

Universidad de las Ciencias Informáticas

Facultad 6



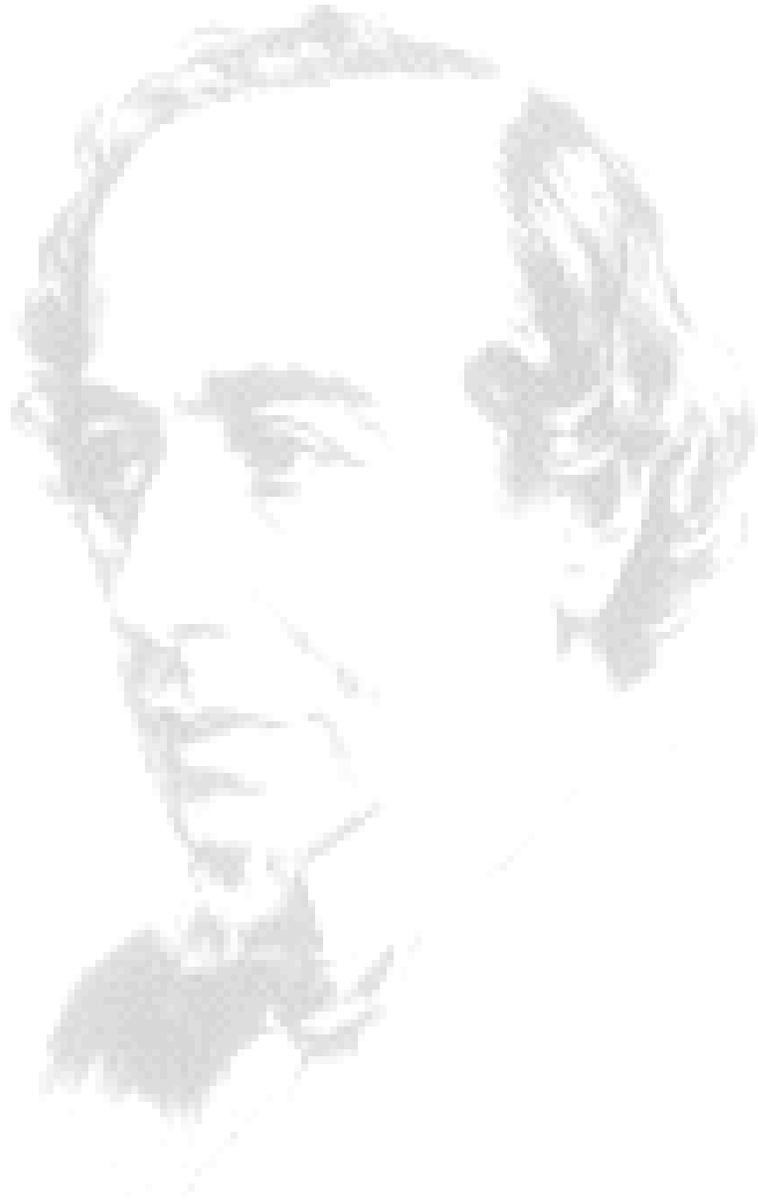
Título: Mercado de datos Económico del Sistema de Información Estadístico Judicial para los Tribunales Populares.

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

***Autores: Odalys Sardiñas Hernández
Andry Benavides Peña***

***Tutores: Ing. Rosayda Valiente Mesa
Ing. Georvys González Rojas***

*La Habana, Cuba
Junio, 2013
“Año 55 de la Revolución”*



"El secreto del éxito es la constancia en el propósito."

Benjamin Disraeli

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 _____.

Odalys Sardiñas Hernández

Firma de Autora

Andry Benavides Peña

Firma de Autor

Rosayda Valiente Mesa

Firma de Tutora

Georvys González Rojas

Firma de Tutor

DATOS DE CONTACTO

Tutora:

Tutora: Ing. Rosayda Valiente Mesa

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Categoría docente: Instructor

Años de experiencia en el tema: 3

Años de graduado: 5

Correo Electrónico: rvaliente@uci.cu

Tutor:

Tutor: Ing. Georvys González Rojas

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Años de experiencia en el tema: 2

Años de graduado: 2

Correo Electrónico: grojas@uci.cu

El afecto, el cariño y la amistad que se recibe de los que te rodean son factores que determinan de forma decisiva el desarrollo y equilibrio de una persona. La cualidad más hermosa que debe revestir un ser humano es reconocer a quienes lo han llevado a conseguir sus objetivos. No es tarea fácil poder escribir en un papel a todos los que me ayudaron en la realización de este trabajo o que compartieron momentos inolvidables de mi vida, pero trataré de agradecerles a todos aquellos que me aportaron su granito de arena o que con su amor y cariño me dieron fuerzas para poder realizar este trabajo.

Agradezco:

A todas las personas que una vez me preguntaron “y la tesis, cómo va?”.

A mis amigos que están dispuestos a extender su mano amiga siempre que sea preciso, que me han apoyado en cada minuto de la vida y que han estado conmigo compartiendo alegrías, tristezas, aventuras y locuras.

Al profesor Monchí con quien me siento eternamente agradecida por haberme brindarme desinteresadamente un poco de su tiempo, ayuda y apoyo.

A mis entrenadores Roly y Salvador quienes me enseñaron y me demostraron que con esfuerzo, sacrificio y perseverancia todo lo que uno se proponga lo puede lograr, que siempre hay que preocuparse menos y ocuparse más.

A mi novio, por ser especial en mi vida y hacerme sentir segura en los momentos más cruciales e importantes de mi vida, por darme todo su amor, comprensión y apoyarme incondicionalmente, por ayudarme a lograr esta meta y confiar en mí.

A mis familiares por haber estado pendientes de mi trabajo a cada instante, por su dedicación, confianza y amor.

A mis padres que son los principales intérpretes de mi vida, esforzándose por convertirme en un ser cada día mejor, por guiarme siempre por el camino correcto, por alentarme, darme seguridad ante todos los acontecimientos de mi vida y estar a mi lado en cada momento. Por enseñarme a luchar incansablemente por un sueño.

Odalys

Agradezco:

A toda mi familia, que a pesar de la distancia siempre estuvieron en mis pensamientos y me apoyaron en todo momento, principalmente mi mamá Xiomara, que nunca me ha dado la espalda por muy difícil que han sido los tiempos, a ella, que durante mucho tiempo ha sido madre y padre a la vez, por eso, todo mi agradecimiento, amor y cariño. A mi hermanita Yanet, que a pesar de ser menor que yo siempre me ha brindado su hombro para que me apoye, me ha dado fuerzas para continuar, y sobre todo, me ha dado mucho amor. Agradezco a mis abuelitos Elsa y Armando, que más que abuelos son mis padres, por haber estado a mi lado cuando más lo he necesitado, por tenerme entre sus brazos y darme cariño, amor y confianza.

A mi novia Ailyn, por haber estado a mi lado durante casi toda la carrera ayudándome, apoyándome y sobre todo, brindándome todo su amor desinteresadamente, por darme fuerzas para continuar en los momentos difíciles, por decirme cada día las palabras mágicas "Te amo".

A todas las personas que de una forma u otra contribuyeron a que este sueño fuera hecho realidad, principalmente los amigos de la infancia, los que he conocido con el pasar del tiempo, mis amigos de la universidad, mi primo "el Yoyo" que sin ellos no habría sido posible lograr que este sueño se hiciera realidad.

A los profesores del departamento Almacenes de Datos, en especial al ingeniero Monchy, que desinteresadamente me brindó parte de su tiempo y conocimiento. Mis tutores, que me apoyaron en todo momento, me indicaron y me guiaron por el camino correcto.

Dedico el presente trabajo de diploma:

A mi mamá Xiomara, que no existen palabras suficientes para decir todo lo que la quiero.

A mi hermanita Yanet, por la personita más sencilla que he podido conocer.

A mis abuelitos Elsa y Armando, que son lo más preciado que tengo en la vida.

A mi novia Ailyn, que ha sabido valorar mi esfuerzo y ha compartido conmigo los momentos malos y buenos.

A mi gran amigo Maikel, que a pesar de no estar presente físicamente, siempre lo tendré en mi corazón, porque los grandes amigos nunca se olvidan.

A mi papá, que por caprichos de la vida hoy no lo puedo abrazar, pero sé que en todo momento ha estado a mi lado y me ha ayudado en los momentos de soledad y tristeza.

Andry

RESUMEN

La presente investigación surge como parte de la colaboración que existe entre el departamento estadístico del Tribunal Supremo Popular y la Universidad de las Ciencias Informáticas (UCI). La principal tarea de este departamento es analizar los principales indicadores estadísticos del tribunal, con el objetivo de contribuir a la toma de decisiones. En esta entidad la información almacenada se ha ido incrementando con el pasar del tiempo. La mayoría de estos datos no son entendibles debido a que provienen de diferentes fuentes, son inconsistentes y no están integrados, trayendo consigo la necesidad de convertir esta información en conocimiento útil. Debido a esto, surge como objetivo fundamental de la investigación: desarrollar el mercado de datos de la materia económica del Sistema de Información Estadístico Judicial para los Tribunales Populares. Para dar respuesta al objetivo trazado se fundamentaron las diferentes herramientas informáticas y metodologías utilizadas durante la investigación. De igual forma, se realizó el análisis, diseño, implementación y pruebas de los subsistemas de integración, almacenamiento y visualización. Como resultado final, se obtuvo el Mercado de Datos poblado y una capa de presentación que contiene todos los reportes que satisfacen las solicitudes de información hechas por los especialistas de la materia económica del Tribunal Supremo Popular.

PALABRAS CLAVE: Almacén de Datos, Herramientas, Mercado de Datos, Metodologías, Tribunal Supremo Popular, Sistema de Información Estadístico Judicial.

Índice

INTRODUCCIÓN.....	1
CAPÍTULO 1: FUNDAMENTOS TEÓRICOS PARA EL DESARROLLO DE UN MERCADO DE DATOS	5
1.1 Sistemas de Información.....	5
1.2 Almacén de datos	6
1.3 Mercado de datos	6
1.3.1 Topología de esquema	7
1.3.2 Modos de almacenamiento	9
1.4 Metodología para el desarrollo de los mercados de datos	10
1.5 Herramientas empleadas para el desarrollo de los mercados de datos.....	12
1.5.1 Herramientas de Modelado	12
1.5.2 Sistema Gestor de Base de Datos	13
1.5.3 Herramientas para la limpieza, estandarización e integración de los datos	14
1.5.4 Herramientas para la Inteligencia de Negocios (BI).....	15
CAPÍTULO 2: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS ECONÓMICO DEL SISTEMA DE INFORMACIÓN ESTADÍSTICO JUDICIAL PARA LOS TRIBUNALES POPULARES	17
2.1 Análisis del Negocio.....	17
2.2 Técnicas de captura de requisitos.....	18
2.3 Reglas del negocio	18
2.4 Especificación de requisitos	19
2.5 Diagrama de Casos de Uso del Sistema	25
2.5.1 Actores del Sistema	26
2.5.2 Especificación de Casos de Uso	26
2.6 Arquitectura del mercado de datos Económico	28
2.7 Diseño del mercado de datos Económico	29
2.7.1 Diseño del Subsistema de almacenamiento.....	29
2.7.2 Diseño del subsistema de integración	35
2.7.3 Diseño del subsistema de visualización	38
2.8 Esquema de seguridad	39

2.8.1 Políticas de respaldo y recuperación	40
CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBAS DEL MERCADO DE DATOS ECONÓMICO DEL SISTEMA DE INFORMACIÓN ESTADÍSTICO JUDICIAL PARA LOS TRIBUNALES POPULARES....	42
3.1 Implementación del subsistema de almacenamiento	42
3.2 Implementación del subsistema de integración de los datos	46
3.2.1 Implementación de las dimensiones lentamente cambiante	47
3.2.2 Implementación de los flujos de transformación	48
3.2.3 Implementación de los trabajos	49
3.2.4 Metadatos	50
3.3 Implementación del subsistema de visualización	50
3.3.1 Implementación de los cubos OLAP.....	51
3.4 Implementación de los reportes candidatos	51
3.4.1 Consultas MDX	52
3.5 Pruebas de software al mercado de datos económico	53
3.5.1 Herramientas de pruebas	54
3.5.2 Resultados de las pruebas aplicadas	55
CONCLUSIONES GENERALES.....	59
RECOMENDACIONES	60
REFERENCIAS BIBLIOGRÁFICAS.....	61
BIBLIOGRAFÍA.....	64
GLOSARIO DE TÉRMINOS	66

Índice de figuras

<i>Figura 1 Esquema en estrellas.....</i>	<i>7</i>
<i>Figura 2 Esquema copo de nieves.....</i>	<i>8</i>
<i>Figura 3 Esquema constelaciones</i>	<i>8</i>
<i>Figura 4 Diagrama de Casos de Uso del Sistema.....</i>	<i>25</i>
<i>Figura 5 Arquitectura del mercado de datos Económico</i>	<i>29</i>
<i>Figura 6 Dimensión monto m3</i>	<i>30</i>
<i>Figura 7 Hecho monto reclamado.....</i>	<i>32</i>
<i>Figura 8 Modelo dimensional</i>	<i>35</i>
<i>Figura 9 Calidad de la fuente de datos.....</i>	<i>36</i>
<i>Figura 10 Distribución de los tipos de datos en la fuente</i>	<i>36</i>
<i>Figura 11 Cantidad máxima de caracteres.....</i>	<i>37</i>
<i>Figura 12 Diseño de las transformaciones.....</i>	<i>38</i>
<i>Figura 13 Mapa de navegación.....</i>	<i>39</i>
<i>Figura 14 Usuarios y roles</i>	<i>40</i>
<i>Figura 15 Transformación del hecho monto reclamado m3.....</i>	<i>49</i>
<i>Figura 16 Trabajo principal de la carga incremental.....</i>	<i>50</i>
<i>Figura 17 Cubo perteneciente al hecho monto reclamado m3</i>	<i>51</i>
<i>Figura 18 Reporte monto reclamado.....</i>	<i>52</i>
<i>Figura 19 Consultas MDX.....</i>	<i>53</i>
<i>Figura 20 Mostrar montos reclamados.....</i>	<i>55</i>
<i>Figura 21 Resultado de las pruebas unitarias</i>	<i>56</i>
<i>Figura 22 Resultado de las pruebas al sistema.....</i>	<i>56</i>

Figura 23 Resultado de las pruebas de aceptación.....57

Figura 24 Resultado general de las pruebas.....58

Índice de tablas

Tabla 1 Actores del sistema y su descripción.....26

Tabla 2 Caso de uso Mostrar monto reclamado.....26

Tabla 3 Matriz dimensional del mercado de datos económico34

Tabla 4 Esquemas y tablas del mercado de datos económico.....42

Tabla 5 Estándares de codificación del mercado de datos económico.....45

INTRODUCCIÓN

Con la ayuda de las Tecnologías de la Información y las Comunicaciones (TIC), la sociedad, las empresas, instituciones y entidades están cada día más conectadas electrónicamente. Se encuentran envueltas en un entorno de constantes cambios y evoluciones que influyen en la inserción de las computadoras en las entidades, elemento fundamental para automatizar sus procesos.

En Cuba, las TIC se han transformado en un mecanismo imprescindible para formar las líneas de desarrollo de la sociedad cubana. En este entorno, las instituciones de la isla se esfuerzan por mejorar y controlar la información histórica que ellas generan, permitiéndoles conocer su estado actual y tomar decisiones a diario. La UCI es uno de los centros con mayor influencia en este perfeccionamiento, la cual cuenta con diferentes centros de investigación y desarrollo, destacándose el Centro de Tecnologías de Gestión de Datos (DATEC), con su departamento Almacenes de Datos, que vinculado con el Tribunal Supremo Popular contribuyen a la mejora de los sistemas de trabajo en este organismo.

El Tribunal Supremo Popular (TSP) es el órgano rector de justicia, responsabilizado en ejercer la máxima autoridad judicial en la República de Cuba. Es el encargado de tomar decisiones judiciales y de dictar normas de obligatorio cumplimiento por todos los tribunales provinciales y municipales del país.

El Departamento Independiente de Estadística Judicial (DIEJ) que pertenece a dicha entidad, tiene la finalidad de presentar cifras estadísticas e indicadores que reflejen el comportamiento y desarrollo de la actividad jurídica del país. Este departamento cuenta con cuatro materias, entre las que se encuentra la Económica, encargada de recoger los datos estadísticos de los conflictos que surjan entre personas naturales o jurídicas, cubanas o extranjeras con representación, bienes o intereses en Cuba.

