no han iniciado proceso.	Por los interesados e instancia.

Tabla # 1 Temas analíticos para Civil y Familia.

Tema Analítico	Proceso de Negocio	de Requerimiento pedido o inferido	Comentario
Atención a la Población	Atención a la Población	Análisis de las personas atendidas en la fiscalía.	Por color de piel, sexo, edad, materia e instancia.
	Reclamaciones	Análisis de las reclamaciones recibidas, tramitadas y pendientes en la fiscalía.	Por color de piel, sexo, edad, materia e instancia.
	Impugnaciones	Análisis de las impugnaciones	Por materia, instancia e

		recibidas, tramitadas y pendientes en la fiscalía.	involucrados.
--	--	--	---------------

Tabla # 2 Temas analíticos para Atención a la Población.

Tema Analítico	Proceso de Negocio	de	Requerimiento pedido o inferido	Comentario
Administrativo	Recursos de Casación Administrativo	de	Análisis de los recursos de casación administrativos.	Dada la participación del fiscal por resolución.
	Procedimientos Contenciosos		Análisis de los procedimientos contenciosos administrativos.	Por causa, instancia y resolución dada.
	Procesos Contencioso		Análisis de los procesos y procedimientos contenciosos administrativos.	Por causa, instancia y resolución dada.
	Revisión Administrativa.		Análisis de las revisiones administrativas recibidas.	Por la resolución dada en los procesos.

Tabla # 3 Temas analíticos para Administrativo.

Por otra parte, a partir del análisis se puede construir una herramienta de la metodología denominada matriz de procesos/dimensiones. Esta matriz tiene en sus filas los procesos de negocio identificado, y en las columnas, las dimensiones identificadas.

2.3.5 TABLAS DE MATRICES PROCESOS/DIMENSIONES

Matriz procesos/dimensiones para el mercado del departamento Administrativo.

Procesos de Negocio	Dimensiones					
	Tiempo	Instancia	Resolución	Motivo	Tipo Bien	Actuación Fiscal
Procedimiento Contencioso	X	X	X	X	X	
Recurso Casación	X	X		X		X
Revisiones Administrativas	X	X	X			
Proceso Contencioso	X	X	X	X	X	X

Tabla # 4 Matriz de procesos/dimensiones Administrativo

Matriz procesos/dimensiones para el mercado del departamento Atención a la Población.

Procesos de Negocio	Dimensiones						
	Tiempo	Instancia	Materia	Persona	Término	Pronunciamiento	Involucrado
Atención a la Población	X	X	X	X			
Reclamaciones	X	X	X	X	X	X	
Impugnaciones	X	X	X		X		X

Tabla # 5 Matriz de procesos/dimensiones Atención a la Población

Matriz procesos/dimensiones para el mercado del departamento Civil y Familia.

Procesos de Negocio	Dimensiones							
	Tiempo	Instancia	Resolución	Actuación Fiscal	Motivo Potestad	Involucrado	Interesado	Asunto
Procesos Contenciosos Sumario	X	X						X
Procesos Contenciosos Ordinarios	X	X	X	X				X
Procesos Sucesorios	X	X				X		
Asuntos Notariales	X	X						X
Expediente Prejudicial	X	X			X	X	X	
Jurisdicción Voluntaria	X	X				X		X
Recursos de Apelación	X	X	X	X				
Recursos de Casación	X	X	X	X				
Expropiación Forzosa	X	X				X		
Revisiones Civiles	X	X						

Tabla # 6 Matriz de procesos/dimensiones Civil y Familia

2.4 MODELADO DIMENSIONAL

Para la creación del modelo dimensional, se realizaron las matrices de cada mercado de datos antes descritas, luego se definirán los procesos del negocio a los cuales se le realizaran los mercado de datos y se establece el nivel de granularidad, posteriormente se describirán las tablas dimensiones y hechos, obteniendo con estas el modelo dimensional de alto nivel.

2.4.1 ELEGIR EL PROCESO DE NEGOCIO

Las áreas de proceso a modelar serían los procesos de negocio pertenecientes a los departamentos Civil y Familia, Administrativo y Atención a la Población.

2.4.2 ESTABLECER EL NIVEL DE GRANULARIDAD

A continuación se muestra la granularidad para el mercado de datos realizado para el departamento Administrativo, especificando el nivel de detalle de los datos en una tabla con la descripción de los datos asociados a las dimensiones, donde se explica el tipo de dato y/o los posibles valores que pueden tomar estos. Las tablas para los demás mercados las podemos encontrar en los anexos.

Departamento Administrativo

Nombre de la dimensión.	Nombre del dato.	Tipo de dato.	Valor nulo.	Descripción.
dim_tiempo	id_tiempo	numeric(19,0)	No	Identificador de la tabla tiempo.
	dia	varchar(20,0)	No	Día en que se realiza una acción.
	mes	varchar(20)	No	Mes en que se realiza una acción.
	trimestre	varchar(20)	No	Trimestre en que se realiza una acción.
	semestre	varchar(20)	No	Semestre en que se realiza una acción.
	anno	varchar(20)	No	Año en que se realiza una

				acción.
	num_mes	numeric(3,0)	No	Mes en que se realiza una acción.
	num_trimestre	numeric(3,0)	No	Trimestre en que se realiza una acción.
	num_semestre	numeric(3,0)	No	Número de semestre en que se realiza una acción.
	fecha	date	No	Fecha en que se realiza el proceso.
dim_motivo	id_motivo	numeric(19,0)	No	Identificador de la tabla motivo.
	descripcion	varchar(255)	No	Define el tipo de motivo a ser representado.
dim_instancia	id_instancia	numeric(19,0)	No	Identificador de la tabla instancia.
	descripcion	varchar(255)	No	Define a que nivel de la fiscalía se realiza el proceso.
dim_resolucion	id_resolucion	numeric(19,0)	No	Identificador de la tabla

				resolución.
	descripcion	varchar(255)	No	Define como está declarado el proceso.
dim_tipo_bien	Id_tipo_bien	numeric(19,0)	No	Identificador de la tabla tipo bien.
	descripcion	varchar(255)	No	Define el tipo de bien que se trata.
dim_actuacion_fiscal_rc	id_actuacion_fiscal_rc	numeric(19,0)	No	Identificador de la tabla actuación fiscal_rc.
	descripcion	varchar(255)	No	Define la actuación que tiene el fiscal en los recursos de casación.
dim_actuacion_fiscal_rf	id_actuacion_fiscal_rf	numeric(19,0)	No	Identificador de la tabla actuación fiscal_rf.
	descripcion	varchar(255)	No	Define la actuación que tiene el fiscal en los procesos contenciosos.

Tabla # 7 Nivel de granularidad para dimensiones de Administrativo

2.4.3 IDENTIFICACIÓN DE LAS DIMENSIONES

Las dimensiones definen como están los datos organizados lógicamente y proveen el medio para analizar el contexto del negocio. Estas se obtuvieron mediante el análisis de los reportes de los departamentos Civil y Familia, Administrativo y Atención a la Población, las dimensiones del negocio y de la base de datos del sistema SIGEF.

Dimensiones comunes para todos los mercados de datos.

Dimensión tiempo: representa el espacio de tiempo en el que se está analizando cada acción realizada en la fiscalía, se define esta dimensión con varias categorías o niveles para su mejor organización, estos son año, semestre, trimestre, mes y día.

Dimensión instancia: define en que instancia de las fiscalías existentes de nuestro país se están obteniendo los datos, las cuales serían Fiscalía General de la República, Provincial y Municipal.

Para el mercado de datos de Administrativo

Dimensión resolución: define cual fue la resolución dada al proceso llevado en la fiscalía en las diferentes instancias las cuales serían Con Lugar, Con Lugar en parte, Sin lugar e Inadmisible.

Dimensión motivo: define quien será representado en el caso, puede ser: un menor, un incapacitado, un ausente o en interés social.

Dimensión tipo bien: define que bien es el que se está representado en el proceso, puede ser vivienda u otro.

Dimensión actuación del fiscal procesos contenciosos: define el tipo de actuación que tiene el fiscal en los procesos contenciosos puede ser demandado.

Dimensión actuación del fiscal en lo recursos de casación: define el tipo de actuación que tiene el fiscal en los recursos de casación puede ser recurrente o no recurrente.

Para el mercado de datos Atención a la Población

Dimensión materia: define de que materia se realiza un proceso determinado.

Dimensión materia atención: define de que materia se realiza el proceso de atención a la población.

Dimensión persona: define los datos relacionados a las personas atendidas en la fiscalía.

Dimensión involucrado: representa personas o agentes que interviene en las impugnaciones, pueden ser promoventes o entidades infractoras.

Dimensión respuesta impugnación: representa las posibles respuestas de las impugnaciones, pueden ser ratificadas o revocadas.

Dimensión pronunciamiento: representa los pronunciamientos por las que se puede tramitar las reclamaciones, pueden ser solución operativa, pronunciamiento no violatorios, orientación jurídica o por resolución.

Dimensión decisión adoptada: representa las decisiones adoptadas en las diferentes reclamaciones, pueden ser archivadas o trasladadas.

Dimensión término: representa si las reclamaciones o impugnaciones están en término o fuera de término.

Para el mercado de datos Civil y Familia

Dimensión interesado: representa agentes que interviene en los expedientes prejudiciales, pueden ser instituciones o particulares.

Dimensión resolución: define cual fue la resolución dada al proceso llevado en la fiscalía en las diferentes instancias las cuales serían Con Lugar, Con Lugar en parte, Sin lugar e Inadmisible.

Dimensión involucrado: representa las personas que van a ser representadas que pueden ser menores de edad, incapacitados, propietario ausente, representación del estado.

Dimensión fiscal recurrente: define el tipo de actuación que tendrá el fiscal en el caso.

Dimensión asunto proceso contencioso sumario: define el tipo de asunto que tratará el proceso contencioso sumario.

Dimensión asunto proceso contencioso ordinario: define el tipo de asunto que tratará el proceso contencioso ordinario.

Dimensión asunto función notarial: representa la descripción del asunto de función notarial que puede ser declaratoria de herederos o perpetua memoria.

Dimensión asunto jurisdicción voluntaria : representa el tipo de asunto que puede ser incapacidad, utilidad y necesidad, tutela, adopción, autorización excepcional para contraer matrimonio, alimento, declaración de ausente, presunción de muerte.

Dimensión motivo patria potestad: define si la patria potestad fue suspendida o privada.

Dimensión escrituras: representa el tipo de escrito que se realiza en los asuntos notariales tratados.

Dimensión adopción: representa si la adopción fue nacional o internacional.

2.4.4 IDENTIFICAR LAS TABLAS DE HECHOS Y MEDIDAS

Los hechos contendrán datos numéricos que se pueden resumir para proporcionar información sobre el historial de las operaciones del negocio. Además de las claves primarias de las tablas de dimensiones asociadas a esta, ofreciéndole al sistema un acceso a la información almacenada en la tabla de hechos, mucho más rápida y eficiente. Estas tablas se obtuvieron mediante el análisis de los reportes de los departamentos Civil y Familia, Administrativo y Atención a la Población y la base de datos de SIGEF.

Hechos

Para el mercado de datos de Administrativo

Hecho Recurso de Casación Administrativo: recogerá todas las medidas relacionadas con los recursos de casación administrativo y participación del fiscal en ellos.

Hecho Procedimiento Contencioso: tendrá todas las medidas relacionadas con la resolución en los procedimientos contenciosos en las diferentes instancias.

Hecho Procesos Contencioso: contendrá las medidas relacionadas con la resolución en los procesos contenciosos y otros procedimientos administrativos que interviene el fiscal en las diferentes instancias.

Hecho Revisión Administrativa: recogerán todas las medidas relacionadas con las revisiones administrativas.

Para el mercado de datos de Atención a la Población

Hecho Atención a la Población: tendrá las medidas relacionadas con la atención a la Población en las diferentes instancias.

Hecho Reclamaciones: guardara todas las medidas relacionadas con los diferentes tipos de reclamaciones tramitadas realizadas a la fiscalía.

Hecho Impugnaciones: tendrá todas las medidas relacionadas con las impugnaciones y su estado en la fiscalía.

Para el mercado de datos de Civil y Familia

Hecho Jurisdicción Voluntaria: recogerá todas las medidas relacionadas con los asuntos en los procesos de jurisdicción voluntaria en las diferentes instancias como las tutelas.

Hecho Asuntos Notariales: tendrá las medidas de las escrituras y documentos notariales según asuntos tratados en la fiscalía

Hecho Recursos Casación: contendrá todas las medidas relacionadas con la resolución y participación del fiscal en los recursos de casación en las diferentes instancias.

Hecho Recursos Apelación: contendrá todas las medidas relacionadas con la resolución y participación del fiscal en los recursos de apelación en las diferentes instancias.

Hecho Expropiación Forzosa: recogerá las medidas relacionadas con los involucrados y resolución en la expropiación forzosa en las diferentes instancias.

Hecho Procesos Sucesorios: tendrá las medidas relacionadas con los involucrados en los procesos sucesorios en las diferentes instancias.

