



Facultad Regional Mártires de Artemisa

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

Título:

**Mercado de datos que contribuya a la toma de decisiones en la dirección de justicia
de la administración provincial de Artemisa.**



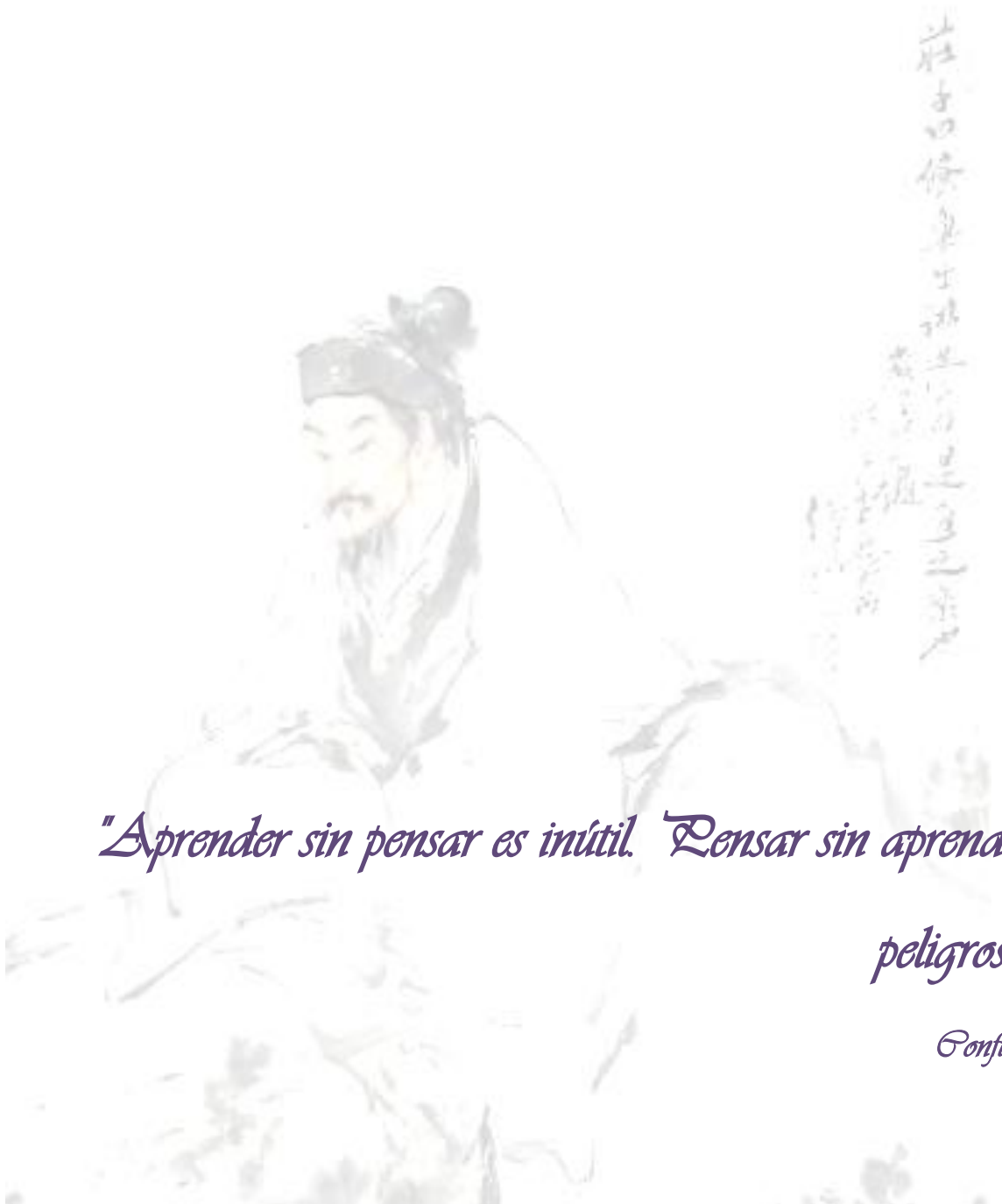
Autora: Yenisley Gonzáles Saez

Tutora: Ing. Mayrín Ramos Maestre

Co-tutor: Ms.C Raúl Edis Riverón Nuñez

Artemisa, junio del 2012

“Año 54 de la Revolución”



*"Aprender sin pensar es inútil. Pensar sin aprender,
peligroso."*

Confucio.