En la actualidad el DIEJ presenta dificultades con el manejo de su información y con el análisis estadístico, que se realiza a través de mecanismos no automatizados, poco confiables y tediosos. La información es almacenada en herramientas basadas en aplicaciones de oficina como Excel y dbf, siendo costoso en cuanto a tiempo y esfuerzo. Existen también limitaciones para recuperar los datos debido a que la información se encuentra almacenada en distintas computadoras. Además, anualmente se genera un gran número de datos en diferentes versiones que obstaculizan el análisis e integración de los mismos.

Estas condiciones, deterioran la calidad de la información en cuanto a disponibilidad e integridad, lo cual impacta negativamente en el análisis estadístico de diferentes variables relacionadas con los datos de la materia económica que se procesan en el Departamento Independiente de Estadística Judicial del Tribunal Supremo Popular.

La situación anteriormente descrita, permite identificar el siguiente problema de investigación: ¿Cómo contribuir a la toma de decisiones en materia económica del Sistema de Información Estadístico Judicial para los Tribunales Populares?

La investigación tiene como **objeto de estudio** los Almacenes de Datos (AD), enmarcada en el **campo de acción** Mercado de Datos para la materia económica del Sistema de Información Estadístico Judicial para los Tribunales Populares (SIEJT).

El **objetivo general** de este trabajo es: Desarrollar el mercado de datos de la materia económica del Sistema de Información Estadístico Judicial para los Tribunales Populares, que contribuya a la toma de decisiones.

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

1. Fundamentar la selección de la metodología, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.
2. Realizar el análisis y diseño del mercado de datos en la materia económica del Sistema de Información Estadístico Judicial para los Tribunales Populares.
3. Realizar la implementación y las pruebas del mercado de datos de la materia económica del Sistema de Información Estadístico Judicial para los Tribunales Populares.

Para el cumplimiento de estos objetivos, se realizarán las siguientes **tareas de la investigación**:

1. Análisis de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos, lo que contribuirá a determinar cuáles se utilizarán durante la investigación.
2. Levantamiento de requisitos del mercado de datos para determinar las necesidades de información.

3. Descripción de los casos de uso del mercado de datos para especificar cada una de las funcionalidades del sistema.
4. Definición de la arquitectura del mercado de datos, que permitirá identificar los principales subsistemas que la componen.
5. Definición de los hechos, las medidas y las dimensiones del mercado de datos para determinar los elementos que forman parte del modelo lógico de datos.
6. Diseño del modelo lógico de datos para así determinar los elementos que componen el modelo físico de los datos.
7. Diseño del subsistema de integración del mercado de datos como guía para la implementación de dicho subsistema.
8. Diseño del subsistema de visualización del mercado de datos, que permite realizar la capa de visualización y el diseño de los cubos OLAP.
9. Implementación del modelo de datos para que queden disponibles las estructuras de la base de datos a la hora de realizar la carga del Mercado de Datos.
10. Implementación del subsistema de integración del mercado de datos para que quede poblado, cargando los hechos y las dimensiones correspondientes.
11. Implementación del subsistema de visualización del mercado de datos con el objetivo de obtener los reportes para los usuarios finales.
12. Diseño de los casos de pruebas del mercado de datos para identificar los elementos que tienen que estar disponibles una vez culminada la implementación.
13. Aplicación de las listas de chequeo al mercado de datos para determinar que la estructura de los artefactos que corresponden a los procesos de Extracción, Transformación y Carga (ETL), tengan la calidad requerida.
14. Aplicación de los casos de prueba al mercado de datos para validar los reportes realizados.

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

El primer capítulo titulado Fundamentos teóricos para el desarrollo de un Mercado de Datos (MD), se centrará en el análisis de las principales características, metas y arquitectura de los almacenes y

mercados de datos, así como en las características de la metodología, técnicas y herramientas a utilizar para el desarrollo de la solución.

El segundo capítulo, Análisis y diseño del MD de la materia económica del Sistema de Información Judicial para Tribunales, aborda aspectos referentes al levantamiento de requisitos y a la definición de las reglas del negocio. Se describen los casos de usos identificados con sus relaciones. Además se tratan aspectos relacionados al modelado de los datos con sus respectivos elementos tales como las dimensiones, los hechos y medidas.

Por último en el tercer capítulo, Implementación y pruebas del MD de la materia económica del Sistema de Información Judicial para Tribunales, se hace referencia a la implementación de la solución. Abordando cómo se realizan los procesos de Extracción, Transformación y Carga (ETL) y los de la capa de inteligencia del negocio de la materia económica del Sistema de Información Estadística Judicial para los Tribunales Populares, teniendo en cuenta los requisitos y necesidades del negocio. De igual forma, este capítulo abordará el tema relacionado con las pruebas, mediante la utilización de las listas de chequeo y los casos de pruebas realizados al mercado de datos; incluyendo la carta de aceptación del cliente.

CAPÍTULO 1: FUNDAMENTOS TEÓRICOS PARA EL DESARROLLO DE UN MERCADO DE DATOS

Este capítulo aborda las principales características, metas y elementos de los almacenes y mercados de datos. Se define cómo se realiza el proceso de gestión de la información correspondiente al mercado de datos de la Sala de lo Económico y se caracterizan la metodología, técnicas y herramientas a utilizar para el desarrollo de la solución.

1.1 Sistemas de Información

Los Sistemas de Información (SI) son un conjunto de componentes que interaccionan entre sí para satisfacer las necesidades de la organización. Dentro de los SI se encuentran los sistemas transaccionales, estratégicos, expertos e inteligencia artificial y los de apoyo a las decisiones. (1)

Sistemas transaccionales: son los primeros que se implantan en las organizaciones. Requieren de mucho manejo de datos para poder realizar sus operaciones y como resultado generan grandes volúmenes de información. (1)

Sistemas estratégicos: su función primordial no es apoyar la automatización de procesos operativos ni proporcionar información para apoyar la toma de decisiones. Apoyan el proceso de innovación de productos y procesos dentro de la empresa, debido a que buscan ventajas respecto a los competidores y una forma de hacerlo es innovando o creando productos y procesos. (1)

Sistemas expertos e inteligencia artificial: los sistemas expertos utilizan las técnicas de razonamiento de la inteligencia artificial para solucionar los problemas que les plantean los usuarios del negocio. (1)

Sistemas de apoyo a las decisiones: suelen introducirse después de haber implantado los sistemas transaccionales más relevantes de la empresa, ya que estos constituyen su plataforma de información. Además, la información que generan sirve de apoyo a los mandos intermedios y a la alta administración en el proceso de toma de decisiones. (1)

Para darle solución a los problemas identificados en el departamento de estadística del TSP se decidió implementar un sistema de apoyo a las decisiones y dentro de este un mercado de datos, el cual va a permitir almacenar los datos históricos que pueden ser extraídos de diversas fuentes,

homogeneizándolos en un sistema que facilite la toma de decisiones. Ello hace posible que toda la información que posee este departamento sea útil para facilitar la estrategia de anticipar y conocer las necesidades, dificultades y problemas que presenta.

1.2 Almacén de datos

Los AD o Data Warehouse (DW) por sus siglas en inglés, surgen tras las dificultades que poseían los sistemas tradicionales y como solución a las necesidades informáticas de las empresas u organizaciones de tener almacenados datos históricos a los que se pudiera acceder para realizar consultas, reportes y ayudar a la toma de decisiones.

Según definió **Bill Inmon**, un AD se caracteriza por ser orientado a temas, integrado, no volátil y variante en el tiempo, para el apoyo a la toma de decisiones. (2):

Integrado: los datos almacenados en el AD deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas.

Orientado a temas: los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales.

Variante en el tiempo: el AD se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

No volátil: un AD existe para ser leído, y no modificado.

Los AD permiten medir y controlar el desarrollo de las variables importantes del negocio. Se basan en “la estructura de información”, es decir, el almacenamiento de la información homogénea y fiable en una estructura basada en la consulta. (3)

1.3 Mercado de datos

Un Mercado de Datos (MD) es “...una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específico.” (4)

Los mercados de datos se encuentran enmarcados dentro de un almacén de datos, representando un área específica del negocio. En ellos se pueden consultar los datos de forma rápida para la toma de decisiones, a diferencia de los almacenes de datos que son utilizados a nivel de organización o empresa, en dependencia de donde se estén utilizando.

Los MD se caracterizan por ser centrados a los requisitos de los usuarios asociados a un departamento o área de negocio y por no contener datos operacionales, siendo diferentes en este sentido de los almacenes de datos debido a que estos contienen datos operacionales. Además su cantidad de información es mucho menor que la de los almacenes de datos, por tal motivo son más sencillos a la hora de utilizarlos y de entender sus datos. (5)

1.3.1 Topología de esquema

En dependencia de las características del negocio y de la información que se posee, los AD se pueden modelar de tres maneras distintas:

Esquema en estrella: consiste en una tabla de hechos en el centro como objeto de análisis y una o varias tablas de dimensión por cada dimensión de análisis que participa en la descripción de ese hecho. En la tabla de hecho se encuentran los atributos que están destinados a medirlo, mientras que en las tablas de dimensión, los atributos se destinan a describir las métricas contenidas en los hechos. En el esquema en estrella la tabla de hechos es la única tabla del esquema que tiene múltiples uniones que la conectan con otras a través de las llaves foráneas (Ver figura 1). (6)



FIGURA 1 ESQUEMA EN ESTRELLAS

Esquema en copo de nieve: El esquema en copo de nieve es un esquema de representación derivado del esquema en estrella, en el que las tablas de dimensión se normalizan en múltiples tablas. Por esta razón, la tabla de hechos deja de ser la única tabla del esquema que se relaciona con otras tablas, y aparecen nuevas uniones (Ver figura 2). (6)

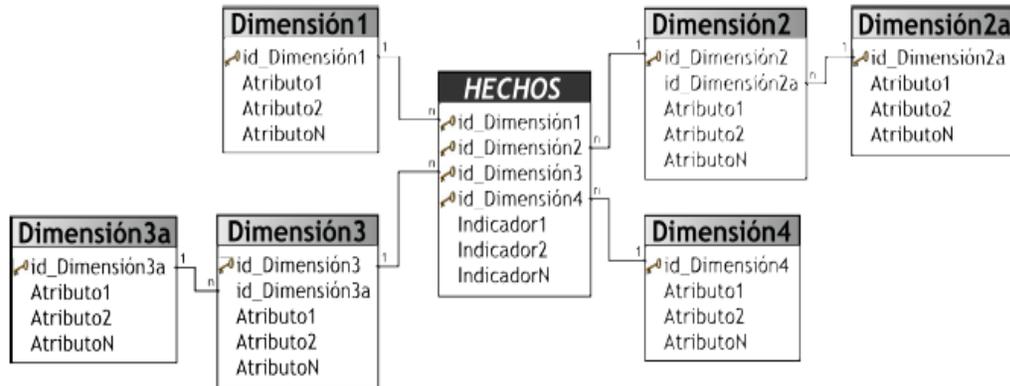


FIGURA 2 ESQUEMA COPO DE NIEVES

Constelación de hechos: es un conjunto de tablas de hechos que comparten algunas tablas de dimensiones (Ver figura 3). (6)

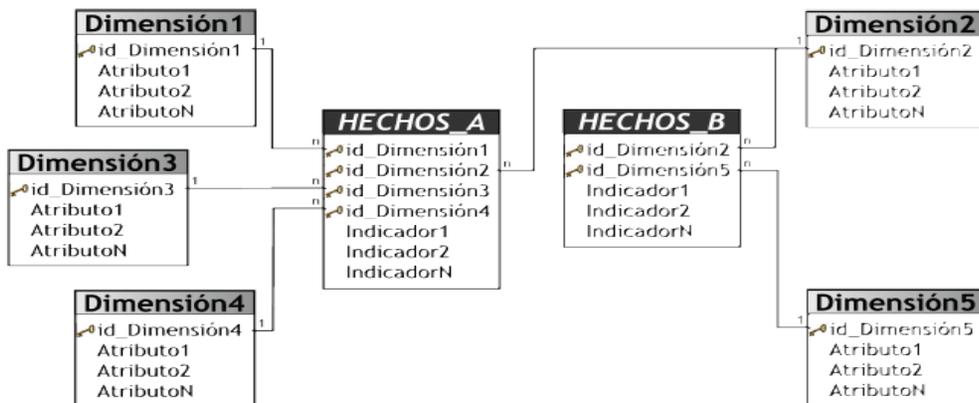


FIGURA 3 ESQUEMA CONSTELACIONES

1.3.2 Modos de almacenamiento

Para poder crear un MD funcional en un área determinada de una empresa, es necesario escoger la estructura más adecuada con el fin de realizar un buen análisis de la información. Dicha estructura puede estar montada sobre un sistema de procesamiento analítico en línea (OLAP, por sus siglas en inglés) o sobre un sistema de procesamiento de transacciones en línea (OLTP, por sus siglas en inglés). Para elegir una de estas estructuras es necesario tener en cuenta las características y los requisitos del área de la empresa, así como de los datos que estos manejan.

Los sistemas OLAP permiten que la información sea analizada de forma rápida, amigable e interactiva y facilitan la construcción de modelos cuantitativos complejos (7). Facilitan la visualización de la información multidimensionalmente usando valores cuantitativos (medidas) y cubos descriptivos. Además deben soportar análisis complejos contra un volumen inmenso de datos y analizarlos desde perspectivas diferentes. (8)

Los sistemas OLAP se pueden definir según su forma de almacenamiento de los datos. Entre los principales sistemas OLAP se encuentran los de procesamiento analítico multidimensional en línea (MOLAP, por sus siglas en inglés), procesamiento analítico relacional en línea (ROLAP, por sus siglas en inglés) y procesamiento analítico híbrido en línea (HOLAP, por sus siglas en inglés).

Los sistemas MOLAP almacenan los datos de forma lógica en una base de datos (BD) multidimensional, los cuales utilizan una arquitectura de dos niveles: uno como base de datos multidimensional, la cual se encarga del manejo, acceso y obtención de los datos y otro como motor analítico que se responsabiliza por la ejecución de los requisitos OLAP. (8)

Los sistemas ROLAP almacenan la información en una BD relacional con forma de estrella. Este sistema posee una arquitectura de tres niveles: El nivel de bases de datos el cual usa bases de datos relacionales para el manejo, acceso y obtención de los datos. El nivel de aplicación el cual es el motor que ejecuta las consultas multidimensionales de los usuarios. Por último se encuentra el motor ROLAP que se integra con los niveles de presentación a través de los cuales los usuarios realizan los análisis OLAP. (8)

Los sistemas HOLAP son híbridos entre ROLAP y MOLAP, donde la configuración más usada o típica es un motor multidimensional basado en una tecnología de almacenamiento relacional.

Por otra parte los sistemas OLTP son BD orientadas al procesamiento de transacciones. Estos pueden involucrar operaciones de inserción, modificación y borrado de datos. El proceso transaccional es típico de las bases de datos operacionales. Se encargan de analizar el negocio y de mejorar los servicios al cliente. (9)

El diseño de OLTP elimina redundancias ya que los datos están normalizados. Este sistema no es óptimo para la toma de decisiones porque proporciona capacidades muy limitadas y los usuarios deben examinar registro a registro para analizar los datos, lo cual impacta negativamente en la racionalidad del tiempo cuando se trata de mucha información. Por otra parte, admiten el acceso simultáneo de muchos usuarios que agregan y modifican datos. (9)

1.4 Metodología para el desarrollo de los mercados de datos

Dos de los enfoques más usados son los planteados por Ralph Kimball y Bill Inmon, siendo este último considerado el padre de los AD por el aporte que brindó a su desarrollo.

Bill Inmon defiende el enfoque descendente (top-down), donde los mercados de datos se crean luego de haber desarrollado el almacén de datos de la empresa. (3)

Por su parte, Kimball defiende el enfoque ascendente (bottom-up) a la hora de diseñar un AD. Plantea que primero se deben desarrollar los mercados de datos de cada una de las áreas de la empresa y luego unirlos, formando así el AD de la empresa en general.