Hecho Procesos Contenciosos Ordinarios: recogerá todas las medidas relacionadas con la resolución y los asuntos tratados en los procesos contenciosos ordinarios por el fiscal en las diferentes instancias.

Hecho Procesos Contenciosos Sumarios: recogerá todas las medidas relacionadas con la resolución y los asuntos tratados en los procesos contenciosos sumarios por el fiscal en las diferentes instancias

Hecho Proceso Revisión Civil: contendrá todas las medidas relacionadas con las demandas de la revisión civil.

Hecho Expediente Prejudicial: recogerá todas las medidas relacionadas con los expedientes y sus involucrados en las diferentes instancias.

2.4.5 IDENTIFICACIÓN DE ATRIBUTOS DE DIMENSIONES Y TABLAS DE HECHOS

A continuación se muestran todas las tablas de dimensiones y de hechos con sus respectivos atributos para cada uno de los mercados de datos.

Administrativo

Nombre de la dimensión	Atributos
dim_tiempo	id_tiempo, num_trimestre, dia, mes, num_mes, anno, trimestre, num_dia, fecha
dim_motivo	id_motivo, descripcion
dim_instancia	id_instancia, descripcion
dim_resolucion	id_resolucion, descripcion
dim_tipo_bien	id_tipo_bien, descripcion
dim_actuacion_fiscal_rf	id_actuacion_fiscal_rf, descripcion
dim_actuacion_fiscal_rc	id_actuacion_fiscal_rc, descripcion

Tabla # 8 Lista de atributos de las tablas de dimensiones de Administrativo.

Nombre del hecho	Atributos
hecho_procesos_contenciosos	id_proceso_contencioso, id_tiempo, id_motivo, id_tipo_bien, id_instancia, id_resolucion, id_actuacion_fiscal_rf, cant_inconformidad

	cant_procesos_contencioso,
hecho_procedimientos_contenciosos	id_procedimiento_contencioso, id_resolucion, id_tipo_bien, id_motivo, id_instancia, id_tiempo,cant_impugnan, cant_procedimientos_contenciosos
hecho_recurso_casacion	id_recurso_casacion, id_tiempo, id_instancia, id_actuacion_fiscal_rc, id_resolucion, cant_recurso_casacion
hecho_revisiones_administrativas	id_revisiones_administrativas, id_tiempo, id_resolucion, cant_revision_administrativa

Tabla # 9 Lista de atributos de las tablas de hechos de Administrativo.

Atención a la Población

Nombre de la dimensión	Atributos
dim_tiempo	id_tiempo, num_trimestre, dia, mes, num_mes, anno, trimestre, num_día, fecha
dim_instancia	id_instancia, descripcion
dim_materia	id_materia, descripcion
dim_materia_ap	id_materia_ap, descripcion_materia
dim_involucrado	id_involucrado, tipo_involucrado
dim_respuesta_impugnacion	id_respuesta_impugnacion, descripcion
dim_termino	Id_termino, descripcion
dim_decision_adoptada	id_decision_adoptada, descripcion
dim_pronunciamiento	id_pronunciamiento, descripcion
dim_persona	id_persona, edad, sexo, raza

Tabla # 10 Lista de atributos de las tablas de dimensiones de Atención a la Población.

Nombre del hecho	Atributos
hecho_atencion_poblacion	id_atencion_poblacion, id_tiempo, id_instancia, id_persona, id_materia_ap, cant_personas_atendidas
hecho_impugnaciones	id_hecho_impugnaciones, id_involucrado, id_materia, id_tiempo, id_instancia, id_respuesta_impugnacion, id_termino, cant_impugnaciones, cant_impugnaciones_resultas
hecho_quejas_reclamaciones	id_quejas_reclamaciones, id_persona, id_materia_ap, id_instancia, id_tiempo, id_termino, id_pronunciamiento, id_decision_adoptada, cant_reclamaciones

Tabla # 11 Lista de atributos de las tablas de hechos de Atención a la Población.

Civil y Familia

Nombre de la dimensión	Atributos
dim_tiempo	id_tiempo, num_trimestre, dia, mes, num_mes, anno, trimestre, num_dia, fecha
dim_instancia	id_instancia, descripcion
dim_involucrado	id_involucrado, descripcion
dim_resolucion	id_resolucion, descripcion
dim_interesado	id_interesado, descripcion
dim_fiscal_recurrente	id_fiscal_recurrente, descripcion
dim_asunto_pccsumario	id_asunto_pccsumario, descripcion
dim_asunto_pccordinario	id_asunto_pccordinario, descripcion

dim_asunto_notarial	id_asunto_notarial, descripcion
dim_asunto_jurisdiccion	id_asunto_jurisdiccion, descripcion
dim_motivo_patriap	id_motivo_patriap, descripcion
dim_escrituras	id_escrituras, descripcion
dim_adopcion	id_adopcion, descripcion

Tabla # 12 Lista de atributos de las tablas de dimensiones de Civil y Familia.

Nombre del hecho	Atributos
hecho_procesos_csumario	id_procesos_csumario, id_tiempo, id_instancia, id_asunto_pcsumario, cant_procesos_su, cant_proceso_su_interes
hecho_procesos_cordinarios	id_procesos_cordinarios, id_asunto_pcordinario, id_tiempo, id_instancia, id_resolucion, cant_procesos_or_interes, cant_procesos_or,
hecho_procesos_sucesorios	id_procesos_sucesorios, id_tiempo, id_involucrado, id_instancia, cant_procesos_sucesorios
hecho_asuntos_notariales	id_asuntos_notariales, id_tiempo, id_asunto_notarial, id_instancia, id_involucrado, id_escrituras, cant_escrituras
hecho_revisiones_civiles	id_revisiones_civiles, id_instancia, id_tiempo, cant_revision, cant_revision_tp
hecho_recursos_apelacion	id_recursos_apelacion, id_tiempo, id_resolucion, id_fiscal_recurrente, id_instancia, cant_recusos_apelacion
hecho_recursos_casacion	id_recursos_casacion, id_fiscal_recurrente, id_tiempo, id_instancia, id_resolucion, cant_recursos_cansacion
hecho_expediente_prejudicial	id_expediente_prejudicial, id_interesado, id_tiempo, id_motivo_patriap, cant_patria_potestadP,

	id_instancia, id_involucrado, cant_expedientes, cant_patria_potestad
hecho_jurisdiccion_voluntaria	id_jurisdiccion_voluntaria, id_involucrado, id_tiempo, id_adopcion, id_asunto_jurisdiccion_voluntaria, id_instancia, cant_jurisdiccion
hecho_expropiacion_forzosa	id_expropiacion_forzosa, id_tiempo, id_instancia, id_involucrado, cant_expropiacion_forzosa

Tabla # 13 Lista de atributos de las tablas de hechos de Civil y Familia.

2.5 PATRONES DE DISEÑO DE BASES DE DATOS PARA EL MD

Los patrones de diseños brindan soluciones estándar a problemas comunes de programación, también constituyen formas de un diagrama de objeto o de un modelo de objeto y pueden describirse como técnicas para flexibilizar los códigos haciéndolo satisfacer ciertos criterios. Los patrones de diseño de una Base de Datos permiten al usuario crear una BD más fortalecida. En la actualidad las bases de datos suelen ser muy grandes y a veces el trabajo con los patrones de diseño hacen que el diseño sea más fácil y que se asegure un resultado satisfactorio.

2.5. 1 PATRÓN DE LLAVES SUBROGADAS

Las llaves subrogadas son de suma importancia para la implementación de los mercados de datos. Las llaves subrogadas son aquellas que se definen artificialmente, son de tipo numérico secuencial, no tienen relación directa con ningún dato y no poseen ningún significado en especial.

Esta clave subrogada debe ser el único campo que sea clave principal de cada tabla de dimensión. Una forma de implementación sería, a través de la utilización de herramientas ETL, mantener una tabla que contenga la clave primaria de la tabla del OLTP y la clave subrogada correspondiente a la dimensión del MD. (21)

El patrón de llaves subrogadas plantea que se genere una llave primara única para cada entidad, en vez de usar un atributo identificador en el contexto dado. Esto permite que las tablas sean más fáciles de consultar a partir del identificador, pues todos tienen el mismo tipo en cada una de las tablas. Por ello cada tabla incluyendo

las tablas hechos tendrá una llave única que la identifique, en caso de las tablas de hechos la llave única estará integrada por las llaves primarias de las tablas de dimensiones que se relacionan con la tabla de hecho.

2.5.2 PATRÓN CONSTELACIÓN DE HECHOS

Se define utilizar el patrón de diseño constelación de hechos para la solución propuesta debido a las necesidades de información. Se selecciona este esquema porque algunas tablas de dimensiones se relacionan con más de una tabla de hecho y los mercados de datos presentan más de un hecho.

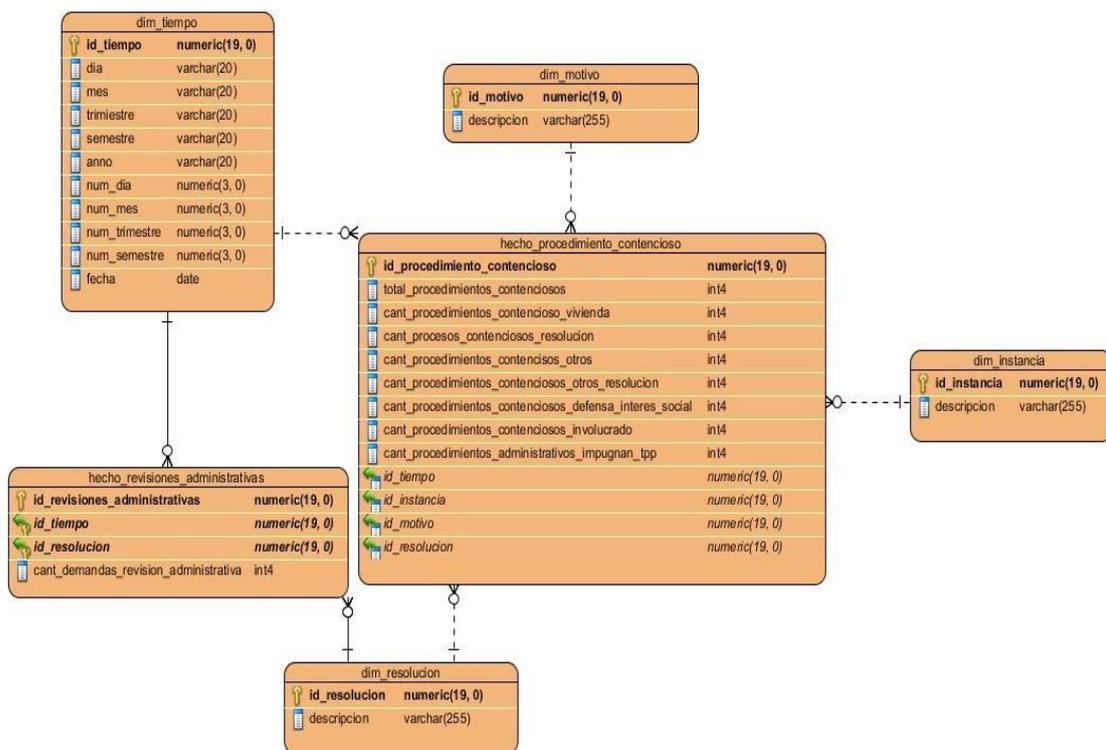


Figura # 7 Esquema representativo de constelación de hechos.

2.6 ESTANDARIZACIÓN DEL CÓDIGO.

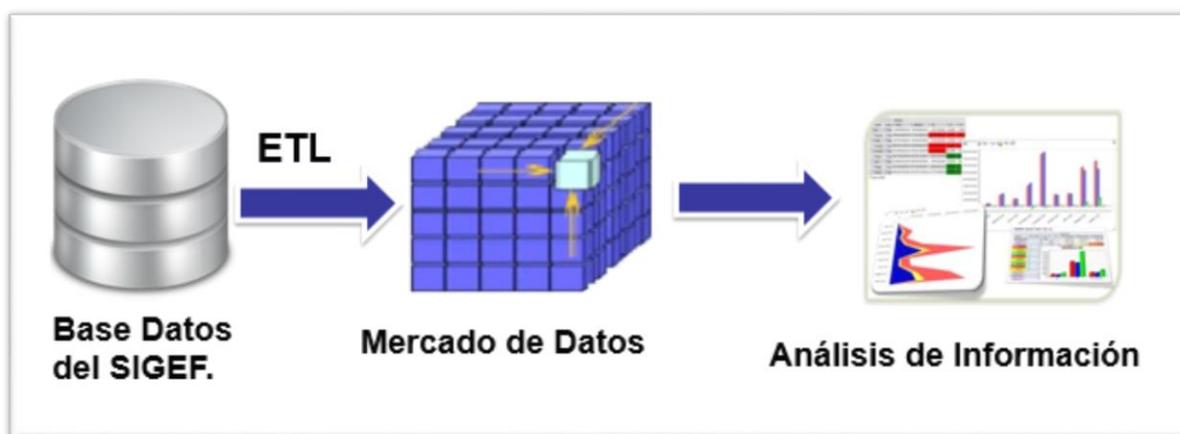
Los estándares de codificación sirven como punto de referencia para los desarrolladores, manteniendo un estilo único de programación. Ayudan a mejorar el proceso de codificación haciéndolo más eficiente en especial cuando de trabajo en grupo se trata.