Declaración de autoría

Declaro que soy el único autor del trabajo “*mercado de datos que contribuya a la toma de decisiones en la dirección de justicia de la administración provincial de Artemisa*” y autorizo a la facultad regional de Artemisa de la Universidad de las Ciencias Informáticas; para que hagan el uso que estimen pertinente con este trabajo.

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

Raúl Edis Riverón Nuñez

Firma del co-tutor

Mayrín Ramos Maestre

Firma del tutor

Yenisley González Saez

Firma del autor

Agradecimientos

*La gratitud es la memoria del corazón, y son muchas las memorias a las cuales les debo agradecer
por ser hoy lo que soy.*

*En primer lugar quisiera agradecer a mis padres y a mi hermano, que me han apoyado en todo
momento durante estos cinco largos años.*

*A mi novio por su amor, comprensión y constante paciencia. A mi familia y a la familia de mi
novio por su cariño, apoyo y constante preocupación.*

*A mi tutora, a mi co-tutor, a mis amigos, a mis profesores y a todos los que facilitaron mi estancia
en la universidad: los que me enseñaron, los que criticaron, los que dejaron huellas a seguir.*

*Gracias a todos aquellos que me abrieron las puertas y me dejaron entrar a este maravilloso
mundo.*

Dedicatoria

Dedico este proyecto de tesis a mis padres Zenaida y Armando, a mi hermano Ruvisey, a mi novio Virgilio, porque han estado conmigo apoyándome, cuidándome y dándome fortaleza para continuar. Es por ellos que soy lo que soy ahora.

Los amo con mi vida.

RESUMEN

Los Almacenes de Datos (AD) tienen gran importancia ya que permiten el análisis de grandes masas de información, mientras que los Mercados de Datos (MD) son en esencia un AD, pero de forma más reducida a un departamento específico.

Con la necesidad de construir un MD para la dirección justicia de la Administración Provincial de Artemisa (APA) surge la presente investigación. En esta dirección existen varias insuficiencias, las cuales provocan la poca seguridad de la información, además dificulta el análisis de los datos para la toma de decisiones.

Este trabajo plasmará a su vez un estudio detallado de las principales herramientas, tecnologías y metodologías encargadas del desarrollo de soluciones referentes a los sistemas de almacenamiento y gestión de datos. Las pruebas realizadas a este mercado se realizaron a través de los casos de prueba basados en casos de uso los cuales arrojaron resultados satisfactorios.

Como resultado se obtuvo un mercado de datos para el apoyo al proceso de toma de decisiones de la dirección justicia. Todo lo expuesto con anterioridad será el principal contenido a tratar en el siguiente material.

Palabras claves: mercado de datos, toma de decisiones, dirección justicia y administración provincial de Artemisa.

Índice General

Introducción.....	1
1 Capítulo 1. Fundamentación teórica	9
1.1 Sistemas de almacenamiento de la información	9
1.1.1 Almacenes de datos.....	10
1.1.1.1 Características de los almacenes de datos.....	11
1.1.1.2 Ventajas y desventajas del almacén de datos.....	12
1.1.2 Mercados de datos.....	12
1.1.2.1 Características de los mercados de datos	13
1.1.2.2 Diferencias entre mercado de datos y almacenes de datos	13
1.2 Arquitectura del diseño	14
1.3 Modelo de datos del almacén de datos	15
1.3.1 Esquema estrella	15
1.3.2 Esquema copo de nieve.....	15
1.3.3 Tabla de hechos.....	16
1.3.4 Jerarquías	16
1.3.5 Niveles	16
1.3.6 Granularidad	17
1.3.7 Modelo dimensional	17
1.4 Procesamiento analítico en línea.....	19
1.5 Tecnología y tendencias actuales.....	20
1.6 Metodologías para el desarrollo de un mercado de datos	22
1.7 Herramientas para el desarrollo de un mercado de datos	23
1.7.1 Visual Paradigm	23
1.7.2 Power Architect	23
1.7.2.1 Versiones que presenta Power Architect.....	24
1.7.3 Pentaho data integration	25
1.7.4 Pentaho BI server 3.8.0.....	26
1.7.5 Mondrian schema workbench 3.2.1	26
Conclusiones	27
2 Capítulo 2: Análisis y diseño del mercado de datos de la dirección de justicia de la administración provincial de Artemisa	28
2.1 Definición del negocio	28
2.2 Temas de análisis.....	29
2.3 Reglas del negocio	29
2.4 Especificación de los Requisitos de Software.....	30
2.4.1 Requisitos informativos	30