Partiendo de lo planteado por Kimball, el centro DATEC propone una metodología para estas soluciones, adaptada a sus características y necesidades. La propuesta de metodología de desarrollo de almacenes de datos planteada por la Ingeniera Yanisbel González Hernández, cuenta con ocho fases fundamentales para su desarrollo:

Estudio preliminar y planeación: se efectúa un estudio de todos los aspectos relacionados con la entidad cliente. Además se llevan a cabo las tareas de planeación del proyecto. (10)

Requisitos: a través de entrevistas con el cliente se determinan los requisitos de información. Se hace levantamiento de las fuentes de datos y el análisis de los requisitos. (10)

Arquitectura: se definen las vistas arquitectónicas de la solución, aspectos como, los subsistemas y componentes, la seguridad, la comunicación y la tecnología a utilizar. (10)

Diseño e Implementación: se define el diseño de las estructuras de almacenamiento de datos, se diseñan los procesos de integración de datos como, el mapa lógico de datos, los cubos OLAP para la presentación de la información, así como el diseño gráfico de la aplicación definido por el cliente. Después se implementan cada uno de los subsistemas (repositorio de datos, integración de datos, presentación de datos). (10)

Prueba: se realizan las pruebas que validan la calidad del producto, comenzando por las Pruebas de Unidad, las Pruebas de Integración y Sistema, hasta llegar a las Pruebas de Aceptación con el cliente final. Esta fase no es la única en la que se realizan pruebas durante el desarrollo del proyecto, en todas las fases hay actividades de aseguramiento de la calidad. (10)

Despliegue: consta de dos etapas, despliegue piloto, donde se configuran los servidores necesarios y se instalan las herramientas según la arquitectura definida, se carga una muestra de los datos en un ambiente controlado, con el fin de mostrarle al cliente final el sistema en funcionamiento. Una vez aceptada la solución por el cliente, se realiza la carga histórica de los datos, puede ser en el mismo entorno que el despliegue piloto u otro, todo depende de las condiciones que establezca el cliente. Además se realiza la capacitación y transferencia tecnológica de la solución a los clientes. El resultado fundamental es la solución desplegada en el entorno real y en correcto funcionamiento. (10)

Soporte y Mantenimiento: comienza cuando la solución está implantada y en explotación, y se ejecuta según el contrato firmado y las condiciones de soporte establecidas. Puede realizarse a través de variados servicios, que pueden ser soporte en línea, vía telefónica, web, correo u otros y el acompañamiento al cliente. Además se realizan las tareas de manteniendo de la aplicación tan

necesarias para este tipo de desarrollo y que garantiza el adecuado funcionamiento y crecimiento del almacén de datos. (10)

Gestión del proyecto: esta fase se ejecuta a lo largo de todo el ciclo de vida del proyecto. Es aquí donde se controla, gestiona y chequea todo el desarrollo, los gastos, las utilidades, los recursos, las adquisiciones, los planes y cronogramas entre otras actividades relacionadas con la gestión de proyectos. Esta fase es la columna vertebral del proyecto y si no se ejecuta de forma continua y correcta, el proyecto puede fracasar. (10)

Esta metodología es flexible y puede ser adaptada al ambiente de cualquier organización que desarrolle un AD. Define claramente las actividades y tareas que deben realizarse para el adecuado desarrollo de un AD. Utiliza un lenguaje estandarizado y ampliamente utilizado para la representación de los requisitos del sistema y los casos de uso, lo cual facilita su utilización. La metodología está concebida para que pueda ser implementada en independencia de las herramientas que se deseen utilizar en el desarrollo del AD (10). Debido a las fortalezas planteadas se ha elegido para esta investigación la propuesta de metodología para el desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio de DATEC, la cual guiará todo el proceso de desarrollo del mercado de datos y será desarrollada hasta la quinta fase. Las fases restantes serán realizadas por otros especialistas vinculados al proyecto de Tribunales.

1.5 Herramientas empleadas para el desarrollo de los mercados de datos

Actualmente las herramientas informáticas están en constante evolución y mejoramiento, lo cual facilita el trabajo de los ingenieros que se encuentran desplegando aplicaciones informáticas, esto no basta para obtener buenos resultados en las soluciones, pues se necesita conocer y escoger correctamente las herramientas que se utilizarán.

1.5.1 Herramientas de Modelado

Las herramientas de modelado constituyen un conjunto de programas los cuales facilitan el trabajo de los ingenieros de software y desarrolladores durante el ciclo de vida de un software o producto. Dentro de

estas herramientas se encuentran las de ingeniería de software asistida por computadoras (CASE, por sus siglas en inglés).

Las herramientas CASE constituyen un conjunto de ayudas para el desarrollo de programas informáticos, por lo que cada día toman más auge a nivel mundial en la planeación y ejecución de proyectos que involucran sistemas informáticos.

Dentro de estas herramientas se encuentran:

- Embarcadero ER/Studio
- EasyCase Profesional
- Case Studio 2
- Rational Rose
- ArgoUML
- Visual Paradigm for UML

En esta investigación se empleará la herramienta **Visual Paradigm for UML 8.0**, la cual permite el modelado visual para todos los tipos de diagramas del Lenguaje Unificado de Modelado (UML, por sus siglas en inglés). Ayuda al desarrollo de software y posibilita una rápida construcción de las aplicaciones con alta calidad. Es factible a la hora de dibujar diagramas de clases, y generar script para diferentes sistemas gestores de base de datos siendo además multiplataforma. (11)

1.5.2 Sistema Gestor de Base de Datos

El sistema gestor de base de datos se define como el conjunto de programas que administran y gestionan la información contenida en una BD. Estos ayudan a mantener la integridad de la información dentro de las BD, controlar la seguridad y privacidad de los datos y la manipulación de los mismos. (12) Dentro de estos sistemas se encuentra PostgreSQL, MySQL, Oracle y SQL Server.

La herramienta que se utilizará como sistema gestor de base de datos es el **PostgreSQL en su versión 9.1** por ser una herramienta multiplataforma que funciona en Linux, Os, Windows, entre otros. Además es de código abierto, por lo que no es manejada por ninguna compañía en específico y no existe una

representación local oficial. PostgreSQL es de bajo costo en su adquisición, no presenta problemas de licencia, brinda soporte gratuito en sus comunidades, puede ser modificado según las necesidades, es de fácil manejo y su instalación es muy sencilla.

En cuanto a la arquitectura de la herramienta, PostgreSQL utiliza un modelo cliente/servidor y usa multiprocesos para garantizar la estabilidad del sistema. El servidor soporta múltiples conexiones concurrentes de clientes. Con este propósito ejecuta nuevos procesos para cada conexión. Este proceso es transparente para los usuarios. (13)

A pesar de las ventajas antes descritas, PostgreSQL necesita de una herramienta para desarrollar y administrar la base de datos, para lo cual se utilizará **PgAdmin en su versión 1.14**. Esta es una herramienta multiplataforma, donde se pueden realizar modificaciones a los datos y manipular las propiedades en el momento. Además tiene soporte en español, lo que facilita su uso. (14)

1.5.3 Herramientas para la limpieza, estandarización e integración de los datos

Para limpieza y estandarización de los datos se utilizará **DataCleaner 1.5.3**, que permite detectar y corregir errores que puedan tener los datos a integrar, ya que trata de entender su estructura, formato y nivel de calidad. Esta herramienta es muy fácil de usar, genera informes y gráficos que permiten al usuario determinar la calidad de los datos de forma muy rápida. Se destaca por ser una herramienta de código abierto y sin coste de licencia.

Dentro de las herramientas más usadas para el proceso de ETL se encuentran Scriptella, Octopus, Talend y Pentaho Data Integration. De estas herramientas se estará utilizando en la presente investigación el **Pentaho Data Integration (PDI) 4.2.0** la cual es una herramienta libre y tiene como objetivo que el proceso de ETL sea más fácil de generar, mantener y desplegar.

PDI incluye cuatro herramientas fundamentales:

1. SPOON: permite diseñar las transformaciones ETL en un entorno gráfico.
2. PAN: permite ejecutar las transformaciones realizadas en SPOON.
3. CHEF: permite crear trabajos teniendo un control sobre estos.

4. KITCHEN: permite la ejecución de los trabajos diseñados con CHEF.

PDI permite evitar grandes cargas de trabajo manual frecuentemente difícil de mantener y de desplegar. Es multiplataforma ya que puede ser usada en Windows, Unix y Linux. Utiliza JavaScript para los cálculos y fórmulas. Se puede realizar scripting con SQL, Java, Shell y fórmulas Open Office. Es una herramienta basada en metadatos, posee una amplia documentación online en la web además de foros de Pentaho y la comunidad Pentaho. Soporta PostgreSQL, Oracle, DB2, SQL Server y MySQL. Para distribuir las tareas de ETL a través de múltiples servidores soporta la arquitectura de procesamiento en paralelo. Todas estas ventajas la hacen ser una herramienta muy usada en el mundo. (15)

1.5.4 Herramientas para la Inteligencia de Negocios (BI)

Se debe recordar que no es suficiente con almacenar datos, es necesario procesarlos para convertirlos en información importante, por lo cual existen distintas herramientas para desarrollar la inteligencia de negocio aplicando técnicas OLAP, destacándose: Pentaho Schema Workbench, Pentaho BI Server, Servidor Mondrian OLAP, Servidor Web Apache Tomcat o Jakarta Tomcat.

Pentaho Schema Workbench 3.2.1 es un diseñador de interfaz que permite crear y probar cubos OLAP Mondrian visualmente, los cuales proveen un mecanismo para buscar datos con rapidez y tiempo de respuesta uniforme independientemente de la cantidad de datos en el cubo. No requiere que un cubo físico real sea construido; sólo que el modelo de metadatos se cree. Además, permite la ejecución de consultas MDX (Expresiones Multidimensionales). (16)

Pentaho BI Server 4.8 provee el soporte y la infraestructura necesarios para crear soluciones de inteligencia empresarial (BI) a problemas de negocios. El marco proporciona los servicios básicos, incluidos autenticación, registro, auditoría y servicios web. La plataforma también incluye un motor de solución que integra reportes, análisis, tableros de comandos y componentes de minería de datos.

BI Server funciona como un sistema basado en administración web de informes, el servidor de integración de aplicaciones y un motor de flujo de trabajo ligero. Está diseñado para integrarse fácilmente en cualquier proceso de negocio. (17)

Servidor Mondrian OLAP 3.0.4 es una de las aplicaciones más importantes de Pentaho, está escrito en java y es de código abierto. Este permite analizar grandes cantidades de datos almacenados en bases de datos de una forma interactiva sin necesidad de escribir las sentencias que para esto harían falta en SQL. Este servidor gestiona la comunicación entre una aplicación OLAP y la base de datos con los datos fuente. Dentro de sus características se destaca la compatibilidad con las expresiones multidimensionales (MDX) y el lenguaje de marcas extensible (XML) para análisis. Además, funciona sobre las bases de datos estándar del mercado: Oracle, DB2, SQL-Server, MySQL, PostgreSQL, lo cual habilita y facilita el desarrollo del negocio basado en la plataforma Pentaho. (18)

Servidor Web Apache Tomcat 6.6 es desarrollado en un entorno abierto y participativo y publicado bajo la licencia Apache versión 2, por miembros de la Apache Software Foundation y voluntarios independientes. El hecho de que Tomcat fue escrito en Java, hace posible que funcione en cualquier sistema operativo que disponga de la máquina virtual Java. Esta herramienta no viene incluida dentro de la suite de Pentaho, pero si la utiliza como servidor debido a que es una herramienta robusta y altamente eficiente. (19)

Por las características descritas, durante la investigación se estarán utilizando como herramientas OLAP para el proceso de inteligencia de negocio el **Pentaho Schema Workbench 3.2.1**, **Pentaho BI Server 4.8**, **Servidor Mondrian OLAP 3.0.4** y el **Servidor Web Apache Tomcat 6.6**.

Conclusiones

Durante el desarrollo del capítulo se describieron los tipos de SI existentes y las características de los almacenes y mercados de datos, definiéndose como mejor solución para la materia económica del departamento de estadística, el desarrollo de un mercado de datos. Se escogió la propuesta de metodología para el desarrollo de soluciones de Almacenes de Datos e Inteligencia de Negocio de DATEC, que facilitará y agilizará el trabajo durante el desarrollo de la solución propuesta. También se estudiaron las características de distintas herramientas, lo cual permitió escoger las que contribuirán a la estabilidad del producto final.

CAPÍTULO 2: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS ECONÓMICO DEL SISTEMA DE INFORMACIÓN ESTADÍSTICO JUDICIAL PARA LOS TRIBUNALES POPULARES

En este capítulo se abordan aspectos referentes al levantamiento de requisitos de la materia económica del Sistema de Información Estadístico Judicial para los Tribunales Populares y a la definición de las reglas del negocio. Se describen los casos de usos identificados con sus relaciones. Además se analiza la arquitectura del mercado con sus respectivos subsistemas, tratando aspectos relacionados con el modelado de los datos, sus dimensiones y hechos, así como el perfilado de las fuentes de datos.

2.1 Análisis del Negocio

El Tribunal Supremo Popular es la institución rectora de la justicia en Cuba, el cual cuenta con un departamento estadístico responsable de recoger y supervisar la gestión estadística judicial del país. El departamento estadístico recoge la información mensualmente en modelos que llegan desde sus diferentes sucursales en cada provincia. Dicha información es agrupada en cuatro materias: Civil, Laboral, Penal y Económico.

La materia económica ha sufrido varios cambios en cuanto a la recogida de los datos en el período del 2002 al 2012. La información que maneja esta materia se encuentra dividida en distintos modelos, que responden a los litigios económicos que existen entre distintas empresas o instituciones del país como son las causales, tramitaciones y resoluciones de procesos ordinarios y ejecutivos. Estos litigios se concentran en distintas metodologías que responden a la forma de organización y recogida de la información en el departamento estadístico del TSP. La metodología 1 recoge los datos almacenados durante la etapa del 2002 al 2008. En la metodología 2 se encuentra la información referente al período del 2009 al 2012 y por último en la metodología 3 se almacenarán los datos del 2013 en adelante.

Frecuentemente a este departamento le solicitan datos estadísticos para ser analizados por los directivos de esta entidad y del país, por lo que es de gran importancia que toda la información esté correctamente estructurada y organizada.

2.2 Técnicas de captura de requisitos

La captura de requisitos es una de las piezas fundamentales del proyecto ya que desde los inicios del desarrollo de sistemas, los ingenieros han presentado problemas con la identificación de los requisitos. Los datos son extraídos de personas que quizás no posean un lenguaje informático, por lo que se debe saber qué es lo que verdaderamente quiere el usuario. Estas técnicas permiten obtener los requisitos con calidad para minimizar los problemas relacionados con la gestión de requisitos.

Existen varias técnicas para la captura de requisitos, dentro de las que se encuentran: entrevistas y cuestionarios, observaciones, discusiones, entre otras.

En el presente trabajo de diploma se utilizaron las entrevistas, observaciones y discusiones, que permitieron la interacción con el usuario, obtener sus necesidades, comprender mejor el problema y definir los objetivos de la posible solución.

2.3 Reglas del negocio

Las reglas del negocio definen políticas o condiciones que debe cumplir un sistema. Estas constituyen la entrada fundamental para los procesos de diseño del almacén, los procesos ETL y la capa de BI. A continuación se muestran algunas reglas y las restantes podrán ser consultadas en el artefacto "DATEC_SIEJT_Reglas de negocio y transformación".

Reglas de variables

- Porcentaje de litigios contractuales no admisibles = $(\text{Litigios contractuales no admisibles} / \text{Litigios contractuales resueltos}) * 100$.
- Porcentaje de litigios contractuales con lugar = $(\text{Litigios contractuales con lugar} / \text{Litigios contractuales resueltos}) * 100$.
- Porcentaje de litigios contractuales con lugar en partes = $(\text{Litigios contractuales con lugar en partes} / \text{Litigios contractuales resueltos}) * 100$.
- Porcentaje de litigios contractuales sin lugar = $(\text{Litigios contractuales sin lugar} / \text{Litigios contractuales resueltos}) * 100$.

- Porcentaje de litigios contractuales desistidos = $(\text{Litigios contractuales desistidos} / \text{Litigios contractuales resueltos}) * 100$.
- Porcentaje de en término para ejecutar por la parte = $(\text{en término para ejecutar por la parte} / \text{ejecuciones pendientes}) * 100$.
- Porcentaje de en trámite para cumplimiento forzoso por el tribunal = $(\text{en trámite para cumplimiento forzoso por el tribunal} / \text{ejecuciones pendientes}) * 100$.
- Porcentaje de los pendientes de oficio a registros públicos = $\text{pendientes de oficio a registros públicos} / \text{ejecuciones pendientes}$.

Reglas de transformación

- Cuando exista un número decimal se redondea hasta 2 números después de la coma.
- Cada vez que exista un campo nulo se pondrá en ese campo 0.

2.4 Especificación de requisitos

Los requisitos son condiciones que debe cumplir un sistema para su correcto funcionamiento. Estos se clasifican en requisitos de información, funcionales y no funcionales.