Para el desarrollo de los mercados se definieron los siguientes estándares de codificación:

- ✓ Los nombres de las entidades y atributos serán con minúscula, separando por “_” cuando exista composición de palabras.
- ✓ Las palabras que tenga ñ esta serán tratadas como “nn”.
- ✓ Las tablas hechos comenzarán con la palabra “hecho”.
- ✓ Las tablas dimensiones comenzarán con “dim”.
- ✓ Los identificadores de las tablas comenzarán su nombre con “id”.
- ✓ Las palabras que lleven tilde se pondrán sin estas en las tablas.

2.8 ARQUITECTURA DE LOS MERCADOS DE DATOS

Teniendo en cuenta las características generales de un mercado de datos se definen los componentes que intervienen en la arquitectura o ambiente. En el siguiente gráfico se muestra la estructura que tendrá el almacenamiento de datos para cada uno de los mercados de datos realizados.



Figura# 8: Arquitectura de los mercados de datos a desarrollar.

El ambiente está formado por diversos elementos que interactúan entre sí y que cumplen una función específica dentro del sistema. Básicamente la forma de operar del esquema superior se resume de la siguiente manera:

- Los datos son extraídos de la base datos SIGEF.
- Los datos son transformados, limpiados e integrados para luego ser cargados en el mercado de datos. Este proceso conocido como ETL es realizado con la herramienta Pentaho Data Integration (PDI).
- La información del mercado de datos es almacenada en cubos multidimensionales.

2.9 DISEÑO FÍSICO DE LAS BASES DE DATOS.

Atención a la Población

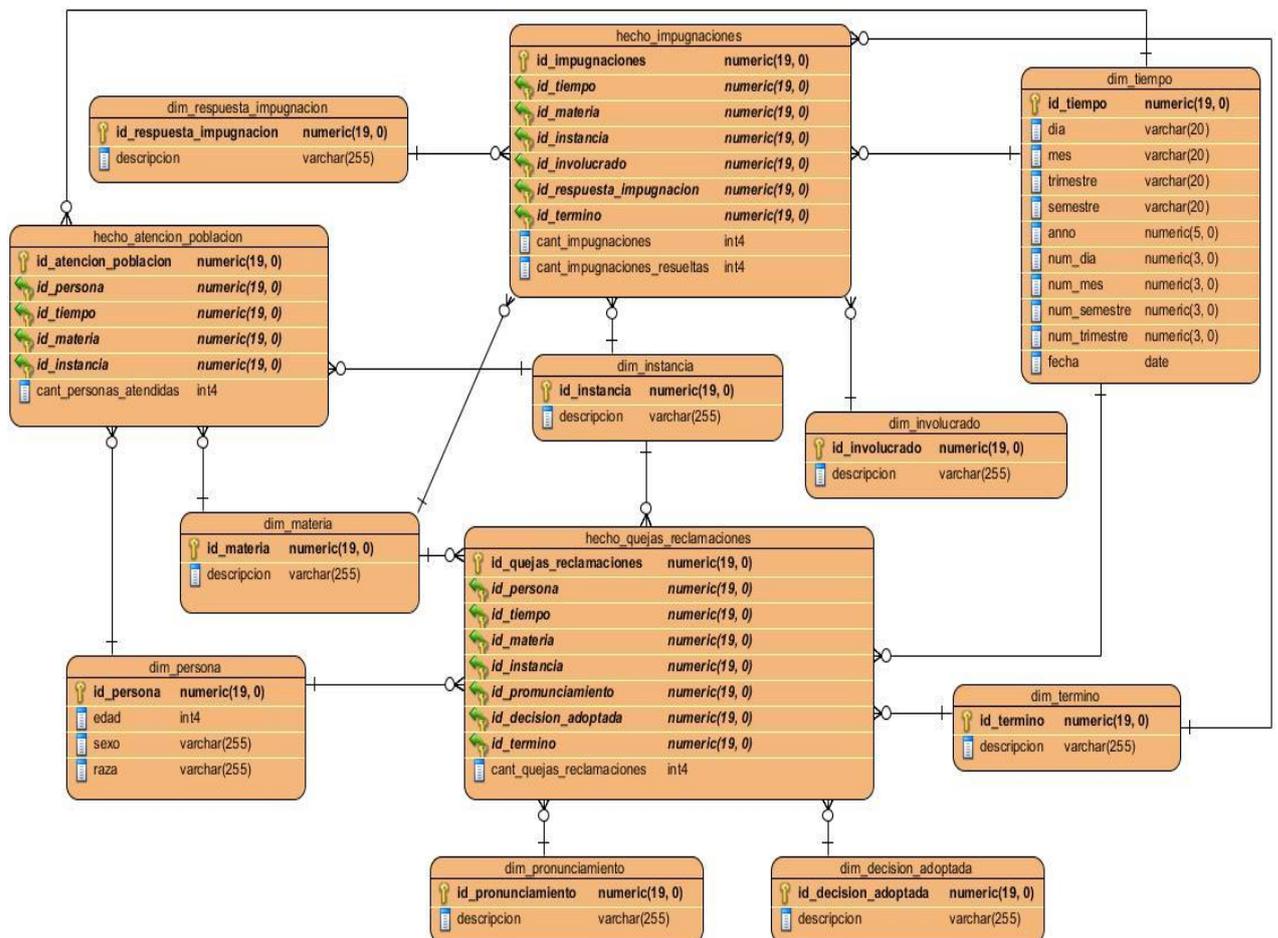


Figura # 9 Diseño físico para Atención a la Población.

Civil y Familia

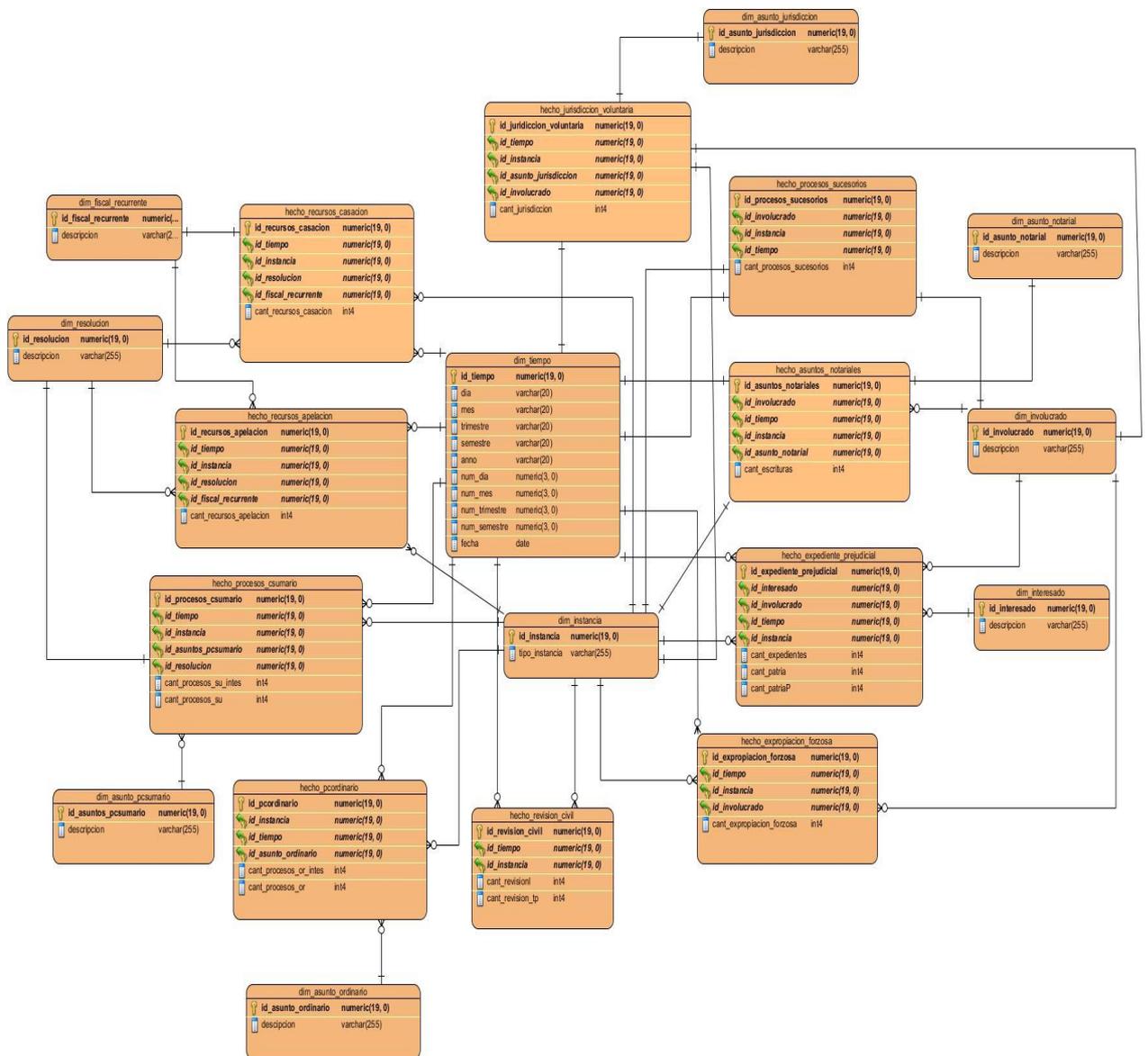


Figura # 10 Diseño físico para Civil y Familia.

Administrativo

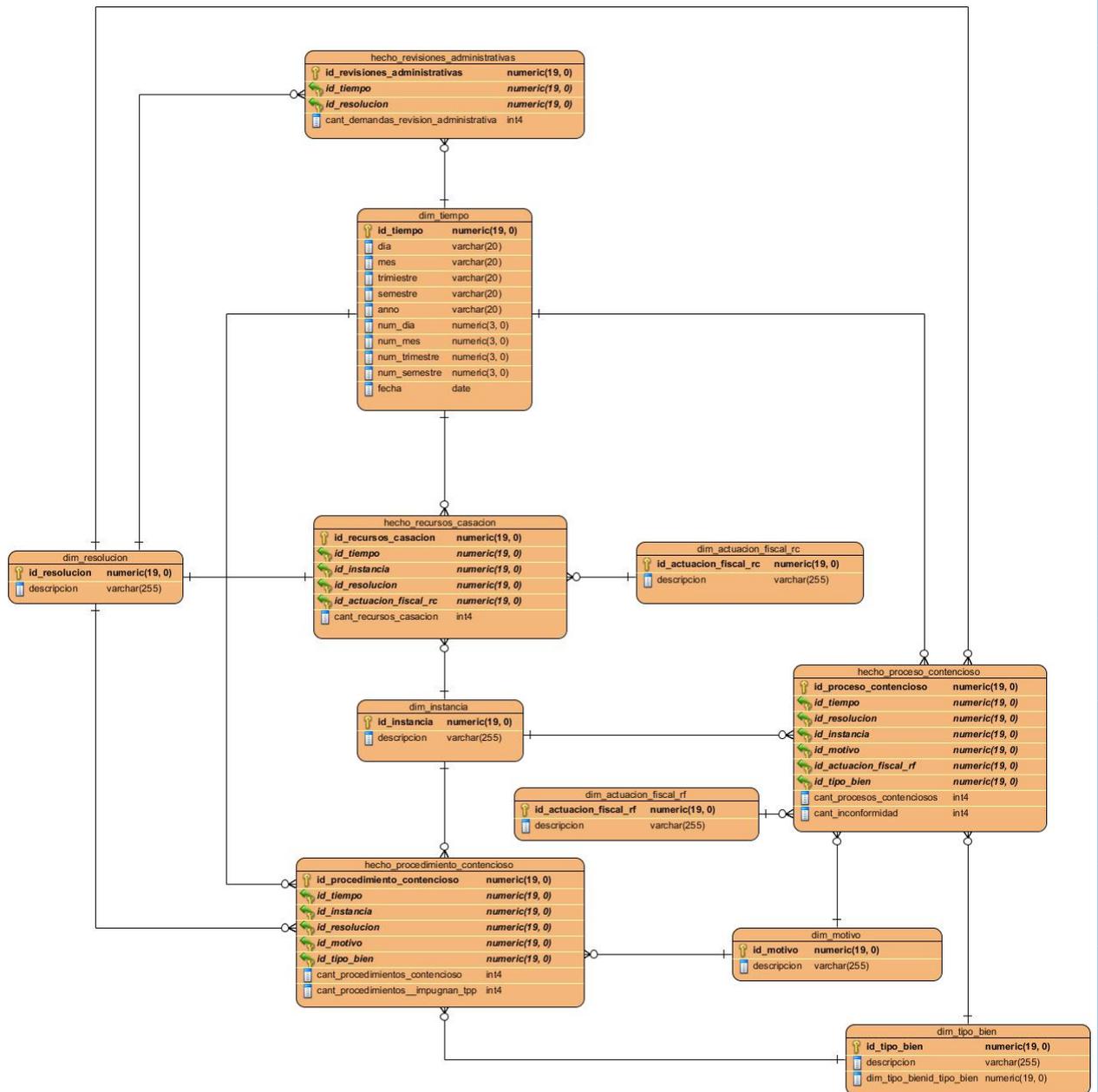


Figura # 11 Diseño físico para Administrativo.