2.4.2	Requisitos funcionales.	35
2.4.3	Requisitos no funcionales.....	36
2.5	Casos de uso del sistema	38
2.5.1	Diagrama de Casos de Uso	39
	Especificación de los Casos de Uso	40
2.5.2	Casos de Uso de Información	40
2.5.3	Casos de Uso Funcionales.....	41
2.6	Diseño de la estructura física de datos	42
2.7	Matriz bus	42
2.8	Modelo de datos dimensional	46
2.8.1	Tabla descriptiva de la Dimensión	46
2.8.2	Medidas presente en la tabla de hecho:	48
2.9	Política de respaldo y recuperación	49
	Conclusiones	50
3	<i>Capítulo 3: Implementación del mercado de datos de la dirección de justicia de la Administración Provincial de Artemisa.....</i>	51
3.1	Implementación de los subsistemas de integración	51
3.2	Implementación de los trabajos.....	54
3.3	Implementación de los subsistemas de visualización	56
3.3.1	Diseño de cubos OLAP	56
3.3.2	Arquitectura de información.....	57
3.3.3	Vistas de análisis.....	58
	Conclusión	60
4	<i>Capítulo 4: Validación del mercado de datos de la dirección de justicia de la administración provincial de Artemisa</i>	61
4.1	Calidad de software	61
4.1.1	Elementos de calidad de software.....	62
4.2	Pruebas de software	62
4.3	Validación y prueba	63
	Conclusiones	65
	<i>Conclusiones generales</i>	<i>66</i>
	<i>Referencias bibliográfica</i>	<i>68</i>
	<i>Bibliografías.....</i>	<i>69</i>
	<i>Glosario de términos.....</i>	<i>72</i>

Introducción

El creciente avance y evolución de las nuevas tecnologías de la información y las comunicaciones (TIC) en el mundo de hoy, trae consigo consecuencias como el aumento, la capacidad de generación y almacenamiento de la información, por lo que las empresas y organizaciones mundiales se han visto en la necesidad de encontrar una nueva solución que registre esta gran cantidad de datos y a su vez, poder extraer información de gran importancia para el desarrollo productivo de estas.

El análisis de información permite tener informes o resultados guardados, validar las fuentes a utilizar, permitiendo que la información a brindar sea confiable y actualizada para la toma de decisiones. Además el análisis debe ofrecer alternativas de decisión a los usuarios y confianza en el analista. Este forma parte del proceso de adquisición y apropiación de los conocimientos acumulados en diferentes fuentes de información. El objetivo del análisis es buscar información que interesa al usuario, a partir de una gran cantidad de datos.

Hoy en día su utilización ha crecido en gran medida porque hacen más fácil el acceso de los usuarios a una gran variedad de información y pueden trabajar en conjunto, por lo que aumentan el valor operacional de las aplicaciones empresariales, en especial la gestión de relaciones con clientes. Al organizar la información se efectúan acciones apropiadas para conseguir un mejor control sobre la producción empresarial, concentrar y normalizar la información para brindar una visión global del comportamiento del negocio.

Con el avance de la tecnología, surge el potencial para resolver problemas de gran importancia. En las grandes empresas, los almacenes de datos han asumido la mayoría de las funciones antes realizadas con consultas. La desaparición total de

las consultas no es práctica ni debe ser el objetivo, puesto que siempre existirán consultas puntuales.

A pesar de las grandes ventajas de los AD, existen algunas importantes barreras para su utilización en empresas de tamaño mediano. Los productos AD han surgido para resolver problemas de análisis existentes de grandes masas de información, en empresas donde una pequeña diferencia en el valor de una variable, puede afectar la cuenta de resultado con unas diferencias de millones de dólares.

Los AD, están dimensionados para grandes empresas, contando con hardware muy potente muchas veces especializado y la masiva intervención de consultores externos, expertos en la realización de la puesta en marcha. Un proyecto de este tipo resulta en todos los aspectos excesivo para un departamento de ventas que necesita analizar la información de 500.000 - 3.000.000 de líneas de pedidos, o una cantidad equivalente de información financiera, que es lo normal para una empresa mediana o pequeña.