Requisitos de información

Los requisitos de información representan toda la información que debe estar disponible para el usuario final. Constituyen una entrada principal para el proceso de inteligencia de negocio y para futuros reportes. A continuación se muestran algunos requisitos de información de la presente investigación. Los restantes requisitos de información podrán ser consultados en el artefacto "DATEC_SIEJT_Especificación de Requisitos".

RI1-.Obtener la cantidad de litigios extracontractuales de la metodología 1 por causales, entidades, tiempo, sala y tribunal provincial.

RI2-.Obtener cantidad de procesos de medio ambiente y otros ilícitos civiles por dpa (división político-administrativa), tiempo, tribunal provincial, sala, entidades demandantes y expedientes.

R13-.Obtener cantidad de procesos ejecutivos de la metodología 1 por dpa, tiempo, tribunal provincial, sala, entidad demandante y expedientes ejecutivos.

R14-.Obtener cantidad de procesos ordinarios contractuales de la metodología 2 las causales, dpa, sujeto demandante, tribunal provincial, sala y tiempo.

R15-.Obtener cantidad de diligencias previas del proceso ejecutivo por dpa, diligencias, tiempo, tribunal provincial, sala y sujeto demandante.

R16-.Obtener cantidad de procesos ordinarios de la metodología 2 por dpa, tiempo, tribunal provincial, sala, sujeto demandante y expedientes.

R17-.Obtener cantidad de procesos ejecutivos en trámites de la metodología 2 por dpa, tramitaciones, tiempo, sujeto demandante, tribunal provincial y sala.

R18-.Obtener cantidad de sentencias por ejecución de sentencias, tiempo, banco, dpa, tribunal provincial y sala.

R19-.Obtener cantidad de asistencias al arbitraje internacional por asistencias al arbitraje, tiempo, sujeto demandante, tribunal provincial, sala y dpa.

R110-.Obtener cantidad de los procesos ordinarios contractuales de la metodología 3 por causales, tiempo, tribunal provincial, sala y dpa.

Requisitos funcionales

Los requisitos funcionales describen lo que el sistema o software debe hacer. Son funcionalidades del sistema enfocadas a las necesidades de información que presentan los usuarios (20). A continuación se muestran algunos de los requisitos funcionales más significativos de la presente investigación. Los requisitos funcionales restantes podrán ser consultados en el artefacto "DATEC_SIEJT_Especificación de Requisitos".

RF1: Realizar almacenamiento y captura de la información

El mercado de datos económico contará con toda la información recopilada en los últimos 10 años, la cual cuenta con un tamaño de 1,00 GB.

RF2: Mostrar resúmenes mensuales

Mensualmente se insertarán los datos correspondientes al mes activo.

RF3: Crear metadatos asociados a la organización.

El sistema debe contar con metadatos que permitan saber que transformación fue ejecutada por última vez, en qué fecha y a qué hora se realizó la actualización de los datos.

RF4: Exportar a PDF y a Excel.

El sistema debe permitirle al usuario exportar los datos a Excel o PDF.

Requisitos no funcionales

Los requisitos no funcionales especifican las cualidades o propiedades del sistema de acuerdo con las necesidades y especificaciones del cliente (20). Están estrechamente relacionados con los requisitos funcionales. A continuación se muestran algunos requisitos no funcionales de la presente investigación. Los requisitos no funcionales restantes podrán ser consultados en el artefacto "DATEC_SIEJT_Especificación de Requisitos".

Requisitos de usabilidad

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

El sistema debe tener una interfaz gráfica uniforme que incluya menú y opciones de navegación al usuario.

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

Los títulos de los componentes de la interfaz, los mensajes para interactuar con los usuarios y los mensajes de error, deben ser en idioma español y tener una apariencia uniforme en todo el sistema. Los mensajes de error deberán ser lo suficientemente informativos para dar a conocer la severidad del error.

RNF3: Establecer tiempo de entrenamiento requerido para que los usuarios sean productivos operando el sistema.

El tiempo de entrenamiento requerido para que usuarios sean productivos operando el sistema deberá ser entre siete y 14 días. Para aquellos usuarios con un nivel avanzado se define como valor máximo siete días. Para lograr el cumplimiento de los tiempos establecidos por parte de los usuarios es necesario un dominio del funcionamiento del negocio en correspondencia con el rol que ocupen.

RNF4: Agilizar el acceso a los reportes.

Agilizar el acceso a los reportes del AD mediante la distribución de la información por áreas de análisis. El usuario podrá acceder de manera rápida a la información que solicita en el área pertinente de acuerdo al objetivo de la solicitud.

RNF5: Diseñar un reporte del Almacén de Datos de manera sencilla.

Un usuario con conocimientos básicos del sistema podrá diseñar un reporte del AD de sencilla sin necesidad de ser un experto en las herramientas requeridas para ello.

Requisitos de eficiencia

RNF6: Garantizar que la obtención de un reporte sea de forma rápida.

El tiempo promedio para la obtención de un reporte será de cinco segundos.

Requisitos de diseño

RNF7: Lograr que los elementos definidos en el almacén tengan una estructura homogénea.

Las estructuras del Almacén de Datos se nombrarán de una manera estándar teniendo en cuenta el tipo de estructura que se maneje. Se definen convenciones de nombrado con el objetivo de manejar un vocabulario común que permita un entendimiento claro y conciso de las estructuras por parte de los desarrolladores. Ver la tabla 1 en el artefacto "DATEC_SIEJT_Especificación de Requisitos".

RNF8: Utilizar los lenguajes de programación definidos durante la investigación.

Para la programación en el Almacén de Datos se utilizará PL/pgSQL como lenguaje dentro del Sistema Gestor de base de datos y para la implementación de los procesos de integración de datos el lenguaje JavaScript. Además se utilizará el lenguaje MDX para realizar las consultas.

Requisitos de software

RNF9: Utilizar el Sistema Gestor de Base de datos definido durante la investigación.

El Gestor de Base de Datos para implementar el Almacén de Datos es el PostgreSQL en su versión 9.1. Como Interfaz de Administración del Gestor de Base de Datos se usará el PgAdmin III en su versión 1.14.

RNF10: Utilizar la herramienta de integración de datos definida durante la investigación.

Para la integración de datos desde el Sistema para la Gestión de Proyectos hacia el Almacén de Datos se usará la herramienta Pentaho Data Integration en su versión 4.2.0.

RNF11: Utilizar las herramientas para la implementación de la capa de inteligencia de negocios definidas durante la investigación.

De la suite de Pentaho, Pentaho BI Suite Community Edition en su versión 4.8, se usarán los siguientes componentes.

- Schema Workbench 3.2.1: es la herramienta gráfica que se utiliza para construir el esquema de metadatos multidimensional que soportará la creación de los reportes multidimensionales.
- Pentaho BI Server 4.8: es el servidor de visualización de la suite, que se encarga de visualizar los

reportes, tableros de control digital, controlar el acceso a la información y unificar en una solución de inteligencia de negocios el uso de las demás herramientas que componen la suite.

- Pentaho Administrator Console 4.8: es la herramienta a usar para administrar el Pentaho BI Server, que permite la administración de las conexiones a las bases de datos, tareas programadas así como los roles y usuarios.

RNF12: Instalar en las estaciones de trabajo el software necesario para el correcto funcionamiento del sistema.

Las configuraciones de software de las máquinas clientes deben contar al menos los siguientes elementos:

- Firefox 2.0 o superior.
- Java Virtual Machine 1.6 o superior y Schema Workbench 4.2.0 en caso de que un usuario capacitado requiera la construcción de esquemas multidimensionales para el diseño de nuevos reportes.

Requisitos de hardware

RNF13: Proporcionar características mínimas de hardware

Los servidores y las estaciones de trabajo deben contar con los siguientes requisitos de hardware para lograr una explotación aceptable del sistema:

Servidores:

- Windows Xp.
- 1 GB RAM.
- 1 Microprocesador dual-core.

Estación de trabajo:

- RAM: al menos 512 MB.
- Espacio en el disco: al menos 1GB.

- Procesador: dual-core o superior.

2.5 Diagrama de Casos de Uso del Sistema

El diagrama de casos de uso (CU) del sistema representa gráficamente a los procesos y su interacción con los actores (21), además es utilizado para ilustrar los requisitos del sistema. En la figura 4 queda reflejado el diagrama de CU de la presente investigación, el cual está compuesto por seis casos de uso de información (CUI), siete casos de uso funcionales (CUF) y tres actores.

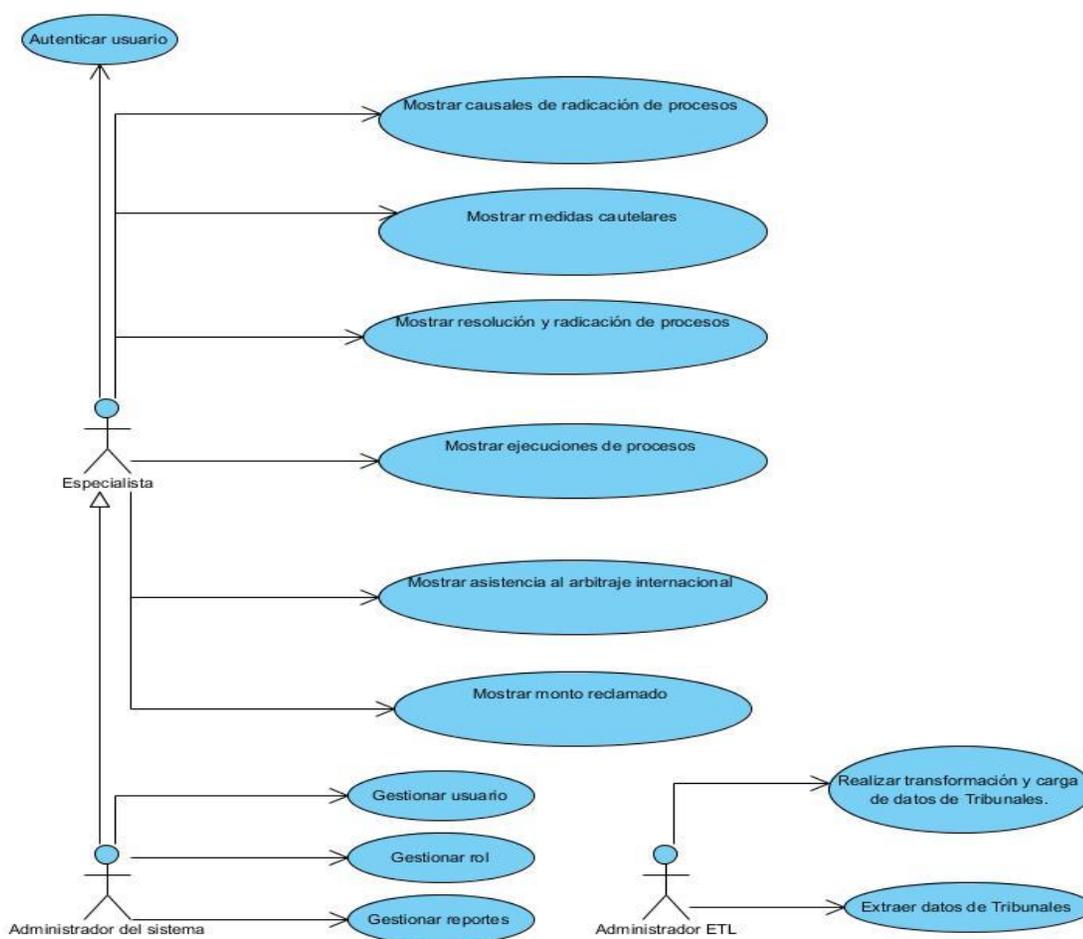


FIGURA 4 DIAGRAMA DE CASOS DE USO DEL SISTEMA

2.5.1 Actores del Sistema

En la siguiente tabla se describen los actores del Mercado de datos económico para el SIEJT (Ver tabla 1).

TABLA 1 ACTORES DEL SISTEMA Y SU DESCRIPCIÓN

Actor	Descripción
Especialista	El especialista es el encargado de inicializar los casos de uso de información: Mostrar causales de radicación de procesos, Mostrar medidas cautelares, Mostrar resolución de procesos, Mostrar asistencia al arbitraje internacional, Mostrar monto reclamado y Autenticar usuario.
Administrador del Sistema	Además de las tareas del especialista, tiene como tarea Gestionar rol, Gestionar reportes y Gestionar usuarios.
Administrador de ETL	Se encarga de extraer los datos de la materia económica, los transforma y los carga.

2.5.2 Especificación de Casos de Uso

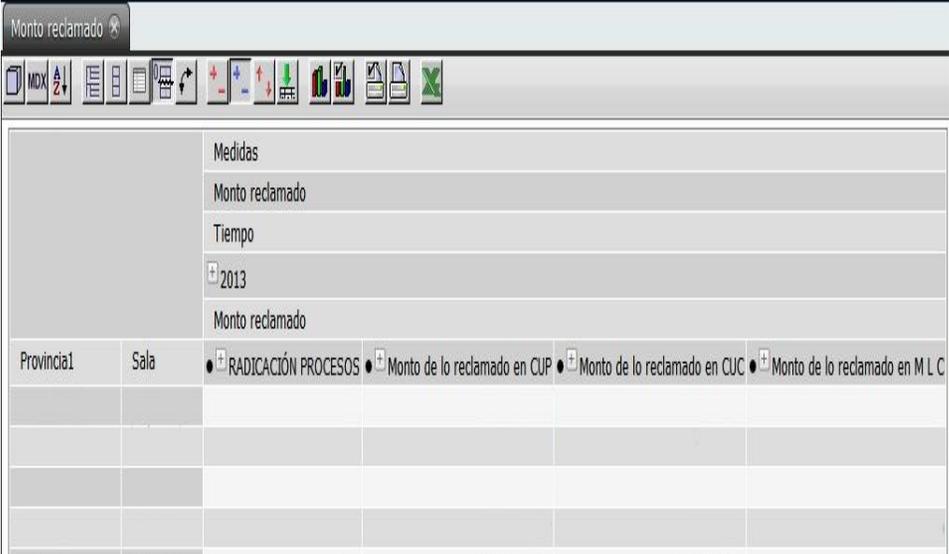
En la tabla 2 se describe el CUI Mostrar monto reclamado del mercado de datos económico. Las descripciones de los casos de uso restantes se encuentran en el artefacto DATEC_SIEJT_Modelo de Casos de Uso.

TABLA 2 CASO DE USO MOSTRAR MONTO RECLAMADO

Caso de Uso	Mostrar monto reclamado
Tipo	Información
Actores	Especialista, Administrador del Sistema
Resumen	El CU comienza cuando el especialista entra al sistema y selecciona mostrar monto reclamado, se realiza un proceso de selección de la información a mostrar y el CU termina cuando se muestra el reporte

Capítulo 2: Análisis y diseño del mercado de datos Económico del Sistema de Información Estadístico Judicial para los Tribunales Populares

	con la información seleccionada.	
Precondiciones	Compleitud del almacén. Carga de los Datos. El usuario debe autenticarse en el sistema.	
Referencias		
Prioridad	Crítico.	
Flujo Normal de Eventos		
Actor	Respuesta del Sistema	
1. El especialista entra al sistema.	2. El sistema muestra el A.A.G.SIEJT.	
3. El especialista despliega el A.A.G.	4. El sistema muestra el A.A. Indicadores de Tribunales.	
5. El especialista despliega el A.A.	6. El sistema muestra los libros de trabajo.	
7. El especialista selecciona el L.T. Indicadores Mensuales.	8. El sistema muestra el listado de reportes asociados a ese L.T.	
9. El especialista selecciona el reporte que desea visualizar.	10. El sistema muestra el reporte y las opciones para configurar los reportes(ver CU Visualizar reporte)	
Opciones de reportes de Mostrar monto reclamado		
Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad
Variables de entrada relacionadas con los hechos hech_monto_reclamado_m3 <ul style="list-style-type: none"> ➤ Tiempo ➤ Monto reclamado ➤ Sala ➤ Provincia 	Variables de salida disponibles en el caso de uso Mostrar monto reclamado. <ul style="list-style-type: none"> • Monto reclamado 	Rango de tiempo en que se solicitan las variables de salida:

		<ul style="list-style-type: none"> • Mensual.
<p>Prototipo de interfaz</p>		
<p>Poscondiciones</p>	<p>Disponibilidad de opciones (cruces de variables) de reportes relacionados con el caso de uso Mostrar monto reclamado.</p>	

2.6 Arquitectura del mercado de datos Económico

La arquitectura trazada para este MD está compuesta por la fuente de datos y tres subsistemas principales. La fuente de datos está formada por archivos Excel y dbf que contienen la información de los conflictos económicos entre diferentes instituciones o empresas del país y se encuentra ubicada en el servidor del departamento estadístico del TSP, con la estructura de carpetas Tribunal Supremo Popular\Económico\Fuentes. El subsistema de integración obtiene la información de la fuente para realizar los procesos de extracción, transformación y carga de los datos al mercado económico. Luego, el subsistema de integración interactúa con el subsistema de almacenamiento donde se guarda la información requerida en la BD tribunal_dwh. Por último, el subsistema de visualización se encarga de

mostrar los datos existentes en la BD a los especialistas de la materia económica del TSP, contribuyendo a la toma de decisiones (Ver figura 5).