2.10 IMPLEMENTAR EL MODELO DIMENSIONAL DETALLADO

Con el objetivo de lograr una mejor organización en el trabajo, evitar las pérdidas de datos y posibilitar una buena ejecución de los pasos siguientes, se realiza un mapeo de datos, indicando de donde se extrajeron los datos en la BD del SIGEF y hacia donde se trasladaron en los mercados de datos Civil y Familia, Administrativo y

Atención a la Población; detallando los campos fuentes y destino, así como el tipo de dato, del mismo se presentan un ejemplo de algunas de las dimensiones del mercado de datos del departamento Administrativo, encontrándose el resto en los Mapas lógicos de datos de cada uno de los mercados de datos.

Fuente de datos			Destino		
Denominación de la fuente de datos:			Denominación del destino de los datos:		
Tabla	Campo	Tipo de dato	Tabla	Campo	Tipo de dato
base.dproceso	fecha_creacion	date	dim_tiempo	fecha	date
base.dproceso	fecha_creacion	date	dim_tiempo	num_dia	numeric(3,0)
base.dproceso	fecha_creacion	date	dim_tiempo	num_mes	numeric(3,0)
base.dproceso	fecha_creacion	date	dim_tiempo	num_trimestre	numeric(3,0)
administrativo.ntipobien	id_tipo	numeric(19,0)	dim_tipo_bien	id_tipo_bien	numeric(19,0)
administrativo.ntipobien	descripcion	varchar(255)	dim_tipo_bien	descripcion	varchar(255)
base.nfallotribunal	id_fallo	numeric(19,0)	dim_resolucion	id_resolucion	numeric(19,0)
base.nfallotribunal	descripcion	varchar(255))	dim_resolucion	descripcion	varchar(255)
administrativo.nmotivoradicacion	id_motivo	numeric(19,0)	dim_motivo	id_motivo	numeric(19,0)
administrativo.nmotivoradicacion	descripcion	varchar(255)	dim_motivo	descripcion	varchar(255)

Figura # 12 Mapeo de datos.

2.11 PRUEBA Y REVISIÓN DEL MODELO

Un vez que tenemos confianza plena en el modelo se revisa el mismo con diferentes audiencias, cada una con diferentes conocimientos técnicos y del negocio. Se revisó con los analistas de los módulos a trabajar y con el usuario.

Se realizan pruebas al modelo haciendo preguntas para ver si se pueden responder con el modelo realizado, un ejemplo para el modelo de Administrativo sería:

¿Cuántas revisiones administrativas se han realizado con lugar en parte?

¿Cuántas revisiones administrativas se han realizado sin lugar?

¿Cuántos recursos de casación sin lugar se han llevado a cabo en la fiscalía general?

¿Cuántos procesos contenciosos han sido en representación de menores?

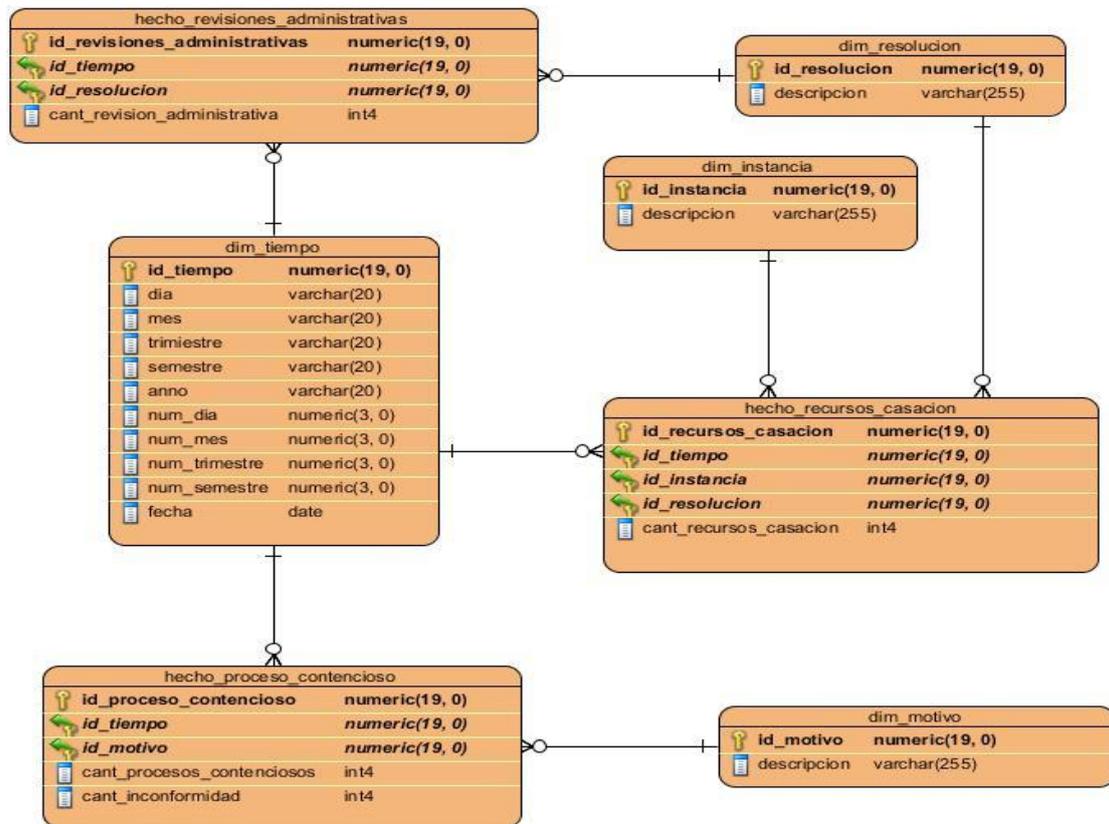


Figura # 13 Fragmento del diseño físico de Administrativo

Como se puede observar en la ilustración la tabla hecho_revisiones_administrativas tiene relaciones con la tabla dim_resolucion además contiene las medida cantidad de revisiones administrativas (cant_revisiones_administrativas).

La tabla hecho_recurso_casacion tiene relación con las tablas dim_instancia, dim_resolucion y tiene la medida cantidad de recursos de casación (cant_recurso_casacion). La tabla hecho_proceso_contencioso se relaciona con dim_motivo y tiene la medida cantidad de procesos contenciosos (cant_procesos_contenciosos). Por lo que se define que el modelo está apto para responder a las preguntas.

2.12 DISEÑO DE LOS SISTEMAS DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE LOS DATOS.

El proceso ETL (Extracción, Transformación, Carga) organiza el flujo de los datos entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un mercado de datos. La

extracción consiste en leer los datos de la base de datos original. La transformación se refiere al hecho de limpiar la información incoherente o no válida, adaptar los tipos de datos y los valores de los campos. Y por último la carga se refiere al hecho de guardar los datos en el formato adecuado dentro del mercado de dato.

Antes de comenzar el diseño de las transformaciones de los datos fue necesario realizar una configuración de la conexión a la base de datos fuente. La cual se puede observar en el Anexo 1.

El próximo paso sería el diseño de la transformación donde se convierten los datos inconsistentes en un conjunto de datos compatibles para que puedan ser cargados en los mercados de datos. Entre los pasos más comunes que se realizaron se encuentran: entrada de tabla, búsquedas en base de datos e insertar/actualizar. En los anexos del 2 al 5 se puede observar los mismos. A continuación se muestra el diseño de la transformación para la dimensión instancia (Figura # 13) siendo está una de las más importantes. Además se muestra la transformación del hecho revisiones administrativas (Figura # 14).



Figura # 14 Transformación para la dimensión instancia.



Figura # 15 Transformación para el hecho revisiones administrativas.

Las transformaciones para los demás hechos y dimensiones se pueden encontrar en los anexos del 6 al 10.

Basándose en las necesidades de información de la fiscalía, se explora la fuente de datos disponible, y se extrae la información que se considere relevante. Para la realización de este proceso se crearon un conjunto de consultas SQL que dan cumplimiento a este objetivo. Las mismas se incorporan al componente entrada de

tabla de Kettle. El conjunto de consultas SQL incorporadas en el paso entrada para cada una de las transformaciones se pueden consultar en los anexos del 21 al 23.

No se realizó una transformación para la dimensión tiempo sino que se crearon tres funciones que al ejecutarlas se llenan las tablas dim_tiempo de los mercados de datos con el rango de tiempo que se define. A la hora de realizar las transformaciones para los hechos se incorpora un paso llamado búsqueda de tiempo que es el encargado de buscar el id_tiempo que le corresponde a una fecha determinada.

2.13 AUTOMATIZACIÓN DEL PROCESO ETL

Teniendo en cuenta la necesidad del cliente en los departamentos, se propone que los procesos de extracción, transformación y carga se realicen simultáneamente y de forma automática, teniendo como objetivo añadir a los mercados de datos aquellos datos que fueron generados después de la última actualización. La misma se llevó a cabo mediante la programación de los trabajos diseñados. Se creó un disparador que permite ejecutar los trabajos de forma automática todos los días primero de cada mes, guardándose en un log los resultados para después ser verificados por el administrador de base datos.

En la Figura # 14 se puede observar un ejemplo de trabajo, en este caso el realizado para revisiones administrativas. Las imágenes de los trabajos restantes se pueden observar del anexo 11 al 13.



Figura # 16 Trabajo Revisiones Administrativas

2.14 DISEÑO DE LAS DIMENSIONES Y CUBOS DE INFORMACIÓN

Una vez terminado el proceso de extracción, transformación y carga de datos a los mercados de datos se procede a realizar el diseño de las dimensiones y los cubos de información en la herramienta Schema-Workbench estableciendo una conexión a los mercados de datos (anexo 14).

Se comienza diseñando las dimensiones definiendo en estas las jerarquías, niveles y propiedades. Las dimensiones no se crean en un cubo específicamente, las mismas

se crean fuera, para que otros cubos puedan utilizarlas de ser necesario. Luego se diseñan los cubos multidimensionales y dentro de estos las medidas y se hace referencia a las dimensiones que utiliza.

En la figura 15 se puede ver una imagen de la creación multidimensional en el Schema-Worbench para el mercado del departamento Administrativo, la imagen de la creación multidimensional de los demás mercados se puede observar en los anexos 14 y 15.

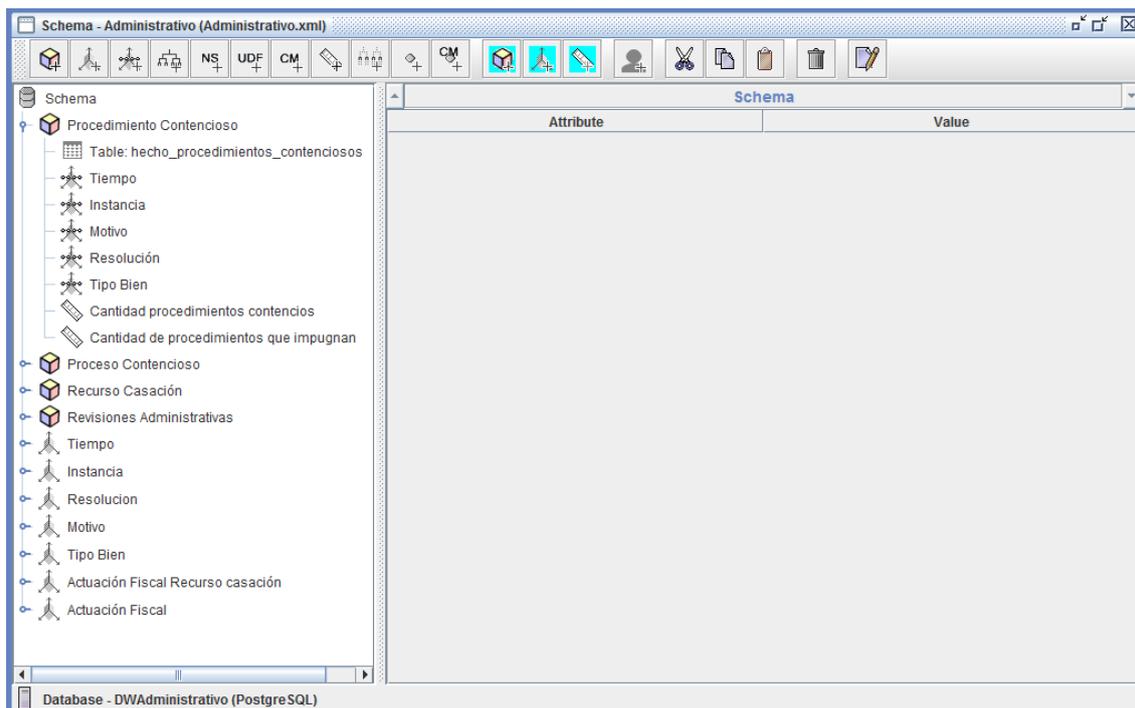


Figura # 17 Esquema multidimensional para el mercado Administrativo.