Para dar solución a problemas, necesidades similares, han surgido los MD, productos que utilizan la tecnología AD adaptada a las necesidades de las empresas medias.

Los esfuerzos de los desarrolladores de productos de MD, junto con las mejoras del índice, precio y rendimiento del hardware, suben constantemente el límite de penetración de los mercados, permitiendo asumir proyectos más y más importantes. La simplicidad de los proyectos y el menor coste en comparación con los AD, significan una ventaja competitiva muy grande a favor de los MD.

El MD tiene gran importancia hoy en día, uno de los principales lugares que se ha manifestado mayor auge ha sido dentro del sector empresarial, principalmente en la rama financiera, como ejemplo de este tema se puede ver las casas de bolsa,

aseguradoras, bancos, etc.

Es muy poco probable que una institución bancaria no cuente en la actualidad con un sistema de este índole. Una de sus principales razones para su uso es que su utilización es simple, se trata de un mercado en donde una decisión mal tomada pueda significar grandes pérdidas monetarias. Esta es una herramienta que permite lograr como ventaja la competitividad.

Cuba a pesar de ser un país subdesarrollado siempre ha puesto su empeño en el desarrollo de los nuevos avances tecnológicos y los diferentes procesos de la informática, a pesar del bloqueo económico, comercial y financiero impuesto por el gobierno de los Estados Unidos. El cuál priva a esta pequeña isla del acceso a la tecnología más avanzada en tan prometedor campo, lo cual limita de manera considerable la labor investigativa.

A pesar de eso, el altísimo nivel profesional de los especialistas cubanos se ha estado poniendo en práctica, actualmente diferentes implementaciones de AD se han desarrollado en el país como es el caso de la red nacional de Genética Médica, la Oficina Nacional de Estadística, Corporación Importadora y Exportadora etc.

Como muestra del desarrollo de estos avances, la Universidad de las Ciencias Informáticas, particularmente la facultad regional “Mártires de Artemisa” se encuentra desarrollando el proyecto AD para la administración provincial de Artemisa. Esta administración tiene como función principal la de preparar y proponer la política integral de la provincia. Además, se encarga de dirigirla y ejecutarla, así como de coordinar y controlar su cumplimiento sobre la base de las estrategias de desarrollo del país.

Este proyecto se está desarrollando con vistas de proveer la APA de una herramienta capaz de analizar los datos de una forma más eficiente y oportuna, en

especial se desarrolla un mercado de datos para la dirección de justicia. La cual cuenta con 3 departamentos y una unidad de prestación de servicios.

La dirección justicia tiene como misión ejecutar y asistir al gobierno del territorio en la implementación de la política jurídica aprobada para su esfera de atención; controlar el cumplimiento de las normas legales y las disposiciones metodológicas que correspondan y cumplir a su nivel con las funciones y tareas del organismo central que le son atribuidas.

Actualmente la dirección de justicia tiene las siguientes insuficiencias:

1. Existen pérdidas y duplicados de la información pues la misma se almacena generalmente en documentos impresos.
2. La información es accesible a cualquier personal no autorizado o no definido por la administración.
3. El análisis de la información resulta engorroso cuando se trata de grandes volúmenes.
4. La información es almacenada en aplicaciones informáticas de oficina (excel, word) o documentos impresos.

Esto trae consigo poca seguridad de la información, además dificulta el análisis de los datos para la toma de decisiones.

Esta situación problemática nos conduce a formular el siguiente **problema científico**: ¿Cómo contribuir a la seguridad y análisis de la información para la toma de decisiones en la dirección de justicia de la administración provincial de Artemisa?

El **objeto de estudio**: Almacén de datos.

Campo de acción: Mercado de datos en el sector justicia de instituciones gubernamentales

Objetivo general de la investigación: Desarrollar un mercado de datos que contribuya a la seguridad y análisis de la información para el apoyo a la toma de decisiones en la dirección justicia de la administración provincial de Artemisa.

Centrándose en los **objetivos específicos** que se definen a continuación:

- Realizar la fundamentación teórica del mercado de datos de la dirección justicia.
- Realizar el análisis y diseño del mercado de datos de la dirección de justicia.
- Implementar el mercado de datos de la dirección de justicia.
- Validar el mercado de datos de la dirección de justicia a través de los casos de prueba.