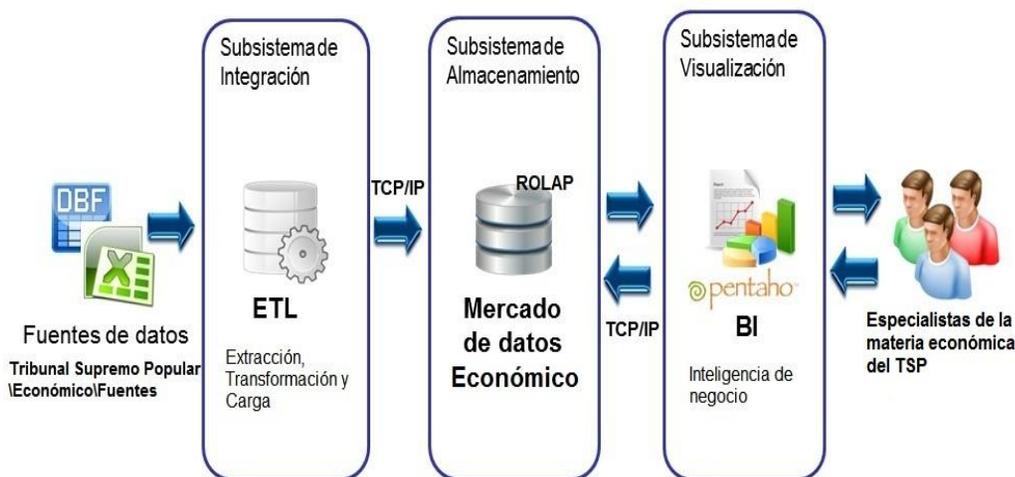


FIGURA 5 ARQUITECTURA DEL MERCADO DE DATOS ECONÓMICO

2.7 Diseño del mercado de datos Económico

Para obtener el diseño completo del mercado de datos Económico, se procede a diseñar los subsistemas de almacenamiento, integración y visualización. Esto permitirá realizar una correcta implementación de la solución.

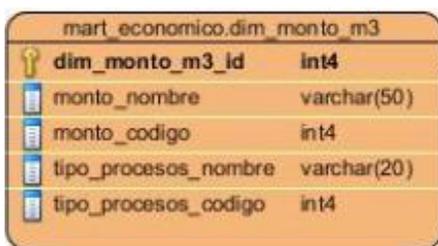
2.7.1 Diseño del Subsistema de almacenamiento

Para lograr un adecuado y correcto funcionamiento del MD se realiza el modelo dimensional de los datos, que contiene las tablas de hechos identificadas en el negocio, las dimensiones y las relaciones que existen entre estas. Se selecciona como modo de almacenamiento el Procesamiento Analítico Relacional en Línea (ROLAP) debido a que el SGBD utilizado no soporta el almacenamiento multidimensional de los datos y permitirá un mejor entendimiento y análisis de la información.

Dimensiones

Las tablas de dimensiones se relacionan con las tablas de hechos, contienen descripciones textuales de la empresa y suelen tener pocas filas y varias columnas. Cada dimensión se define por su llave primaria que es única. Esta llave sirve como base para la integridad referencial con cualquier tabla de hechos a la que se una. Las dimensiones pueden estar compuestas por jerarquías o niveles, que permiten detallar o agregar la información de forma dinámica. (22)

En la investigación se identificaron 25 dimensiones. A continuación se muestra la dimensión “dim_monto_m3” que describe los montos y los tipos de procesos que se analizan en el mercado de datos económico. El resto de las dimensiones se podrán encontrar en el artefacto “DATEC_SIEJT_Económico_Especificación del modelo de datos” (Ver figura 6).



mart_economico.dim_monto_m3	
 dim_monto_m3_id	int4
 monto_nombre	varchar(50)
 monto_codigo	int4
 tipo_procesos_nombre	varchar(20)
 tipo_procesos_codigo	int4

FIGURA 6 DIMENSIÓN MONTO M3

Lista de dimensiones

- D1-. dim_diligencias_m2
- D2-. dim_dpa_provincia
- D3-. dim_temporal_mes
- D4-. dim_ejecuciones_sentencias_m2
- D5-. dim_entidad_demandante_m1

D6-. dim_tipo_sala

D7-. dim_tramitaciones_m2

D8-. dim_entidades_demandantes_m1

D9-. dim_causales

D10-. dim_expedientes

D11-. dim_asistencia_arbitraje_m2

D12-. dim_sujeto_demandante_m2

D13-. dim_monto_m3

D14-. dim_partes_subordinadas_m3

D15-. dim_entidades_m1

D16-. dim_ejecuciones_pendientes_final_m3

D17-. dim_expediente_ejecucion_m3

D18-. dim_procesos_m3

D19-. dim_expedientes_ejecutivos_m1

D20-. dim_sectores_m3

D21-. dim_tribunal_provincial

D22-. dim_medidas_cautelares_solicitadas_m3

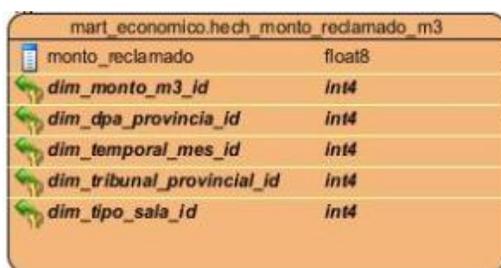
D23-. dim_banco_m2

D24-. dim_comportamiento_procesos_m3

D25-.dim_estado_ejecuciones_m3

Hechos

Una tabla de hechos es la tabla principal del modelo dimensional donde se almacenan las medidas numéricas del rendimiento de la empresa. Contiene dos o más llaves foráneas que son las claves principales de las dimensiones con las que se relaciona. (22) Durante el diseño se identificaron 20 hechos, a continuación se muestra el hecho “hech_monto_reclamado_m3” donde se recoge mensualmente el monto reclamado por los demandantes de las distintas provincias del país, el resto se podrá encontrar en el artefacto “DATEC_SIEJT_Económico_Especificación del modelo de datos” (Ver figura 7).



mart_economico.hech_monto_reclamado_m3	
monto_reclamado	float8
dim_monto_m3_id	int4
dim_dpa_provincia_id	int4
dim_temporal_mes_id	int4
dim_tribunal_provincial_id	int4
dim_tipo_sala_id	int4

FIGURA 7 HECHO MONTO RECLAMADO

Lista de hechos

H1-. hech_diligencias_proceso_ejecutivo_m2

H2-. hech_resolucion_procesos_m2

H3-. hech_resolucion_procesos_ma_otros_m1

H4-. hech_ejecucion_sentencias_m2

H5-. hech_asistencia_arbitraje_internacional_m2

H6-. hech_monto_reclamado_m3

H7-. hech_partes_subordinadas_mismo_organismo_m3

H8-. hech_tramitaciones_procesos_ejecutivos

H9-. hech_resolucion_procesos_contractuales_m1

H10-. hech_resolucion_procesos_ordinarios_m3

H11-. hech_causales_litigios_extracontractuales_m1

H12-. hech_causales_ordinarios_contractuales_m3

H13-. hech_causales_m2

H14-. hech_resolucion_procesos_ejecutivos_m1

H15-. hech_medidas_cautelares_solicitadas

H16-. hech_comportamiento_radicacion_m3

H17-. hech_control_ejecuciones_m3

H18-. hech_antiguedad_ejecuciones_pendientes_m3

H19-. hech_comportamiento_procesos_m3

H20-. hech_estado_ejecuciones_m3

Matriz Dimensional o Matriz Bus

La matriz BUS tiene como objetivo mostrar gráficamente la relación de las dimensiones con los hechos del MD. A través de la matriz se podrá verificar que no existan dos hechos que compartan exactamente

Capítulo 2: Análisis y diseño del mercado de datos Económico del Sistema de Información Estadístico Judicial para los Tribunales Populares

las mismas dimensiones en el mercado de datos. Además se pueden detectar las dimensiones que necesitan un análisis especial debido a su relación con varios hechos y la prioridad de carga de las dimensiones. En la tabla 3 queda reflejada la matriz dimensional que corresponde a la presente investigación.

TABLA 3 MATRIZ DIMENSIONAL DEL MERCADO DE DATOS ECONÓMICO

D/H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
D1	X																			
D2	X	X	X	X	X	X	X	X	X	X	X	X	X	X						X
D3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
D4				X																
D5									X					X						
D6	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
D7								X												
D8			X																	
D9										X	X	X								
D10		X	X						X	X										
D11					X															
D12	X	X			X			X					X							
D13						X														
D14							X													
D15										X										
D16																		X		
D17																	X			
D18																X				
D19														X						
D20																X	X	X		
D21	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
D22															X					
D23				X																
D24																			X	
D25																				X

Modelo Dimensional

Luego de identificar las tablas de hechos y de dimensiones, y una vez confeccionada la matriz bus para definir las relaciones entre las tablas, se realiza el diseño del modelo dimensional. En la figura 8 se muestra el modelo dimensional de la presente investigación, compuesto por 20 tablas de hechos y 25 tablas de dimensiones. La topología que se utilizó para realizar el modelo de datos fue la constelación de hechos debido a que esta presenta varias tablas de hechos que comparten dimensiones entre ellas.

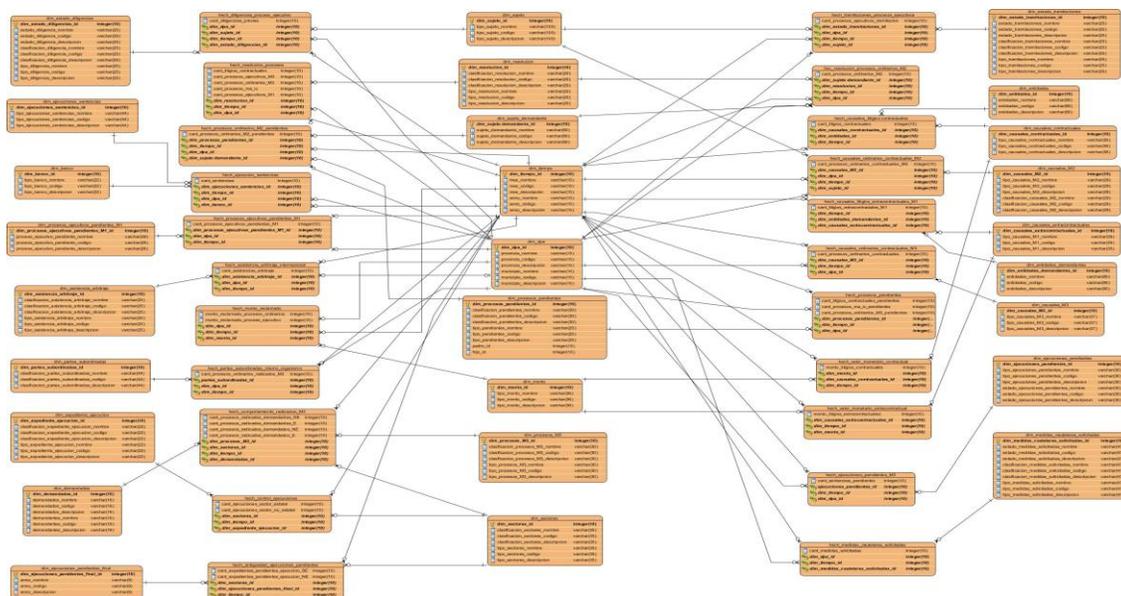


FIGURA 8 MODELO DIMENSIONAL

2.7.2 Diseño del subsistema de integración

Para realizar con calidad el diseño del subsistema de integración es necesario llevar a cabo el perfilado de los datos. El perfilado permite lograr un mejor entendimiento de los datos y verificar la existencia de valores nulos y duplicados. Se definen nuevas reglas del negocio que posteriormente pasarán a ser las reglas de transformación, las cuales se llevarán a cabo en los procesos de ETL.

Después de haber realizado el perfilado de los datos a la fuente económica del departamento de estadística judicial del TSP, se pudo determinar que en el período del 2002 al 2012 la calidad de los datos

se comportó según muestra la figura 9. Con estos resultados se identificaron nuevas reglas de transformación, que ayudaron a realizar el diseño e implementación del subsistema de integración.

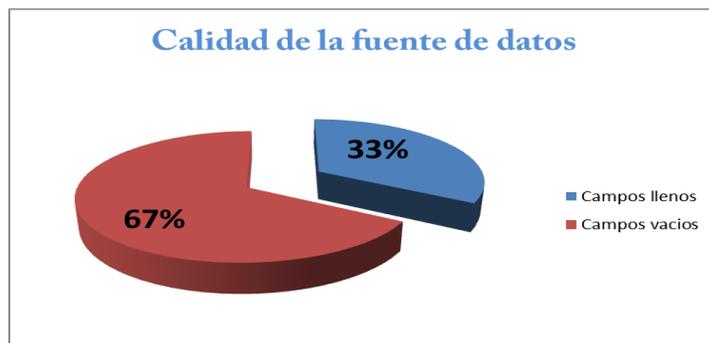


FIGURA 9 CALIDAD DE LA FUENTE DE DATOS

De igual forma, permitió obtener la distribución de los tipos de datos y la cantidad máxima de caracteres en la fuente económica del TSP, como se muestra en las figuras 10 y 11 respectivamente, lo que contribuyó a realizar el diseño del modelo dimensional.



FIGURA 10 DISTRIBUCIÓN DE LOS TIPOS DE DATOS EN LA FUENTE

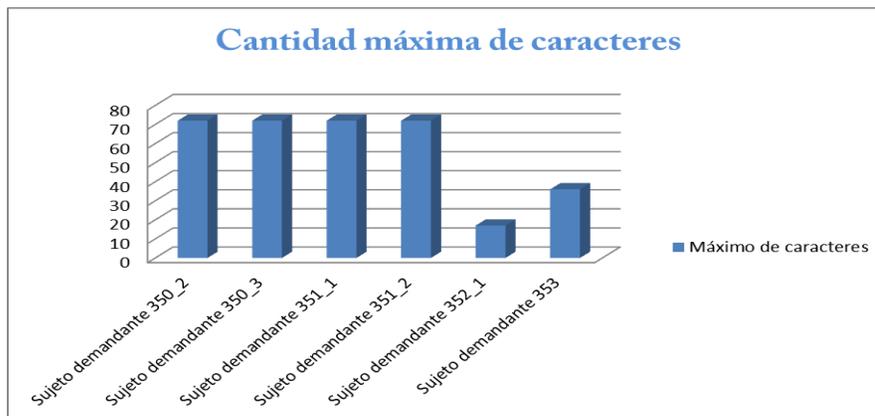


FIGURA 11 CANTIDAD MÁXIMA DE CARACTERES

Para un mejor entendimiento del sistema se realizó el diccionario de datos, el cual tiene como objetivo contribuir a la documentación del sistema fuente y especificar el significado de cada variable en el negocio con los valores que pueden tomar. En el diseño se identificaron 19 variables que corresponden a cada una de las dimensiones del mercado de datos económico, lo que ayudó a realizar las transformaciones.

Diseño de las transformaciones

Las transformaciones comienzan estableciendo y verificando la conexión con la BD. Además se realiza la extracción de los datos contenidos en los archivos Excel y dbf, las cuales se encuentran en el servidor del departamento estadístico con la estructura de carpetas Tribunal Supremo/Económico/Fuentes. Se validan y solucionan un grupo de restricciones definidas por el analista. Luego se obtienen los códigos de los indicadores y se buscan los identificadores (id) de las dimensiones en la fuente de datos y se validan. Si algunos de los id no se encuentran, la información es enviada hacia un archivo Excel para que el especialista realice las modificaciones necesarias. Luego de solucionado el error, el administrador de ETL procede a cargar la información modificada. En caso de no existir errores se insertan todos los datos en la tabla que tiene como salida dicha transformación. (Ver figura 12).