2.15 VISUALIZACIÓN DE LOS RESULTADOS

A cada cubo se le realiza una consulta MDX para especificar que dimensiones y medidas se deben mostrar. A continuación se muestra como se visualizan estos datos en el navegador en el mercado de datos Administrativo para el cubo Procesos Contenciosos.

Tiempo	Instancia	Motivo	Resolucion	Tipo Bien	Measures			
[-] Todos	[+] Todos	[+] Todos	[+] Todos	[+] Todos	● Cantidad procedimientos contenciosos 2			
[-] 2014	[-] Todos	[+] Todos	[+] Todos	[+] Todos	2			
		Fiscalia General	[+] Todos	[+] Todos	[+] Todos	2		
			Representación de menores	[+] Todos	[+] Todos	1		
				Con lugar	[+] Todos	Vivienda	1	
			Representación por interés social	[+] Todos	[+] Todos	1		
		Sin lugar		[+] Todos	Vivienda	1		
		[+] Trimestre 1	[-] Todos	[+] Todos	[+] Todos	[+] Todos	2	
				Fiscalia General	[+] Todos	[+] Todos	[+] Todos	2
					Representación de menores	[+] Todos	[+] Todos	1
						Con lugar	[+] Todos	Vivienda
Representación por interés social	[+] Todos				[+] Todos	1		
	Sin lugar			[+] Todos	Vivienda	1		

Figura # 18 Visualización de los datos del cubo Procesos Contenciosos para el mercado Administrativo.

Conclusiones del capítulo

En este capítulo se realizó un análisis de los diferentes requerimientos que existen en el negocio. Se detallaron aspectos relacionados con el diseño como la arquitectura que tendrá la solución, los estándares para el diseño y patrones concluyendo utilizar el patrón llave subrogada. Se identificaron 29 dimensiones y 17 hechos, además de las uniones establecidas entre las mismas que dieron como resultado el diseño de los mercados de datos. Permitiendo la implementación del proceso de extracción, transformación y carga de los datos. Se explicó cómo se realiza el diseño del modelo multidimensional y cómo este es mostrado en reportes que responden a los requerimientos del cliente.

3. CAPÍTULO 3 PRUEBAS DE VALIDACIÓN

3.1 INTRODUCCIÓN

Dentro del desarrollo de una solución informática la realización de las pruebas es un elemento importante para garantizar su éxito. El mecanismo que se seguirá para ello en los mercados de datos será la realización de pruebas de carga y estrés, además de las pruebas de volumen y carga, las cuales validaran la utilización de los mercados de datos. En este punto se analiza el rendimiento de los sistemas que se ha construido, al

dar respuesta a distintos pedidos de información accediendo a la base de datos que se encuentra en el servidor PostgreSQL.

3.2 PRUEBA DE CARGA Y ESTRÉS

La prueba de carga y estrés es para validar y valorar la aceptabilidad de los límites operacionales de un sistema bajo carga de trabajo, enfocada a evaluar cómo el sistema responde bajo condiciones anormales. Para realizar la prueba se utiliza la herramienta JMeter en su versión 2.9. La herramienta posee dos tipos de generación de carga, indirecta a través de una aplicación y directa que se basa fundamentalmente en la ejecución de consultas grabadas en la taza o log del servidor de base de datos. La que se utiliza para las pruebas de los mercados de datos es la directa.

A continuación se muestra la cantidad de usuarios que tendrán los mercados de datos en los distintos niveles de la FGR.

Instancia	Cantidad de usuarios	
Fiscalía general de la república	25 (real)	50(prueba)
Fiscalías provinciales	12(real)	25(prueba)

Tabla # 14 Cantidad de usuarios en las diferentes instancias de las fiscalías.

Es preciso definir un rango de tiempo aceptable para las consultas, el cual queda definido de la siguiente forma:

Tipo de respuesta	Tiempo
Aceptable	Menor o igual a los 5 segundos
No aceptable	Mayor a los 5 segundos

Tabla # 15 Tiempo de respuestas de las consultas.

Las pruebas consisten en realizar una consulta y ver los tiempos de respuestas que se tienen para varios usuarios conectados concurrentemente, las consultas utilizadas para los mercados de datos se muestran a continuación:

Administrativo

Select

```
hecho_procesos_contenciosos.id_proceso_contencioso
FROM
hecho_procesos_contenciosos
INNER JOIN dim_tiempo on (hecho_procesos_contenciosos.id_tiempo = dim_tiempo.id_tiempo)
INNER JOIN dim_motivo on (hecho_procesos_contenciosos.id_motivo = dim_motivo.id_motivo)
INNER JOIN dim_resolucion on (hecho_procesos_contenciosos.id_resolucion =
dim_resolucion.id_resolucion)
INNER JOIN dim_instancia on (hecho_procesos_contenciosos.id_instancia = dim_instancia.id_instancia)
INNER JOIN dim_tipo_bien on (hecho_procesos_contenciosos.id_tipo_bien = dim_tipo_bien.id_tipo_bien)
INNER JOIN dim_actuacion_fiscal_rf on (hecho_procesos_contenciosos.id_actuacion_fiscal_rf =
dim_actuacion_fiscal_rf.id_actuacion_fiscal_rf)
```

Atención a la Población

Select

```
hecho_atencion_poblacion.id_atencion_poblacion
FROM
hecho_atencion_poblacion
INNER JOIN dim_tiempo on (hecho_atencion_poblacion.id_tiempo = dim_tiempo.id_tiempo)
INNER JOIN dim_persona on (hecho_atencion_poblacion.id_persona = dim_persona.id_persona)
INNER JOIN dim_materia_ap on (hecho_atencion_poblacion.id_materia_ap =
dim_materia_ap.id_materia_ap)
INNER JOIN dim_instancia on (hecho_atencion_poblacion.id_instancia = dim_instancia.id_instancia)
```

Civil y Familia

Select

```
hecho_expediente_prejudicial.id_expediente_prejudicial
FROM
hecho_expediente_prejudicial
INNER JOIN dim_tiempo on (hecho_expediente_prejudicial.id_tiempo = dim_tiempo.id_tiempo)
INNER JOIN dim_interesado on (hecho_expediente_prejudicial.id_interesado =
dim_interesado.id_interesado)
INNER JOIN dim_involucrado on (hecho_expediente_prejudicial.id_involucrado =
dim_involucrado.id_involucrado)
INNER JOIN dim_instancia on (hecho_expediente_prejudicial.id_instancia = dim_instancia.id_instancia)
INNER JOIN dim_motivo_patriap on (hecho_expediente_prejudicial.id_motivo_patriap =
dim_motivo_patriap.id_motivo_patriap)
```

En el informe agregado del JMeter se muestra una serie de datos los cuales son:

Muestras: cantidad de páginas (hilos) que simula la cantidad de usuarios que están interactuando con el sistema desde la misma URL.

Media: media del tiempo total que demora la petición en cargarse.

Mediana: tiempo promedio que ha tardado en cargarse las páginas.

Min: tiempo mínimo que ha demorado en cargarse una página.

Max: tiempo máximo que ha demorado en cargarse una página.

Desv.Estándar: para cada valor de la serie obtenemos el cuadrado de su distancia respecto a la media (calculando (valor – media)). Para esta nueva serie de valores calculamos su media y le hacemos la raíz cuadrada. Conceptualmente es una medida como de agrupado respecto a la media. Nos dice como de representativa de la serie es el valor de la media. La desviación estándar tiene la particularidad de que se expresa en la misma unidad de medida que los valores de la serie.

Rendimiento: cantidad de solicitudes por segundo.

%Error: por ciento de error de las páginas que no se llegaron a cargar de manera satisfactoria.

Kb/Seg: velocidad de carga de las páginas (22)

A continuación se muestran los resultados que arrojaron las pruebas para el mercado perteneciente al departamento Administrativo, los resultados de los otros mercados se pueden ver en los Anexos.

Administrativo

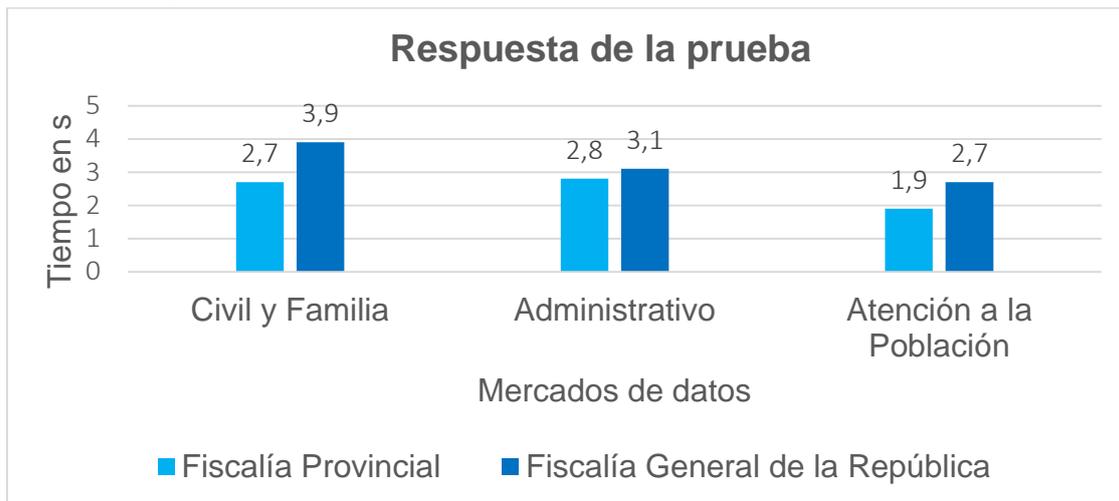
Etiqueta	# Muestras	Media	Mín	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Media de Bytes
Petición JDBC	250	235	11	1446	295,54	0,00%	88,3/sec	14,40	167,0
Total	250	235	11	1446	295,54	0,00%	88,3/sec	14,40	167,0

Figura # 19 Resultados del plan de pruebas simulando cantidad de usuarios de la Fiscalía Provincial

Etiqueta	# Muestras	Media	Mín	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Media de Bytes
Petición JDBC	500	393	11	1788	378,34	0,00%	100,1/sec	16,32	167,0
Total	500	393	11	1788	378,34	0,00%	100,1/sec	16,32	167,0

Figura # 20 Resultados del plan de pruebas simulando cantidad de usuarios de la Fiscalía General

Una vez concluida la prueba se pudo observar que el tiempo de respuesta es aceptable ya que fue menor a los 5 segundos para la cantidad de usuarios conectados concurrentemente.



3.3 PRUEBAS DE VOLUMEN Y CARGA.

La prueba de volumen y carga consiste en insertar una gran cantidad de datos a los mercados de datos para ver si alcanzan límites que los hagan fallar, identificando de esta manera la carga máxima que los mercados de datos pueden manejar por un periodo de tiempo dado. La prueba permite verificar que la aplicación funciona adecuadamente con el máximo de datos esperados en la base de datos. Para la prueba se utilizó la herramienta Datanamic Data Generator para PostgreSQL en su versión 5.0, la cual permitió llenar los mercados de datos con una cantidad determinada de datos. A continuación la figura # 18 muestra el resultado al llenar el mercado de datos de Administrativo:

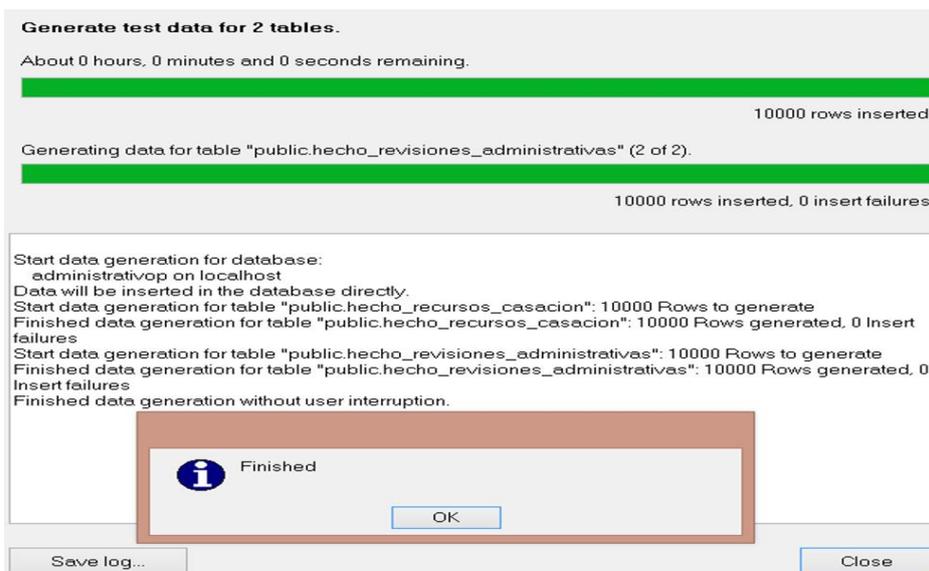


Figura # 21 Prueba de volumen y carga para Administrativo.