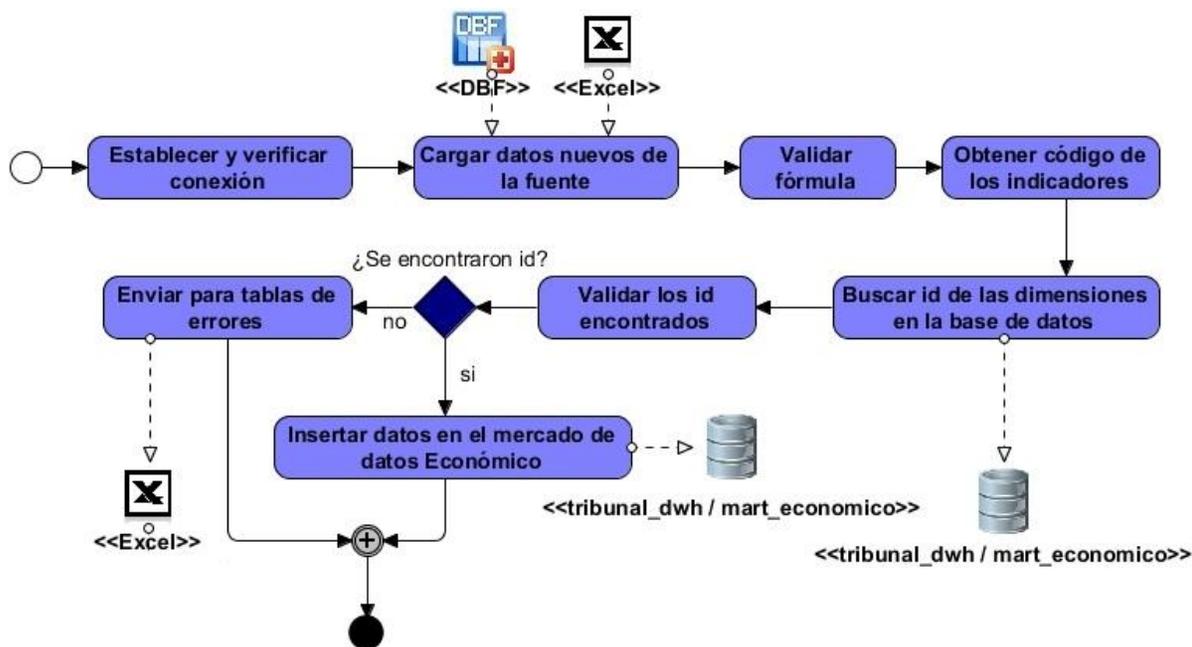


FIGURA 12 DISEÑO DE LAS TRANSFORMACIONES

2.7.3 Diseño del subsistema de visualización

La información que maneja la materia económica se encuentra dividida en distintos modelos, que se pueden concentrar en un área de análisis general (A.A.G) y siete áreas de análisis (A.A) principales. Las A.A DPA (1976-2010) y DPA (2011-actualidad), están organizadas en tres A.A, Metodología 1, Metodología 2 y Metodología 3, donde se identificaron 27 libros de trabajos (LT) y 29 reportes (Ver figura 13).

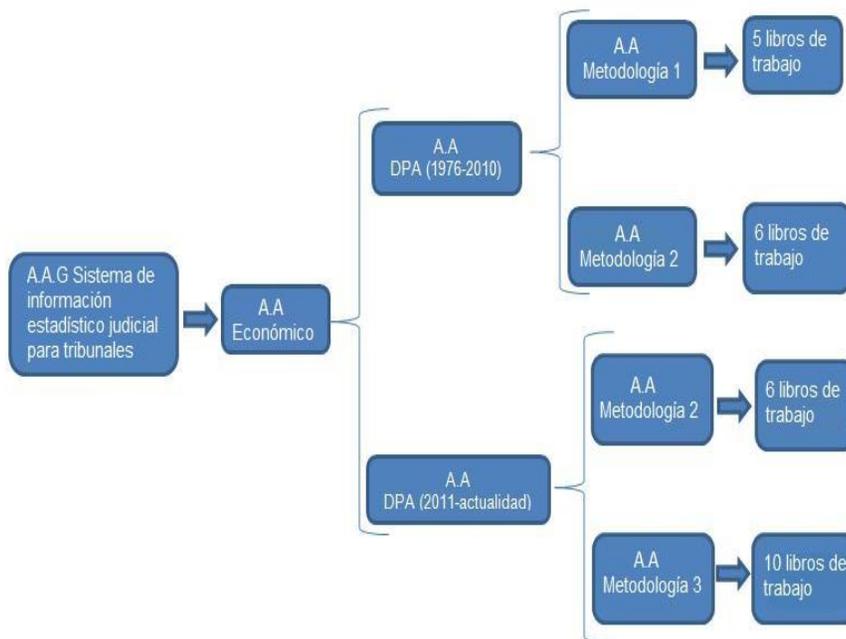


FIGURA 13 MAPA DE NAVEGACIÓN

2.8 Esquema de seguridad

Para lograr mayor seguridad de la información en el mercado de datos económico, fueron creados usuarios a los cuales se les asignó roles con los permisos necesarios para cada uno.

Seguridad en la BD

Los usuarios que tienen acceso a la base de datos son los administradores de ETL y los especialistas. Los administradores tienen acceso total a la BD y son los encargados de realizar los procesos de ETL. Por su parte, los especialistas se encargan de visualizar los datos en la BD y analizarlos.

Seguridad en el sistema

Un elemento primordial para todos los sistemas es la seguridad, por lo que se definen los usuarios y los roles que pueden acceder a la capa de visualización del MD económico, así como las acciones que estos

pueden realizar sobre los datos. A continuación se muestran los usuarios y los roles del sistema (Ver figura14).

Se creó el rol admin que es el administrador del sistema, el cual tiene privilegio de crear nuevos usuarios y asignarle los permisos debidos. Además puede acceder a las A.A que desee para realizar consultas a la información que estas poseen. El administrador tiene permisos para crear y eliminar reportes.

Los usuarios con rol de analista son los especialistas, los cuales pueden acceder al A.A Económico y a sus respectivos LT para visualizar los reportes que sean de su interés.

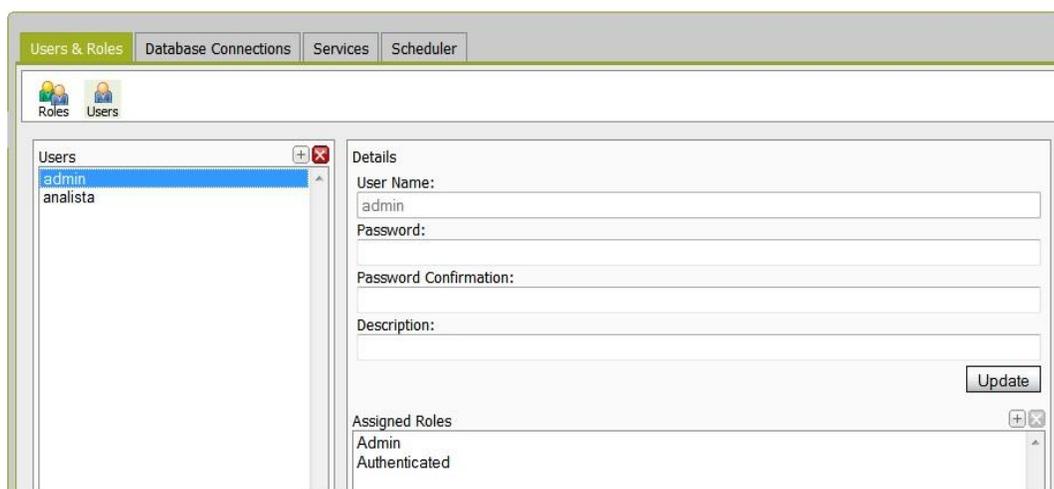


FIGURA 14 USUARIOS Y ROLES

2.8.1 Políticas de respaldo y recuperación

Para garantizar que no exista la pérdida de la información del mercado de datos económico, se realizan salvadas mensuales de la información que está en la BD y se verifica que la información almacenada es la correcta. Además después de cada carga se genera un backup de la BD y se guarda en varios lugares evitando así la pérdida de la información en caso de que se dañe la BD o sufra modificaciones involuntarias. En este proceso se involucran las tablas de hechos y dimensiones, ya que contienen toda la información del MD.

Conclusiones

Durante todo el capítulo se realizó el análisis y diseño del mercado de datos económico del Sistema de Información Estadístico Judicial para los Tribunales Populares identificándose los requisitos de información, requisitos funcionales y requisitos no funcionales. Se realizó el diagrama de casos de uso del sistema compuesto por los actores y los casos de usos de información y funcionales, lo que permitió representar gráficamente la interacción que existe entre los actores y los CU. Se diseñó el subsistema de almacenamiento, permitiendo la realización del modelo dimensional donde se encuentra la relación de hechos y dimensiones. Además se diseñó el subsistema de integración, el cual apoyará todo el proceso de ETL. Se analizó el subsistema de visualización obteniendo como resultado las áreas de análisis con las que contará la presente investigación, los libros de trabajo y las vistas de análisis.

CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBAS DEL MERCADO DE DATOS ECONÓMICO DEL SISTEMA DE INFORMACIÓN ESTADÍSTICO JUDICIAL PARA LOS TRIBUNALES POPULARES

Este capítulo hace referencia a la implementación de la solución. En este se abordarán temas relacionados con los subsistemas de almacenamiento, integración y visualización de la materia económica del Sistema de Información Estadístico Judicial para los Tribunales Populares, teniendo en cuenta los requisitos y necesidades del negocio. De igual forma, este capítulo aborda el tema relacionado con las pruebas, mediante la utilización de las listas de chequeo y los casos de pruebas realizados al mercado de datos; incluyendo la carta de aceptación del cliente.

3.1 Implementación del subsistema de almacenamiento

Implementación del modelo de datos

Los datos se almacenan en tablas ubicadas en esquemas en la BD, lo que ayuda a la distribución y manipulación de los mismos. En el mercado de datos económico se definieron cuatro esquemas para una mejor organización de las tablas de la BD:

- **dimensiones:** agrupa cuatro tablas de dimensiones compartidas del almacén de datos para los tribunales populares.
- **mart_economico:** contiene 20 tablas de hechos y 21 tablas de dimensiones que son propias del mercado de datos económico.
- **errores:** se encuentran las tablas que recogen la información cuando no aparece algún identificador.
- **metadatos:** se encuentran las tablas que recogen la información de los metadatos y de los datos que no cumplan con alguna de las reglas de negocio.

La tabla 4 muestra los esquemas y las tablas correspondientes a cada uno de ellos.

TABLA 4 ESQUEMAS Y TABLAS DEL MERCADO DE DATOS ECONÓMICO

Esquemas	Tablas
Dimensiones	dim_dpa_provincia

	dim_tribunal_provincial dim_tipo_sala dim_temporal_mes
mart_economico	closure_causales closure_expedientes dim_causales dim_expedientes dim_asistencia_arbitraje_m2 dim_tramitaciones_proceso_ejecutivo_m2 dim_diligencias_previas_m2 dim_ejecucion_sentencias_m2 dim_sujeto_demandante_m2 dim_banco dim_expedientes_ejecutivos dim_expedientes_ejecucion dim_monto dim_comportamiento_radicacion dim_control_ejecuciones dim_antiguedad_ejecucion dim_estado_ejecuciones dim_subordinada_mismo_organismo dim_entidad_demandante_m1 dim_entidades_demandantes_m1 dim_entidades_m2 dim_entidad dim_estado_ejecuciones_m3 hech_causales_m2 hech_asistencia_arbitraje_m2 hech_diligencias_previas_m2

	<p> hech_ejecucion_sentencias_m2 hech_resolucion_procesos_m2 hech_tramitaciones_procesos_ejecutivos_m2 hech_causales_extracontractuales_m1 hech_resolucion_procesos_ejecutivos_m1 hech_resolucion_procesos_ma_otros_m1 hech_resolucion_procesos_contractuales_m1 hech_causales_ordinarios_contractuales_m3 hech_monto_reclamado_m3 hech_comportamiento_radicacion_m3 hech_resolucion_procesos_ordinarios_m3 hech_resolucion_procesos_ejecutivos_m3 hech_control_ejecuciones_m3 hech_antiguedad_ejecuciones_pendientes_m3 hech_comportamiento_radicacion_procesos_m3 hech_estado_ejecuciones_m3 hech_medidas_cautelares_solicitadas </p>
errores	<p> hech_antiguedad_ejecuciones_pendientes_m3 hech_causales_ordinarios_contractuales_m3 hech_comportamiento_radicacion_m3 hech_comportamiento_radicacion_procesos_m3 hech_control_ejecuciones_m3 hech_estado_ejecuciones_m3 hech_medidas_cautelares_solicitadas hech_monto_reclamado_m3 hech_resolucion_procesos_ejecutivos_m3 hech_resolucion_procesos_ordinarios_m3 </p>
Metadatos	<p> md_carga_historica md_cdc </p>

	md_mercado md_registro_hist_fichero md_temporal md_transformacion
--	--

Estándares de codificación

Los estándares de codificación permiten el entendimiento de un proyecto, ya que estos describen los esquemas y tablas utilizadas. A continuación se muestran los estándares de codificación utilizados en el MD económico (Ver tabla 5).

TABLA 5 ESTÁNDARES DE CODIFICACIÓN DEL MERCADO DE DATOS ECONÓMICO

Estructura	Descripción	Ejemplo
Esquemas	El nombre del esquema donde se almacenan los datos del mercado económico estará compuesto por la palabra mart y el nombre del mercado separados por un guión bajo.	mart_economico
Identificadores	Todos los identificadores tendrán el nombre de la tabla a la que pertenecen y una cadena que demuestra que son identificadores.	dim_<nombre_tabla>_id
Medidas	Todas las medidas presentes en las tablas de hechos estarán definidas por un atributo	cant_<nombre_medida>

	“cant” separado por un guión bajo y después el nombre que se corresponde con las mismas.	
Tablas de dimensiones	Todas las tablas de dimensiones tendrán una cadena que demuestra que son dimensiones y el nombre que se corresponde con las mismas.	dim_<nombre_dimension>
Tablas de hechos	Todas las tablas de hechos tendrán una cadena que demuestra que son hechos y el nombre que se corresponde con el mismo.	hech_<nombre_hecho>

3.2 Implementación del subsistema de integración de los datos

El subsistema de integración es el encargado de realizar la extracción, limpieza, estandarización e integración de la información, preparándola para la carga al mercado de datos (22). Ralph Kimball definió 34 subsistemas que se clasifican en cuatro grupos: limpieza, extracción y conformación, entrega y gestión.

Los subsistemas utilizados para la implementación de la solución se describen a continuación (22):

- Perfilado de datos: consiste en la exploración de los datos para verificar su calidad y si cumple los estándares conforme a los requisitos.
- Sistema de extracción: permite la extracción de los datos desde la fuente origen a la fuente destino.

- Rastreo de eventos de errores: captura todos los errores que proporcionan información valiosa sobre la calidad de los datos y permiten mejorar los mismos.
- Dimensiones lentamente cambiantes (SCD, por sus siglas en inglés): implementa la lógica para crear atributos de variabilidad lenta a lo largo del tiempo.
- Llaves subrogadas: permite crear claves subrogadas independientes por cada tabla.
- Creador de cubos OLAP: permite alimentar de datos a esquemas OLAP desde los esquemas dimensionales relacionales.
- Programador de trabajos: permite gestionar las transformaciones mediante trabajos.
- Control de versiones: permite realizar el control de versiones de los procesos de ETL y sus metadatos asociados.
- Repositorio de metadatos: captura los metadatos de los procesos ETL, de los datos del negocio y de los aspectos técnicos.

3.2.1 Implementación de las dimensiones lentamente cambiante

Cuando los datos de una dimensión tienden a cambiar al transcurrir el tiempo, se está en presencia de una dimensión lentamente cambiante. (23) Kimball define tres tipos de SCD (tipo 1, 2 y 3), aunque en la actualidad se han creado nuevos tipos (tipo 0, 4 y 6).

Tipo 0: no tiene en cuenta la gestión de los cambios históricos. Algunos datos serán sobrescritos y otros no, sin ninguna planificación.

Tipo 1: no guarda los cambios históricos y no necesita que se añadan nuevos campos a la tabla. Es utilizado en caso de que no sea importante mantener la información histórica.

Tipo 2: esta estrategia es utilizada cuando se necesita almacenar los cambios históricos, para ello se añadirá una nueva fila y se deberán complementar los datos del historial de cambios. Esta técnica permite guardar ilimitada información.

Tipo 3: en esta estrategia, en lugar de emitir una nueva fila cuando se produce algún cambio, se crea una nueva columna con los valores que se desea mantener. Permite guardar una limitada información de cambios.

Tipo 4: se utiliza en combinación con algún otro tipo y su función es almacenar en una nueva tabla la información de los cambios históricos realizados en una dimensión.

Tipo 6: combina las SCD de tipo 1, 2 y 3.

En el desarrollo de la aplicación se utilizaron las SCD de tipo 2 y 3. La dimensión tipo_sala se implementó de tipo 2 permitiendo conocer las fechas y las versiones de los cambios ocurridos con el transcurso del tiempo. Por su parte, con el objetivo de establecer comparaciones entre los datos antiguos y actuales se implementaron las dimensiones tribunal_provincial y dpa_provincia de tipo 3.

3.2.2 Implementación de los flujos de transformación

La transformación es el elemento básico dentro del proceso de ETL, está compuesta por pasos enlazados entre sí a través de saltos. Los pasos constituyen el elemento más pequeño dentro de las transformaciones y los saltos hacen que la información fluya entre los diferentes pasos.

Como se puede apreciar en la figura 15, los primeros pasos de la transformación son para obtener la nueva información de la entrada dbf perteneciente al modelo “Monto de lo reclamado de las demandas presentadas”. Luego se validan un grupo de restricciones del negocio y se normalizan los datos, además se realiza una búsqueda para obtener el id de las dimensiones con las que se relaciona el hecho. El próximo paso es el encargado de verificar que se hayan encontrado todos los id, es decir que no hayan campos de id vacíos. Si algunos de los id no se encuentran, la información es enviada hacia un archivo Excel para que el especialista realice las modificaciones necesarias. En caso de no existir errores se insertan los datos en la BD tribunal_dwh y se obtiene toda la información de la transformación para guardarla en las tablas de metadatos correspondiente a ella.

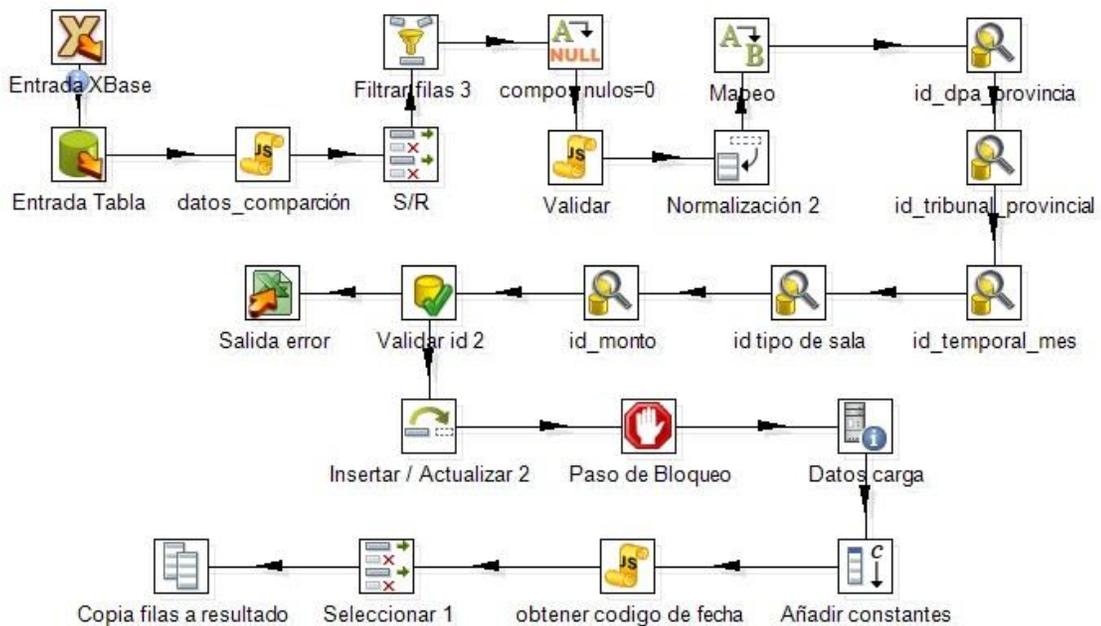


FIGURA 15 TRANSFORMACIÓN DEL HECHO MONTO RECLAMADO M3

3.2.3 Implementación de los trabajos

Un trabajo es un conjunto de tareas que tiene como objetivo realizar una acción determinada. Estos permiten ejecutar varias transformaciones o trabajos previamente diseñados para lo que organiza una secuencia de ejecución de estos.

En la presente investigación se realizaron dos trabajos principales. Uno para la carga histórica que está compuesta por tres trabajos y una transformación. El otro para la carga incremental que está compuesta por dos trabajos, donde se ejecutan primeramente las transformaciones correspondientes a la carga de las dimensiones particulares del mercado de datos, y posteriormente las cargas de los hechos identificados. A continuación se muestra el trabajo realizado para la carga incremental (Ver figura 16).



FIGURA 16 TRABAJO PRINCIPAL DE LA CARGA INCREMENTAL

3.2.4 Metadatos

Los metadatos describen y apoyan a diferentes datos y estructuras, tales como reglas de negocio y los procesos que manipulan los datos. Estos se agrupan en tres categorías principales (23):

Metadatos de negocio: permite obtener la información sobre las descripciones del negocio. Incluyen descripciones de datos que no están relacionadas con la implementación del software.

Metadatos técnicos: es la representación de los aspectos técnicos de los datos. Describen el funcionamiento de un sistema o el modo en que se relacionan sus componentes.

Metadatos de proceso: representan las estadísticas sobre los resultados de la ejecución del propio proceso de ETL.

En la presente investigación se utilizaron los metadatos de proceso, los que permiten obtener la información sobre los procesos de transformación tales como la cantidad de filas cargadas correctamente y las filas rechazadas. Además se utilizaron los metadatos técnicos que permiten conocer aspectos sobre la gestión de la carga histórica, como la fecha de la última carga y cuáles fueron los modelos que se cargaron.

3.3 Implementación del subsistema de visualización

Luego de realizar el diseño del subsistema de visualización se procede a su implementación, la cual depende fundamentalmente de los cubos OLAP, las vistas de análisis y los reportes candidatos.

3.3.1 Implementación de los cubos OLAP

Los cubos poseen un mecanismo para la búsqueda de los datos con un tiempo de respuesta uniforme independientemente de la complejidad de la búsqueda y de la cantidad de datos en el cubo. En estos se representan los hechos y las dimensiones asociadas. Los hechos definidos en un MD indican la cantidad de cubos OLAP que deben ser creados para la visualización de la información.

El presente trabajo de diploma cuenta con 20 cubos que responden a los hechos identificados en el mercado de datos económico y 25 dimensiones. En cada uno de los cubos se encuentran las dimensiones con las que se relaciona y las medidas que responden a este. La implementación fue realizada en la herramienta Pentaho Schema Workbench en su versión 3.2.1, la cual genera un fichero MDX con los cubos y las dimensiones implementadas. En la figura 17 se muestra el cubo del hecho **monto reclamado** con las dimensiones asociadas y la medida que responde a este hecho.



FIGURA 17 CUBO PERTENECIENTE AL HECHO MONTO RECLAMADO M3

3.4 Implementación de los reportes candidatos

Los reportes candidatos muestran la información que se cargó a la BD a través del proceso de ETL. Mediante estos reportes se podrá obtener la información necesaria para realizar los análisis deseados, lo que facilita la toma de decisiones. En la presente investigación se implementaron 29 reportes, los cuales muestran información de los litigios de la materia económica del TSP. Para esta implementación se utilizó

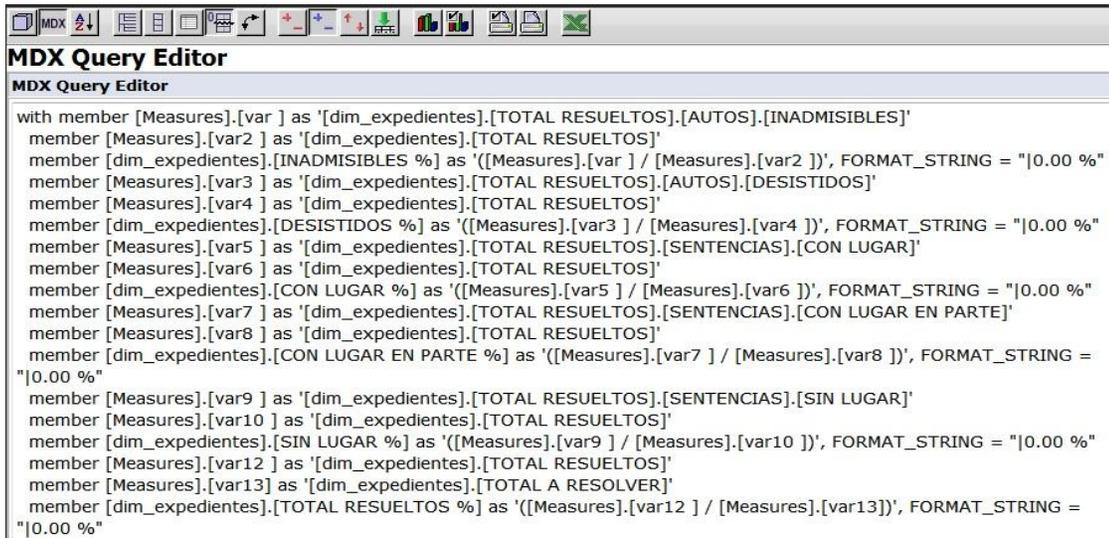
la herramienta Pentaho BI Server. A continuación se muestra el reporte **Monto reclamado** (Ver figura 18). En la imagen se pueden observar los indicadores que componen dicho reporte y la medida que responde a este. En el indicador provincia se encuentran todas las provincias del país, en sala se tiene el nombre de la sala a la que responden los datos del reporte (en este caso la sala económica) y Monto reclamado que contiene la radicación de procesos por cada una de las provincias con el monto que representa.

		Medidas			
		Monto reclamado			
		Tiempo			
		+ 2013			
		Monto reclamado			
Provincia1	Sala	● + RADICACIÓN PROCESOS	● + Monto de lo reclamado en CUP	● + Monto de lo reclamado en CUC	● + Monto de lo reclamado en M L C
Pinar del Río	Sala Econ.	63	3.661.990,15	5.862,6	0
Artemisa	Sala C.A.L.E.	30	2.881.771,85	220.428,784	184.018,21
La Habana	Sala Econ.	332	338.731.358,7	12.465.814,77	5.718.903,11
Mayabeque	Sala Econ.	19	13.689.520,32	9.634,32	0
Matanzas	Sala Econ.	76	2.151.256,53	1.551.431,26	0
Villa Clara	Sala Econ.	67	16.825.456,73	2.517.938,92	0
Cienfuegos	Sala Econ.	61	5.311.535,78	3.596.525,53	0
Sancti Spiritus	Sala Econ.	69	1.800.184,14	142.704,69	0
Ciego de Ávila	Sala Econ.	133	8.360.075,75	1.614.351,05	54.321,92
Camagüey	Sala Econ.	110	14.237.096,09	6.959.063,99	0
Las Tunas	Sala Econ.	43	2.684.230,07	15.822,87	0
Holguín	Sala Econ.	39	6.048.858,13	447.680,45	0
Granma	Sala Econ.	93	52.240.055,32	178.777,11	1.382,08
Santiago de Cuba	Sala Econ.	192	10.642.752,63	1.078.434,89	0
Guantánamo	Sala Econ.	24	713.643,27	6.498,37	6.109,17

FIGURA 18 REPORTE MONTO RECLAMADO

3.4.1 Consultas MDX

Las consultas MDX se utilizan para generar reportes basados en datos históricos, ayudando así a la toma de decisiones. (24) Además permiten cambiar la estructura del cubo, dándole la vista deseada. Las consultas mencionadas pueden ser simples o complejas, dependiendo de la dificultad de cada operación. A continuación se muestran algunas consultas MDX utilizadas en la presente investigación, donde se calculó el por ciento de los procesos desistidos, inadmisibles, con lugar, con lugar en parte, sin lugar y del total de procesos resueltos (Ver figura 19).



```
with member [Measures].[var ] as '[dim_expedientes].[TOTAL RESUELTOS].[AUTOS].[INADMISIBLES]'
member [Measures].[var2 ] as '[dim_expedientes].[TOTAL RESUELTOS]'
member [dim_expedientes].[INADMISIBLES %] as '([Measures].[var ] / [Measures].[var2 ])', FORMAT_STRING = "|0.00 %"
member [Measures].[var3 ] as '[dim_expedientes].[TOTAL RESUELTOS].[AUTOS].[DESISTIDOS]'
member [Measures].[var4 ] as '[dim_expedientes].[TOTAL RESUELTOS]'
member [dim_expedientes].[DESISTIDOS %] as '([Measures].[var3 ] / [Measures].[var4 ])', FORMAT_STRING = "|0.00 %"
member [Measures].[var5 ] as '[dim_expedientes].[TOTAL RESUELTOS].[SENTENCIAS].[CON LUGAR]'
member [Measures].[var6 ] as '[dim_expedientes].[TOTAL RESUELTOS]'
member [dim_expedientes].[CON LUGAR %] as '([Measures].[var5 ] / [Measures].[var6 ])', FORMAT_STRING = "|0.00 %"
member [Measures].[var7 ] as '[dim_expedientes].[TOTAL RESUELTOS].[SENTENCIAS].[CON LUGAR EN PARTE]'
member [Measures].[var8 ] as '[dim_expedientes].[TOTAL RESUELTOS]'
member [dim_expedientes].[CON LUGAR EN PARTE %] as '([Measures].[var7 ] / [Measures].[var8 ])', FORMAT_STRING =
"|0.00 %"
member [Measures].[var9 ] as '[dim_expedientes].[TOTAL RESUELTOS].[SENTENCIAS].[SIN LUGAR]'
member [Measures].[var10 ] as '[dim_expedientes].[TOTAL RESUELTOS]'
member [dim_expedientes].[SIN LUGAR %] as '([Measures].[var9 ] / [Measures].[var10 ])', FORMAT_STRING = "|0.00 %"
member [Measures].[var12 ] as '[dim_expedientes].[TOTAL RESUELTOS]'
member [Measures].[var13 ] as '[dim_expedientes].[TOTAL A RESOLVER]'
member [dim_expedientes].[TOTAL RESUELTOS %] as '([Measures].[var12 ] / [Measures].[var13 ])', FORMAT_STRING =
"|0.00 %"
```

FIGURA 19 CONSULTAS MDX

3.5 Pruebas de software al mercado de datos económico

Las pruebas de software constituyen una fase importante en el desarrollo de cualquier producto. Tienen el objetivo de identificar dificultades en temas de implementación, o de manera general, en la usabilidad del sistema y demuestran hasta qué punto funciona el software. Proporcionándole así alta calidad al producto final.

Existen pruebas que se utilizan para el proceso de calidad del software. A continuación, se describen las pruebas aplicadas al MD económico:

Pruebas unitarias: permiten probar el correcto funcionamiento de un componente o subsistema específico y son desarrolladas por los propios desarrolladores durante la implementación. (10)

Pruebas de integración: permiten verificar la correcta integración de los componentes y subsistemas que conforman la solución. Ponen a prueba la vista arquitectónica del sistema definida en una infraestructura de desarrollo. Estas pruebas son ejecutadas por los arquitectos de software. (10)

Pruebas de sistema: permiten validar el cumplimiento de los requisitos de información y funcionales definidos por los clientes. Son las pruebas más cercanas a la realidad del cliente debido a que los probadores utilizan el sistema de la misma manera que será usado por los clientes. (10)

Pruebas de aceptación: Estas pruebas son realizadas por el cliente para verificar que se cumple con los requisitos planteados por él mismo y validar su conformidad con el producto. (10)

En el presente trabajo de diploma, las pruebas unitarias se realizaron durante el desarrollo de cada uno de los subsistemas, donde se fueron identificando errores a los que se les daba solución. Las pruebas de integración fueron realizadas después de implementado el subsistema de visualización, donde se verificó que los datos cargados al mercado de datos económico corresponden con la información que se encuentra en la fuente. Por su parte, las pruebas de sistema se llevaron a cabo con el objetivo de encontrar incoherencias entre la descripción de los casos de prueba y el sistema. Por último, se realizaron las pruebas de aceptación, donde se le demostró al cliente que la solución cumple con los requisitos iniciales.

3.5.1 Herramientas de pruebas

Casos de prueba

Los casos de prueba (CP) son un conjunto de valores de entrada, precondiciones de ejecución y resultados esperados, desarrollados con un objetivo particular o condición de prueba, tal como ejercitar un camino de un programa particular o para verificar que se cumple un requisito específico. (25) El diseño de un buen caso de prueba es el eslabón fundamental del proceso de prueba, pues mediante ellos se encontrarán la mayor cantidad de defectos posibles.