Al introducir los datos no se presentaron problemas ni errores.

Con la base de datos llena se realizó la generación de carga con la herramienta JMeter e su versión 2.9, quedando definido la cantidad de usuarios y los tipos de respuesta de la siguiente forma:

Instancia	Cantidad de usuarios
Fiscalía general de la república	25
Fiscalías provinciales	12

Tabla # 16 Cantidad de usuarios concurrentes en las diferentes instancias de las fiscalías.

Tipo de respuesta	Tiempo
Aceptable	Menor o igual a los 5 segundos
No aceptable	Mayor a los 5 segundos

Tabla # 17 Tiempo de respuestas de las consultas.

A continuación se muestran los resultados que arrojaron las pruebas para el mercado perteneciente al departamento Administrativo, los resultados de los otros mercados se pueden ver en los Anexos.

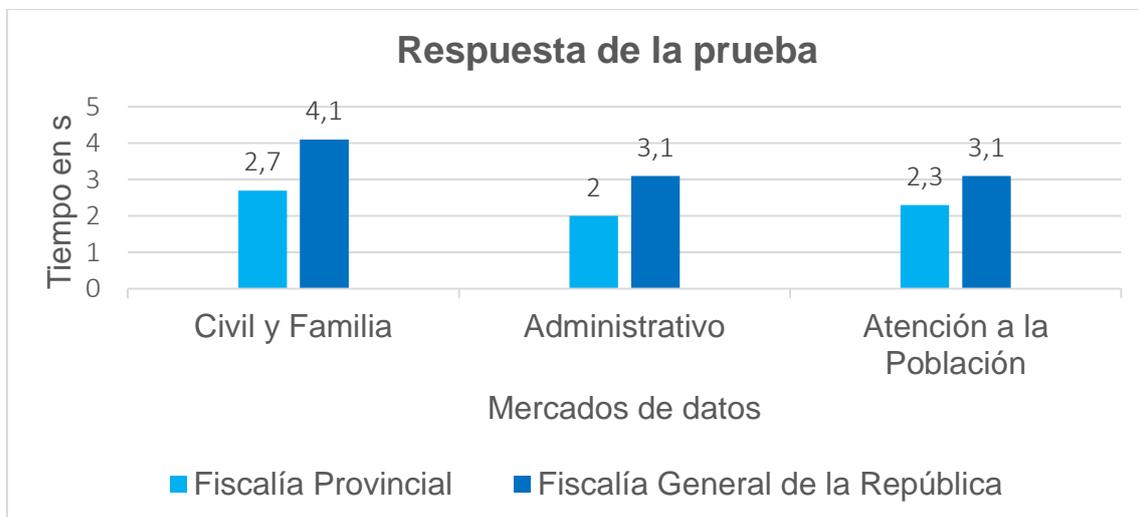
Etiqueta	# Muestras	Media	Mín	Máx	Desv. Están...	% Error	Rendimiento	Kb/sec	Media de Byt...
Petición JDBC	12	1704	442	1961	387,20	0,00%	5,9/sec	227,42	39181,0
Total	12	1704	442	1961	387,20	0,00%	5,9/sec	227,42	39181,0

Figura # 22 Resultados del plan de pruebas simulando la cantidad de usuarios de la Fiscalía Provincial

Etiqueta	# Muestras	Media	Mín	Máx	Desv. Están...	% Error	Rendimiento	Kb/sec	Media de Byt...
Petición JDBC	25	2305	1761	2898	378,56	0,00%	8,3/sec	316,43	39181,0
Total	25	2305	1761	2898	378,56	0,00%	8,3/sec	316,43	39181,0

Figura # 23 Resultados del plan de pruebas simulando la cantidad de usuarios de la Fiscalía General

Una vez concluida la prueba se pudo observar que el tiempo de respuesta es aceptable ya que fue menor a los 5 segundos para la cantidad de usuarios conectados concurrentemente.



Lo anteriormente planteado garantiza que el SGBD utilizado y el diseño implementado soportan completamente el almacenamiento de los niveles de información requeridos para los mercados de datos.

3.4 CONCLUSIONES DEL CAPÍTULO.

En este capítulo se realizaron un conjunto de pruebas que validan la solución propuesta, demostrando que el mercado de datos está apto para recibir el volumen de datos que se espera surjan de los procesos tenidos en cuenta en la investigación y que es capaz de mantener el máximo de conexiones estimadas de manera concurrentemente. Además se midió el tiempo de ejecución de las consultas de acceso a los datos y las mismas cumplen con los umbrales de aceptación definidos por los clientes.

CONCLUSIONES

En el presente trabajo se establece una solución que facilita el análisis de la información y tiene como objetivo el apoyo al proceso de toma de decisiones. La obtención de una propuesta de solución, donde se definieron las herramientas y tecnologías necesarias permitió el desarrollo de los mercados de datos para los departamentos Civil y Familia, Administrativo y Atención a la Población.

Se realizó el análisis y diseño para dar cumplimiento a todo lo requerido en el negocio y para satisfacer las necesidades del cliente. Se logró la limpieza, integración y consistencia de los datos mediante el desarrollo de los procesos de ETL. La representación de la información se garantizó mediante el diseño de las dimensiones y los cubos que permitieron obtener mercados de datos poblados con información para ser consultada por parte de los usuarios.

Se realizaron además pruebas a la solución garantizando su correcto funcionamiento. Es importante destacar que todos los objetivos del trabajo se cumplieron en el tiempo estimado, teniendo en cuenta la planificación realizada para cada una de las tareas.

RECOMENDACIONES

Con el objetivo de mejorar la solución antes planteada se proponen las siguientes recomendaciones:

- Continuar el desarrollo de la solución implementando mercados de datos para los restantes módulos del Sistema de Gestión Fiscal.
- Integrar todos los mercados de datos realizados en un almacén de datos único para todo el sistema.
- Utilizar la metodología y las herramientas seleccionadas en la investigación para el desarrollo de futuros mercados de datos. (3)

BIBLIOGRAFÍA

1. Inmon, W. H. *Building the Data Warehouse*. NEW YORK, EEUU : John Wiley , 2002. ISBN: 0-471-08130-2.
2. Kimball, Ralph. *The Data Warehouse Lifecycle Toolkit*. New York : s.n., 1998.
3. Ricardo Dario Bernabeu. HEFESTO DATA WAREHOUSING: Investigación y Sistematización. Córdoba : s.n., 19 de julio de 2010. Vol. 2.
4. Rivadera, Gustavo R. La metodología de Kimball para el diseño de almacenes de datos . *La metodología de Kimball para el diseño de almacenes de datos* . Buenos Aires : s.n., 2010.
5. Classora Technologies. *Classora Technologies*. [En línea] 25 de junio de 2013. [Citado el: 6 de 11 de 2013.] <http://blog.classora.com/>. B27411289.
6. Ing. Bernabeu R. Dario. DataPrix. *DataPrix*. [En línea] 6 de mayo de 2009. [Citado el: 18 de febrero de 2014.] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse/34-datawarehouse-manager>.
7. Flores, Jesús Alberto Cogque. Base de Datos Para Aplicaciones. *Base de Datos Para Aplicaciones*. [En línea] 16 de abril de 2013. [Citado el: 18 de febrero de 2014.] <https://sites.google.com/site/jesusalbertocogqueflores/unidad-v/ejemplo-datamart>.
8. Datamart S.A. *Datamart S.A.* [En línea] [Citado el: 18 de febrero de 2014.] <http://www.datamart.cl/>.
9. Buenas Tareas. [En línea] marzo de 2010. [Citado el: 18 de febrero de 2014.] <http://www.buenastareas.com/ensayos/Ejemplo-Datamart/173004.html> . 173004.
10. Buenas Tareas.com. *Buenas Tareas.com*. [En línea] 2013. [Citado el: 18 de febrero de 2014.] <http://www.buenastareas.com/ensayos/Almac%C3%A9n-De-Datos/26623982.html>. 26623982.
11. Infomed. *Informática Salud 2013*. [En línea] 2013. <http://www.informatica2013.sld.cu/index.php/informaticasalud/2013/paper/view/195>. 195.
12. DATAPRIX. *DATAPRIX*. [En línea] 13 de julio de 2010. [Citado el: 19 de mayo de 2014.] <http://www.dataprix.com/empresa/productos/pentaho-bi-suite>.
13. Integrando Datos. *Integrando datos*. [En línea] Pentaho. [Citado el: 19 de febrero de 2014.] <http://integrandodatos.blogspot.com/2013/06/pentaho-data-integration-instalacion.html>.

14. pentaho. *pentaho*. [En línea]
<http://community.pentaho.com/projects/mondrian/>.
15. postgresql-es. *postgresql-es*. [En línea] 2009.
http://www.postgresql.org.es/sobre_postgresql.
16. guia-ubuntu. *guia-ubuntu*. [En línea] 10 de marzo de 2008. [Citado el: 19 de febrero de 2014.] http://www.guia-ubuntu.com/index.php?title=PgAdmin_III.
17. Ecured. *Ecured*. [En línea] [Citado el: 19 de febrero de 2014.]
http://www.ecured.cu/index.php/Visual_Paradigm.
18. Apache Software Foundation. [En línea] 1999-2014. <http://tomcat.apache.org/>.
19. DATANAMIC. *DataNamic Tools for Database Developers*. [En línea] febrero de 2014. [Citado el: 19 de mayo de 2014.] <http://www.datanamic.com/>.
20. Softpedia. *Softpedia*. [En línea] 6 de enero de 2014. [Citado el: 18 de febrero de 2014.] <http://www.softpedia.es/programa-JMeter-161347.html>.
21. Bernabeu, Ricardo Dario. Dataprix. *Dataprix*. [En línea] 27 de enero de 2010. [Citado el: 23 de abril de 2014.] <http://www.dataprix.com/blogs/bernabeu-dario/claves-subrogadas>.
22. conocimiento, Consultoría de áreas de. JMeter. *Manual de usuario de Apache Jmeter*. 2009. V 1.2.
23. Dario, Ing. Bernabeu Ricardo. *Metodología propia para la Construcción de un Data Warehouse*. . Cordoba : s.n., 2009.
24. Torres, Liudmila Padrón. mailxmail. *mailxmail*. [En línea] 09 de 1 de 2006. [Citado el: 27 de 11 de 2013.] <http://www.mailxmail.com/curso-almacenes-datos-importancia-estandar/arquitectura-dw>.
25. Wang, John. *Encyclopedia of Data Warehousing*. Montclair State University, USA : Advisory Board, 2006. ISBN 1-59140-559-9.
26. Business Intelligence - Almacenes de Datos - ETL. *Business Intelligence - Almacenes de Datos - ETL*. [En línea] 2006. [Citado el: 6 de 11 de 2013.] http://etl-tools.info/es/bi/almacenedatos_esquema-copo-de-nieve.htm.
27. Ing.Cañizares., Mauricio Guanche. *Desarrollo de un DataMart para la obtención de las razones financieras de los subsistemas de*. La Habana : s.n., 2011.
28. LUIS ORTIZ. Webquest.es. *Webquest.es*. [En línea] 10 de octubre de 2010. [Citado el: 18 de febrero de 2014.] <http://www.webquest.es/wq/data-warehouse-y-datamart>.
29. Méndez, Geidy Acosta. Informática Salud 2013. *Infomed*. [En línea] 2013. <http://www.informatica2013.sld.cu/index.php/informicasalud/2013/paper/view/195>. 195.

30. Revista de Arquitectura e Ingeniería. [En línea] 2013.
<http://www.redalyc.org/articulo.oa?id=193915954004>.

31. Oracle. *Oracle Business Intelligence Enterprise*. [En línea] Oracle, 27 de abril de 2011. [Citado el: 19 de febrero de 2014.]
<http://www.oracle.com/lad/corporate/press/pr-lad-110427-367397-esa.html>.