En el presente trabajo de diploma se utilizaron seis CP, los cuales contribuyeron a la mejora de la solución. A continuación se muestra el CP correspondiente a los montos reclamados (Ver figura 20). El resto de los CP podrán ser consultados en el expediente de proyecto de la materia económica.

Escenario	Descripción	Variables de Entrada	Variables de Salida	Respuesta del sistema	Flujo central
EC 1: Monto reclamado.	Muestra la información de los montos reclamados de los procesos ordinarios.	Tiempo	Cantidad de montos reclamados de los procesos ordinarios	El sistema muestra todas las variables disponibles para los análisis, ubicados en las filas y las columnas que pueden ser visualizadas para cada reporte.	Se abre la aplicación.
Tribunal provincial		Se autentifica.			
Sala		Se entra al sistema.			
Provincia		Se despliega hacia la derecha el componente ubicado en el lateral izquierdo que contiene el navegador.			
Monto reclamado		Se selecciona el área de análisis general Sistema de Información Estadístico Judicial para los Tribunales			
					Se selecciona el área de análisis A.A. Económico
					Se selecciona el área de análisis A.A. DPA (2011-actualidad)
					Se selecciona el área de análisis A.A. Metodología 3
					Se selecciona el libro de trabajo de LT 3501 Modelo radicación de procesos y monto reclamado.
					En la parte inferior izquierda

FIGURA 20 MOSTRAR MONTOS RECLAMADOS

Listas de chequeo

Las listas de chequeo tienen como objetivo obtener los puntos satisfactorios e insatisfactorios de los artefactos del expediente de proyecto correspondiente a la implementación de los procesos de integración. Estas están compuestas por un conjunto de preguntas previamente definidas por el centro en forma de cuestionario.

Las listas de chequeo fueron aplicadas al expediente de proyecto, lo que permitió verificar la calidad, estructura y semántica de los documentos que lo componen.

3.5.2 Resultados de las pruebas aplicadas

Las pruebas se realizaron en conjunto con la especialista del TSP y la especialista del departamento AD y después de ser aplicadas se obtuvo los siguientes resultados:

Durante la aplicación de las pruebas unitarias se encontraron 15 no conformidades (NC), de ellas 10 de complejidad baja y 5 de complejidad media como se muestra en la figura 21.



FIGURA 21 RESULTADO DE LAS PRUEBAS UNITARIAS

Al realizar las pruebas de integración no se detectaron NC, lo que demostró que los datos cargados al mercado económico coinciden con la información de la fuente.

Las pruebas de sistema arrojaron 15 NC, de ellas ocho de complejidad baja, seis media y una alta. Este resultado queda reflejado en la figura 22.



FIGURA 22 RESULTADO DE LAS PRUEBAS AL SISTEMA

Después de realizar las pruebas de aceptación con el cliente, fueron identificadas cinco NC, de ellas dos de complejidad baja, dos media y una alta. En la figura 23 se muestra el resultado de la realización de dicha prueba.



FIGURA 23 RESULTADO DE LAS PRUEBAS DE ACEPTACIÓN

La figura 24 refleja los resultados obtenidos al concluir la fase de pruebas donde fueron detectadas 35 NC, de ellas 19 de complejidad baja, 13 media y tres alta. Las de complejidad baja respondían a errores ortográficos encontrados en las descripciones de los libros de trabajo. Las NC encontradas de complejidad media correspondían a la no coincidencia de lo descrito en los CP con la aplicación y las de complejidad alta estaban asociadas al no cumplimiento de algunos de los requisitos funcionales.



FIGURA 24 RESULTADO GENERAL DE LAS PRUEBAS

Después de darle tratamiento a las NC encontradas estas quedaron solucionadas. Luego se realizó una segunda iteración de las pruebas de aceptación, donde el cliente quedó satisfecho con la solución dada, por lo que se obtuvo la carta de aceptación.

Conclusiones

Durante el capítulo se realizó la descripción de la implementación del subsistema de almacenamiento, el subsistema de integración y del subsistema de visualización. Mediante la implementación del subsistema de almacenamiento la información quedó organizada en tres esquemas, que contribuyen a la calidad de la implementación del subsistema de integración. Para lograr la carga de los hechos y las dimensiones se realizaron 42 transformaciones y dos trabajos principales los cuales posibilitaron la ejecución de dichas transformaciones. A través de la implementación del subsistema de visualización quedó definida el área de análisis general, las áreas de análisis y sus libros de trabajo, lo que facilita la visualización de la información y contribuye a la toma de decisiones. Se realizaron las pruebas unitarias, de integración, de sistema y de aceptación al mercado de datos económico, obteniendo varias NC las cuales fueron resueltas satisfactoriamente, lo que ayudó a mejorar la calidad de la solución.

CONCLUSIONES GENERALES

- Para desarrollar el mercado de datos económico se escogió la metodología y las herramientas que contribuyeron a la estabilidad del producto final.
- Se realizó el análisis y diseño de la materia económica, identificándose 20 tablas de hechos y 25 tablas de dimensiones.
- Se diseñaron e implementaron los subsistemas de almacenamiento, integración y visualización, obteniéndose el Mercado de Datos Económico del Sistema de Información Estadístico Judicial para los Tribunales Populares.
- Se realizaron pruebas unitarias, de integración, de sistema y de aceptación al mercado de datos económico, asegurando la calidad del producto final.

RECOMENDACIONES

Al concluir el presente trabajo se recomienda:

- Realizar mejoras en la capa de visualización para lograr mayor facilidad de uso por parte de los usuarios.
- Integrar el MD económico al Sistema de Información de Gestión Estadística (SIGE) para realizar la carga incremental de los datos y general reportes de forma dinámica.

REFERENCIAS BIBLIOGRÁFICAS

1. Scribd. [En línea] [Citado el: 5 de junio de 2013] Clasificación de los sistemas de información
<http://es.scribd.com/doc/6839764/Clasificacion-de-Los-Sistemas-de-Informacion>
2. Inmon, Willian H. Building the Data Warehouse, Fourth Edition. IN 46256
3. DataPrix. [En línea] [Citado el: 5 de diciembre de 2012] ¿Qué es un Almacén de Datos?
<http://webcache.googleusercontent.com/search?q=cache:2v9EU25lotwJ:www.dataprix.com/que-es-un-datawarehouse+bill+inmon+%2B+almacen+de+datos&cd=3&hl=es&ct=clnk&gl=cu>
4. Amparo Griñan García y Yurena Ávila Domínguez. [En línea] [Citado el: 5 de diciembre de 2012] Almacén de datos (Data Warehouse)
https://docs.google.com/viewer?a=v&q=cache:VXgGhSS6RfEJ:www.ucasal.net/templates/unid-academicas/ingenieria/apps/5-p56-rivadera-formateado.pdf+articulo+de+kimball&hl=es&gl=cu&pid=bl&srcid=ADGEEShD9OulRMdeW-FpT_n2ZFPaOcETeUdwrUJXJePvpnYi15WshyhPUOxqqWMzMAM_zu6t3dVJT-JOkhQeVmJSVHyf1glr9LufG8_IEL5brmBeSgN70W7u24pMBmph7BJDRn_6eK0z&sig=AHIEtbRGr1gfihY0C1S9yf82A7zoMLi-oQ
5. Sinnexus. [En línea] [Citado el: 5 de diciembre de 2012.]
http://www.sinnexus.com/business_intelligence/datamart.aspx.
6. Casales Cabrera, María Evelia. Data Warehouse (Almacenes de Datos). 2009.
7. Curto, Josep. Information Management. Diseño de un data warehouse: estrella y copo de nieve. [En línea] 19 de noviembre de 2007. [Citado el: 29 de mayo de 2013.]
<http://informationmanagement.wordpress.com/2007/11/19/disenio-de-un-data-warehouse-estrella-y-copo-de-nieve/>.
8. Cox Alvarado, Alexander. [Citado el: 2 de febrero de 2012.] Propuesta de un álgebra para modelar bases de datos multidimensionales. Cartago, Instituto Tecnológico de Costa Rica. : s.n., 2000.

9. kerwa. [En línea] [Citado el: 25 de Noviembre de 2012.] <http://www.kerwa.ucr.ac.cr/bitstream/handle/10669/619/TM%2024-3%20art%206.pdf?sequence=1>.
10. Hernández, Yanisbel González. PROPUESTA DE METODOLOGIA DE DASARROLLO DE ALMACENES DE DATOS. La Habana, Cuba : s.n., 2012 .
11. Visual Paradigm. [En línea] [Citado el: 10 de diciembre de 2012.] <http://www.visual-paradigm.com/product/vpuml/>.
12. ALVAREZ. S. desarrollo web. Sistemas gestores de bases de datos. [En línea] [Citado el: 10 de diciembre de 2012.] <http://www.desarrolloweb.com/articulos/sistemas-gestores-bases-datos.html>.
13. Sobre PostgreSQL. Copyright 2009-2012 PosgreSQL-es. [En línea] [Citado el: 15 de diciembre de 2012.] http://www.postgresql.org.es/sobre_postgresql.
14. The pgAdmin Development Team. [En línea] [Citado el: 12 de diciembre de 2012.] www.pgadmin.org/.
15. Pentaho Data Integration. [En línea] [Citado el: 12 de diciembre de 2012.] <http://www.pentaho.com/explore/pentaho-data-integration/> Pentaho Data Integration.
16. Mondrian. [En línea] [Citado el: 10 de enero de 2013.] <http://mondrian.pentaho.com/documentation/workbench.php>.
17. Sitio web [En línea] [Citado el: 14 de enero de 2013.] <http://www.summan.com/pentaho/pentaho-bi-platform-server> Certificado ISO 9001:2008. 20 años en el medio, prestando Soluciones y Servicios para Gestión Documental e Infraestructura Informática. Cra 43C N0. 10 - 23 Medellín - Colombia.
18. El servidor OLAP Mondrian. [En línea] [Citado el: 14 de enero de 2013.] <http://pentaho.almacendatos.com/mondrian.html>.
19. The Apache Software Foundation. [En línea] [Citado el: 18 de enero de 2013.] <http://tomcat.apache.org/>.

20. Ralph R. Young. The requirements engineering handbook. 2004 ARTECH HOUSE. INC 685
21. Pressman, Roger S. *Ingeniería de Software.Un enfoque práctico*. 6ta Edición. s.l. : McGraw-Hill Companies, 2007. ISBN: 8448132149.
22. Kimball, Ralph y Ross, Margy. The Data Warehouse Toolkit. s.l.: Wiley Computer Publishing. ISBN: 0-471-20024-7.
23. Kimball, Ralph y Ross, Margy. The Data Warehouse ETL Toolkit. sl.: Wiley Computer Publishing. ISBN: 0-431-20025-8.
24. Consultas MDX básicas (MDX). [En línea] [Citado el: 12 de mayo de 2013.] <http://msdn.microsoft.com/es-es/library/ms144785.aspx>.
25. EEE-Std. IEEE Standard Glossary of Software Engineering Terminology. 2002. www.ces.com.uy/documentos/AutomCACIC07.pdf

BIBLIOGRAFÍA

1. Ruiz, Oscar Marcos Amelunge. Análisis y Diseño de un Data Mart para la gestión de reportes de RRHH de la Empresa de Agua SA. Santa Cruz, Bolivia : s.n., 2010.
2. Casales Cabrera, María Evelia. Data Warehouse (Almacenes de Datos). 2009.
3. Cox Alvarado, Alexander. Propuesta de un álgebra para modelar bases de datos multidimensionales. Cartago, Instituto Tecnológico de Costa Rica. : s.n., 2000.
4. ALVAREZ. S. desarrollo web. Sistemas gestores de bases de datos. [En línea] <http://www.desarrolloweb.com/articulos/sistemas-gestores-bases-datos.html>.
5. Mondrian. [En línea] <http://mondrian.pentaho.com/documentation/workbench.php>.
6. The Apache Software Foundation. [En línea] <http://tomcat.apache.org/>.
7. Sinnexus. [En línea] http://www.sinnexus.com/business_intelligence/datamart.aspx.
8. Hernández, Yanisbel González. PROPUESTA DE METODOLOGIA DE DASARROLLO DE ALMACENES DE DATOS. La Habana, Cuba : s.n., 2012.
9. Rivadera, Gustavo R. La metodología de Kimball para el diseño de almacenes de datos (Data warehouses). 2010.
10. Inmon, William, Strauss, Derek y Neushloss, Genia. The Architecture for the Next Generation of Data Warehousing.
11. Castelán García, Leopoldo y Ocharan Hernández , Jorge Octavio. Diseño de un Almacen de datos basado en Data Warehaouse Engineering Process (DWEP) y HEFESTO. Veracruz, Mexico : s.n.
12. Definición de. Definición de Sistema de Información. [En línea] URL: <http://webcache.googleusercontent.com/search?q=cache:GMz8iOgyRiEJ:definicion.de/sistema-de-informacion/+sistema+de+informacion&cd=5&hl=es&ct=clnk&gl=cu>.
13. Ministerio de Educación superior. Sistemas de información y estadísticas universitarias . [En línea] URL: http://www.mes.edu.cu/index.php?option=com_content&task=view&id=24.
14. Sistema de información sobre categorías farmacológicas . Pérez, Daisy . n.2, Ciudad de la Habana Revista Cubana de Farmacia, 2001, Vol. v.35.

15. Pressman, Roger S. *Ingeniería de Software. Un enfoque práctico*. 6ta Edición. s.l. : McGraw-Hill Companies, 2007. ISBN: 8448132149.
16. Durán, Juan A. de Mula y González, Antonio Molina. EL SISTEMA DE INFORMACIÓN ESTADÍSTICA DE ANDALUCÍA. Andalucía : s.n.
17. IEEE-Std. IEEE Standard Glossary of Software Engineering Terminology. 2002. www.ces.com.uy/documentos/AutomCACIC07.pdf.
18. El servidor OLAP Mondrian. [En línea] <http://pentaho.almacen-datos.com/mondrian.html>.
19. LUJAN MORA, Sergio. Data Warehouse Design with UML Tesis (Doctorado). España, Universidad de Alicante, 2005.
20. KIMBALL, Ralph y ROSS, Margy. The Data Warehouse Lifecycle Toolkit. 2a. ed. Canadá: Wiley Publishing, Inc, 2002.
21. Kimball, Ralph y Ross, Margy. The Data Warehouse ETL Toolkit. sl.: Wiley Computer Publishing. ISBN: 0-431-20025-8.
22. Scribd. [En línea] Clasificación de los sistemas de información <http://es.scribd.com/doc/6839764/Clasificacion-de-Los-Sistemas-de-Informacion>
23. Pentaho Mondrian Documentation [En línea] <http://mondrian.pentaho.com/documentation/mdx.php>
24. Adamson, Christopher Mastering Data Warehouse Aggregates: Solutions for Star Schema Performance. Canadá: Wiley Publishing, 2006
25. Ecured [En línea] http://www.ecured.cu/index.php/Patrones_de_Casos_de_Uso
26. Kimball, Ralph y Ross, Margy. The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence. Wiley Publishing, Inc., Indianapolis, 2010
27. Wrembel, Robert y Koncilia, Christian. Data Warehouses and OLAP: Concepts, Architectures and Solutions. Estados Unidos, 2006
28. Sobre PostgreSQL. Copyright 2009-2012 PostgreSQL-es. [En línea] http://www.postgresql.org.es/sobre_postgresql.

GLOSARIO DE TÉRMINOS

A continuación se presentan los términos que podrían resultar de difícil comprensión, nuevos al lector o de diversos significados dependiendo del contexto que se analice. Esta sección tiene como objetivo facilitar la comprensión del contenido expuesto en el documento.

BI: proceso que permite realizar la capa de visualización de datos de un AD.

Dimensión: característica de un hecho que permite su análisis posterior en el proceso de toma de decisiones y brinda una perspectiva adicional a un hecho dado.

Hecho: operación que se realiza en el negocio la cual está estrechamente relacionada con el tiempo y es objeto de análisis para la toma de decisiones.

Indicador: generalmente es un valor numérico y representan lo que se desea analizar concretamente

JDBC: protocolo de conexión de Java a base de datos (del inglés Java Data Base Connectivity).

Lista de chequeo: instrumento de medición y evaluación que consiste básicamente en un formulario de preguntas referentes al atributo de calidad que se está probando y de las características del documento en el caso de la documentación.

No conformidad: defecto, error o sugerencia que se le hace al equipo de desarrollo una vez encontrada alguna dificultad en lo que se está evaluando.

TSP: Tribunal Supremo Popular.