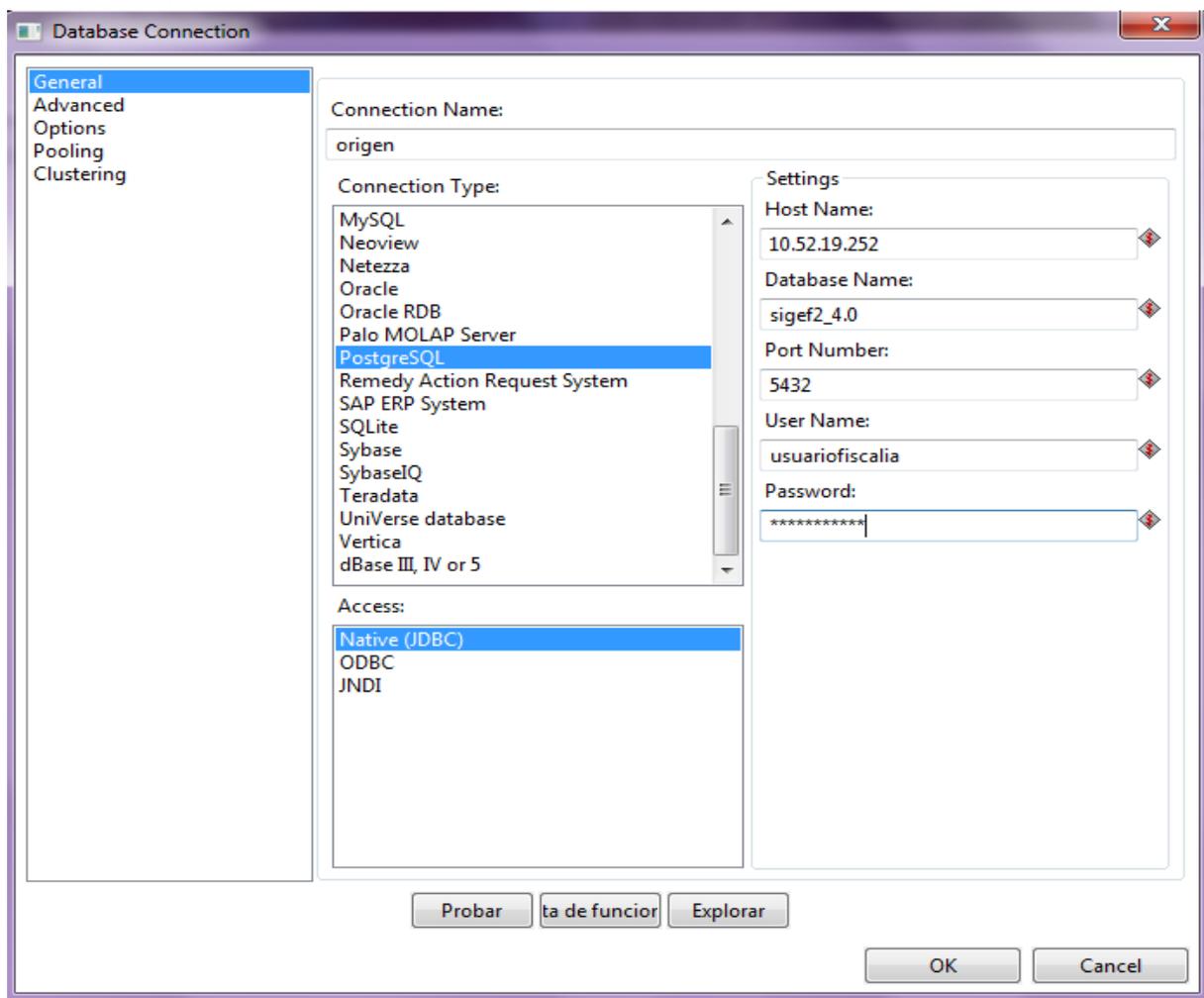
32. Descargas de Software. [En línea] [Citado el: 19 de febrero de 2014.]
http://www.freedownloadmanager.org/es/downloads/Generador_de_Datos_de_SME_2005_para_PostgreSQL_37202_p/.

33. Infante, Ing. Dennis Alba. Pasos para crear cubos con Mondrian Schema Workbench. *Pasos para crear cubos con Mondrian Schema Workbench*. 2009.

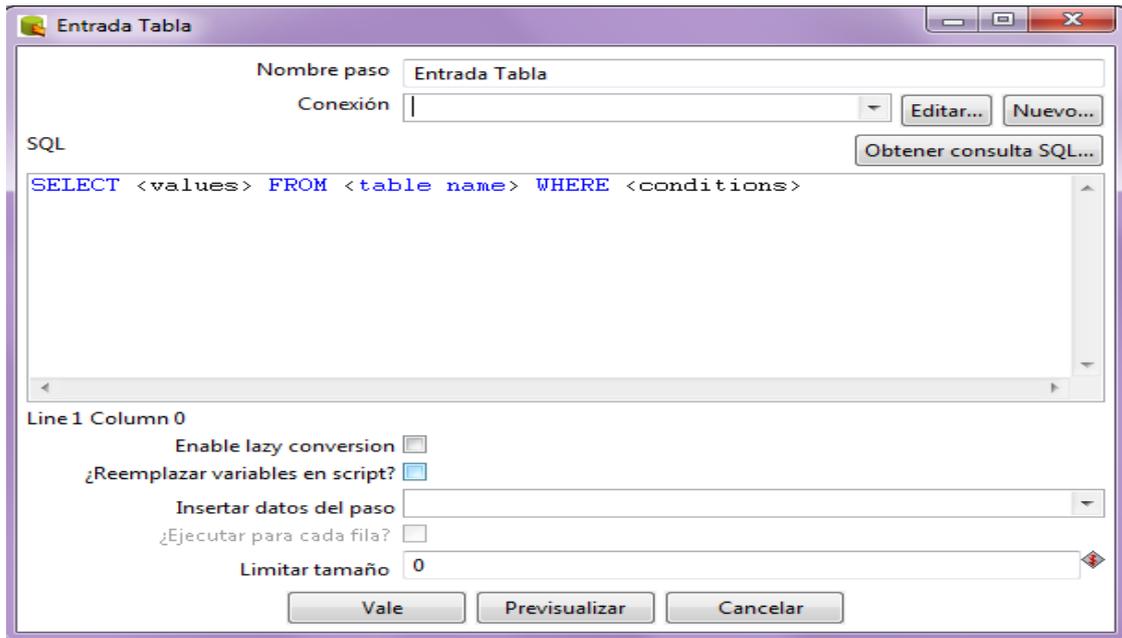
34. Datanamic. Datanamic. *Datanamic*. [En línea] 2014. [Citado el: 02 de 05 de 2014.] <http://www.datanamic.com/>.

ANEXOS

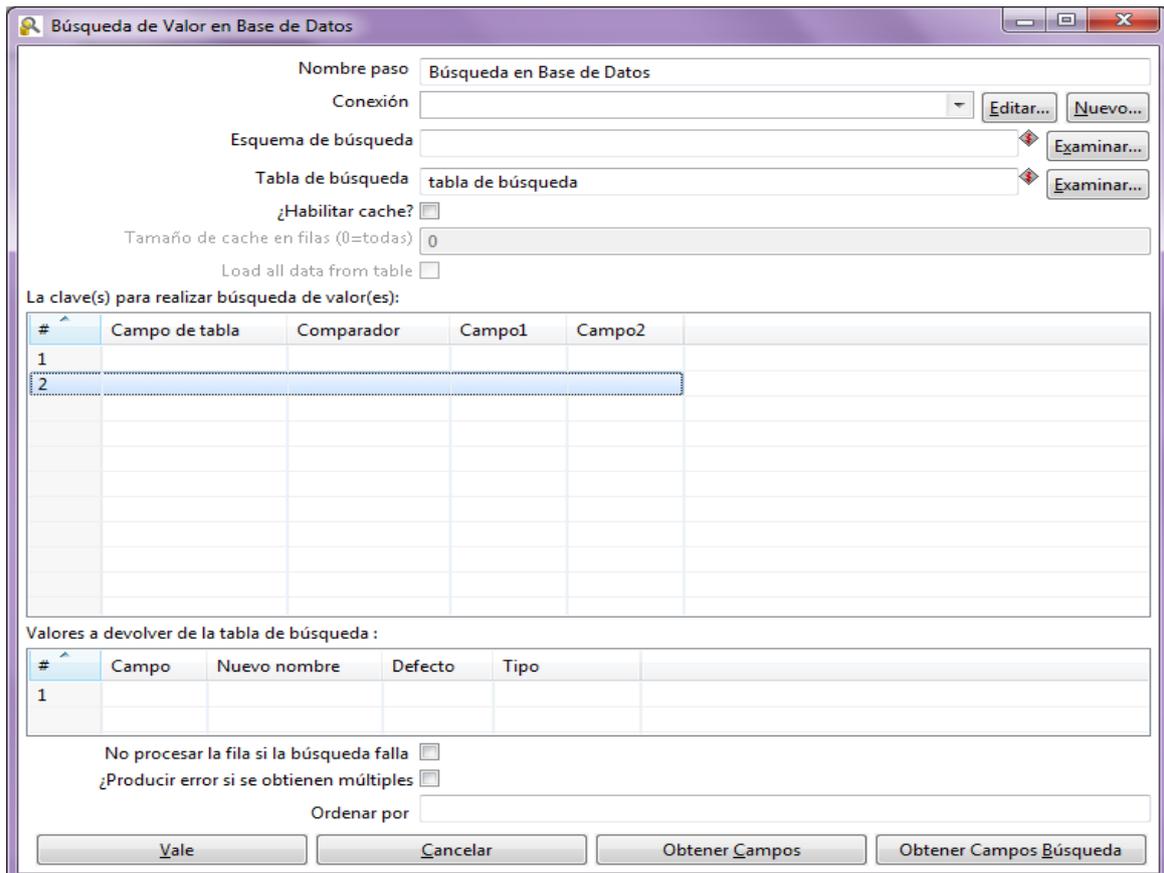
Anexo 1: Conexión con base de datos PostgreSQL en Kettle Pentaho Data Integration.



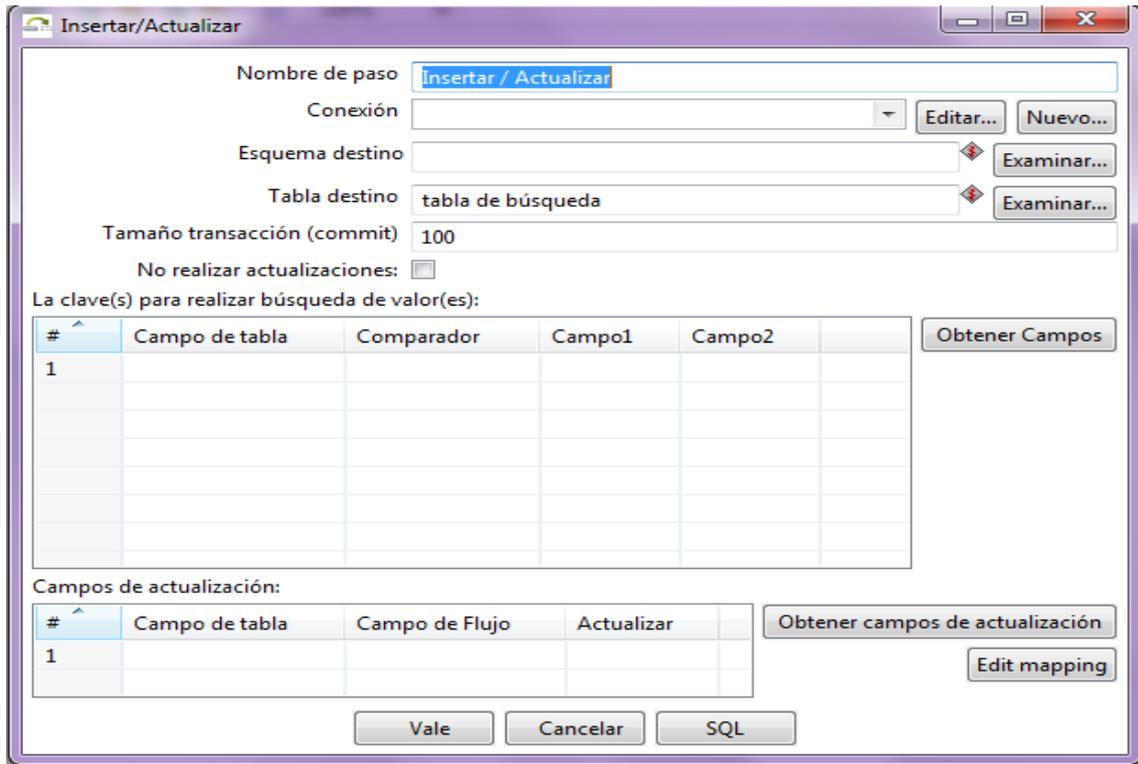
Anexo 2 Componente para la entrada de datos.



Anexo 3 Componente para realizar búsqueda en base de datos.



Anexo 4 Componente para insertar o actualizar los valores



Anexo 5 Transformaciones realizadas para las dimensiones de Administrativo

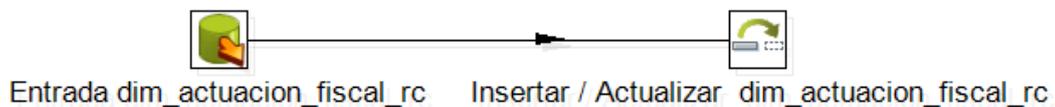


Figura # 24 Transformación realizada para la dimensión actuación del fiscal para recurso casación.



Figura # 25 Transformación realizada para la dimensión actuación del fiscal para proceso sucesorio.



Figura # 26 Transformación realizada para la dimensión motivo.



Figura # 27 Transformación realizada para la dimensión resolución



Figura # 28 Transformación realizada para la dimensión tipo de bien.

Anexo 6 Transformaciones realizadas para las dimensiones de Atención a la Población

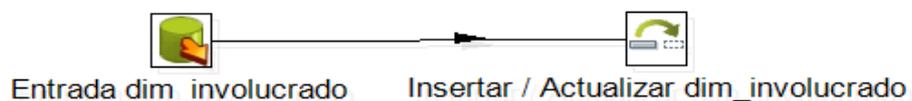


Figura # 29 Transformación realizada para la dimensión entidad involucrada.

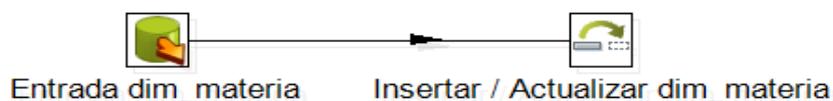


Figura # 30 Transformación realizada para la dimensión materia.



Figura # 31 Transformación realizada para la dimensión persona.

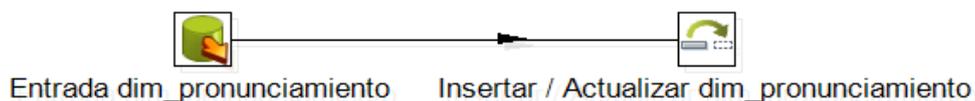


Figura # 32 Transformación realizada para la dimensión pronunciamento.



Entrada dim_respuesta_impugnacion Insertar / Actualizar dim_respuesta_impugnacion

Figura # 33 Transformación realizada para la dimensión respuesta impugnación.

Anexo 7 Transformaciones realizadas para las dimensiones de Civil y Familia



Entrada dim_asunto_psumario Insertar / Actualizar dim_asunto_psumario

Figura # 34 Transformación realizada para la dimensión asunto de proceso contencioso sumario.



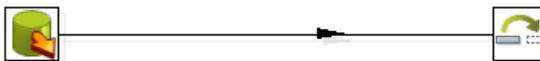
Entrada dim_asunto_jurisdiccion Insertar / Actualizar dim_asunto_jurisdiccion

Figura # 35 Transformación realizada para la dimensión asunto de jurisdicción voluntaria.



Entrada dim_asunto_notarial Insertar / Actualizar dim_asunto_notarial

Figura # 36 Transformación realizada para la dimensión asunto notarial



Entrada dim_asunto_pordinario Insertar / Actualizar dim_asunto_pordinario

Figura # 37 Transformación realizada para la dimensión proceso contencioso ordinario.



Entrada dim_fiscal_recurrente Insertar / Actualizar dim_fiscal_recurrente

Figura # 38 Transformación realizada para la dimensión fiscal recurrente.

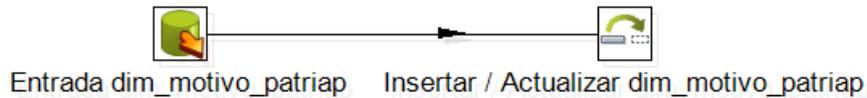


Figura # 39 Transformación realizada para la dimensión motivo de patria potestad.

Anexo 8 Transformaciones realizadas para los hechos de Administrativo



Figura # 40 Transformación realizada para el hecho procedimientos contenciosos.



Figura # 41 Transformación realizada para el hecho procesos contenciosos.



Figura # 42 Transformación realizada para el hecho recursos de casación.

Anexo 9 Transformaciones realizadas para los hechos de Atención a la Población

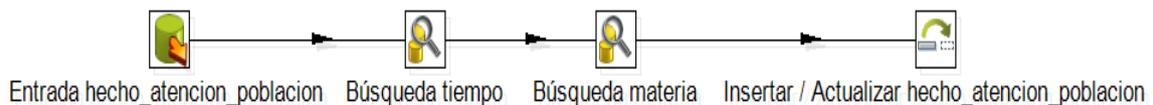


Figura # 43 Transformación realizada para el hecho atención población.



Figura # 44 Transformación realizada para el hecho impugnaciones.



Figura # 45 Transformación realizada para el hecho reclamaciones.

Anexo 10 Transformaciones realizadas para los hechos de Civil y Familia



Figura # 46 Transformación realizada para el hecho asuntos notariales



Figura # 47 Transformación realizada para el hecho expropiación forzosa.



Figura # 48 Transformación realizada para el hecho expediente prejudicial.



Figura # 49 Transformación realizada para el hecho jurisdicción voluntaria.



Figura # 50 Transformación realizada para el hecho procesos contenciosos ordinarios.



Figura # 51 Transformación realizada para el hecho procesos contenciosos sumarios.



Figura # 52 Transformación realizada para el hecho procesos sucesorios.



Figura # 53 Transformación realizada para el hecho recursos de apelación.

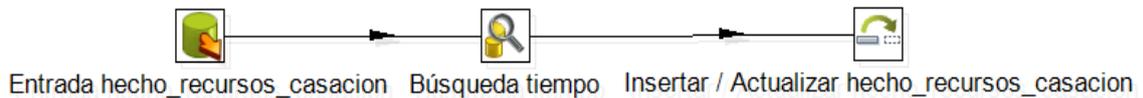


Figura # 54 Transformación realizada para el hecho recursos de casación.



Figura # 55 Transformación realizada para el hecho revisiones civiles.

Anexo 11 Trabajos realizados para Administrativo.

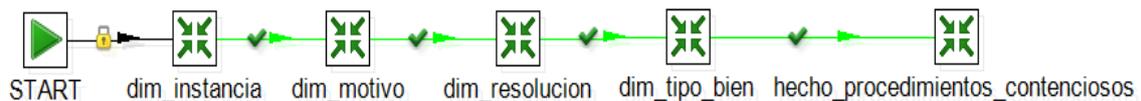


Figura # 56 Trabajo realizado para el hecho procedimiento contencioso.



Figura # 57 Trabajo realizado para el hecho procesos contenciosos.

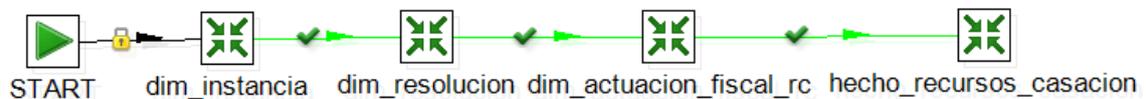


Figura # 58 Trabajo realizado para el hecho recursos de casación.

Anexo 12 Trabajos realizados para Atención a la Población.



Figura # 59 Trabajo realizado para el hecho atención población.



Figura # 60 Trabajo realizado para el hecho impugnaciones.

Anexo 13 Trabajos realizados para Civil y Familia.

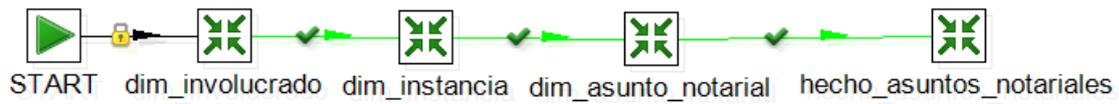


Figura # 61 Trabajo realizado para el hecho asuntos notariales.



Figura # 62 Trabajo realizado para el hecho expropiación forzosa.

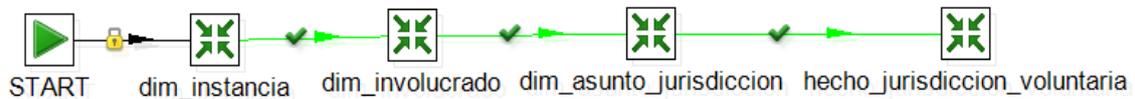


Figura # 63 Trabajo realizado para el hecho jurisdicción voluntaria.



Figura # 64 Trabajo realizado para el hecho proceso contencioso ordinario.



Figura # 65 Trabajo realizado para el hecho proceso contencioso sumario.



Figura # 66 Trabajo realizado para el hecho procesos sucesorios.



Figura # 67 Trabajo realizado para el hecho recursos de apelación.

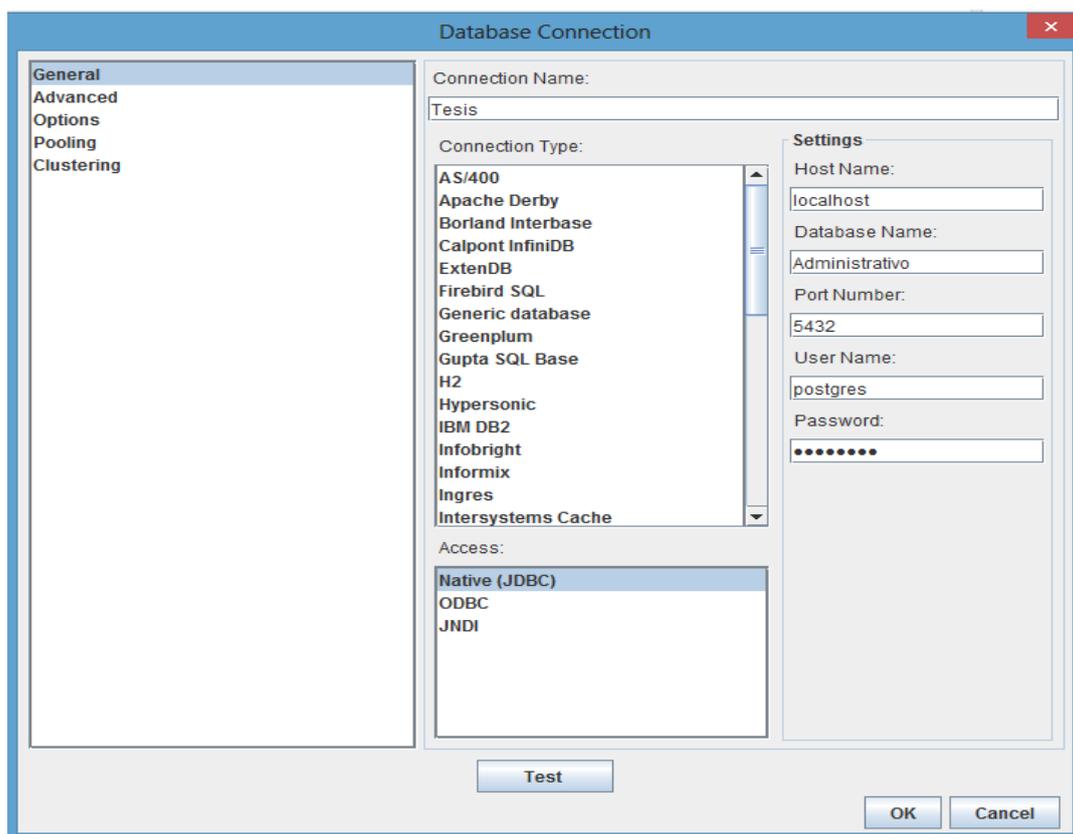


Figura # 68 Trabajo realizado para el hecho recursos de casación.



Figura # 69 Trabajo realizado para el hecho revisiones civiles.

Anexo 14 Conexión al mercado de datos.



Anexo 15 Esquema multidimensional para el mercado Atención a la Población

GLOSARIO DE TÉRMINOS.

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

SQL: es un lenguaje formal declarativo para manipular información en una base de datos

COUNT: función definida en el lenguaje SQL para contar, también se puede encontrar definida en otros lenguajes.

CASE: función definida en el lenguaje SQL para condiciones, también se puede encontrar definida en otros lenguajes.

MDX: es el acrónimo de MultiDimensional eXpressions, es un lenguaje de consulta para bases de datos multidimensionales sobre cubos OLAP

UML: lenguaje unificado de modelado (UML, por sus siglas en inglés, Unified Modeling Language) es el lenguaje de modelado de sistemas de software más conocido y utilizado en la actualidad.

Nivel de granularidad: la granularidad representa el nivel de detalle al que se desea almacenar la información sobre el negocio que se esté analizando.

Requerimiento: es una necesidad documentada sobre el contenido, forma o funcionalidad de un producto o servicio.

XML: son las siglas en inglés de Extensible MarkupLanguage, es un metalenguaje extensible de etiquetas desarrollado por World Wide Web Consortium (W3C).

Servlet: programas de aplicación que es ejecutado por un servidor.

Java: es un lenguaje de programación orientado a objetos desarrollado por Sun Microsystems.

C++: es un lenguaje de programación orientado a objetos.

Automatización: sistema de producción en el que se usan máquinas en lugar de mano de obra; también puede verse como un proceso para lograr operaciones automáticas.

Llave primaria: es un término de bases de datos que constituye una identificación unívoca, esta puede ser un atributo o una combinación de estos.

Java Server Page (JSP): es una tecnología Java que permite generar contenido dinámico para la web.

Código abierto: es el término con el que se conoce al software distribuido y desarrollado libremente, tiene un punto de vista más orientado a los beneficios prácticos de compartir el código que a las cuestiones morales y/o filosóficas.