Para dar cumplimiento a los objetivos se definen las siguientes **tareas de investigación:**

1. Establecimiento de los fundamentos teóricos metodológicos que permitan el desarrollo del MD.
2. Caracterización de los procesos de Análisis y Diseño del mercado de datos.
3. Desarrollo de los subsistemas de integración y visualización del MD.
4. Validación del mercado de datos a través de los casos de prueba.

Idea a defender: Con el desarrollo del mercado de datos se contribuirá a la seguridad y análisis de la información en la toma de decisiones de la dirección justicia de la administración provincial de Artemisa.

Variable independiente: Mercado de datos para la dirección de justicia de la administración provincial de Artemisa.

Variables dependientes: La seguridad y el análisis de la información para la toma de decisiones en la dirección de justicia de la administración provincial de Artemisa.

Los métodos utilizados en esta investigación fueron: Los teóricos y los empíricos.

Métodos teóricos:

Método analítico–sintético, este método posibilita la descomposición de los componentes para estudiar por separado cada elemento y sus relaciones. Permitiendo el análisis de documentos, materiales, y temas relacionados con el mejoramiento del desarrollo del MD. Permitió definir los conceptos fundamentales del tema y se usó para especificar la metodología a utilizar. Para este método se hace necesario partir de un análisis de toda la información que fue recopilada, mediante otros métodos de investigación, como la entrevista; para luego centrarla y sintetizarla para elaborar el procedimiento adecuado.

Método histórico–lógico: que tiene como principal objetivo, analizar la evolución y surgimiento de los AD, posibilita documentar los temas relacionados con el desarrollo del MD en el sector justicia de instituciones gubernamentales.

Estos dos métodos fueron utilizados para el estudio del estado del arte de los procesos, soluciones y herramientas de integración de datos.

Dentro de los **métodos empíricos** utilizados se encuentra:

La **encuesta**, se realizó mediante un cuestionario previamente elaborado por el investigador, caracterizado por una serie de interrogantes y las respuestas son plasmadas de acuerdo al criterio del encuestado.

La **entrevista** se realiza mediante una conversación entre el investigador y el entrevistado con el objetivo de que el entrevistado proporcionara la información

necesaria para obtener los requisitos funcionales y no funcionales que se necesitan para el desarrollo del mercado de datos.

La **población** se encuentra constituida por 20 trabajadores de la dirección justicia de la APA, con una **muestra** de 11 especialistas y 2 directivos de esta dirección justicia, representando el 65% de la población actual.

Aporte práctico: consiste en la creación de un mercado de datos para el apoyo a la toma de decisiones en la dirección de justicia de la administración provincial de Artemisa.

Para lograr un mejor entendimiento, este trabajo quedó estructurado de la siguiente manera:

Una introducción en la cual se hace una explicación sintetizada de lo que va a tratar la investigación y se despliega además el diseño teórico metodológico.

Capítulo 1: Fundamentación teórica de los almacenes de datos.

En este capítulo se abordará toda la fundamentación teórica sobre los principales conceptos, metodologías y herramientas para el desarrollo del mercado de datos, que permitirá a la dirección de justicia apoyar el proceso de toma de decisiones.

Capítulo 2: Análisis y diseño del mercado de datos de la dirección de justicia de la administración provincial de Artemisa.

Se definen los requerimientos del sistema, se diseña el diagrama de casos de uso del sistema, el modelo de datos, los subsistemas de integración y visualización.

Capítulo 3: Implementación del mercado de datos de la dirección de justicia de la administración provincial de Artemisa.

En este capítulo se realizarán todas las transformaciones necesarias como Extraer, Transformar y Cargar (ETL), proceso que deja toda la información lista para la fase

siguiente, Inteligencia del Negocio (BI); donde se realizará la implementación del modelo de datos, los cubos OLAP y la política de seguridad de los usuarios.

Capítulo 4: Validación del mercado de datos de la dirección de justicia de la administración provincial de Artemisa.

En este capítulo se aplicarán los casos de prueba basados en los casos de uso para validar el mercado de datos.

Luego de la estructuración de estos capítulos se presentan las conclusiones, recomendaciones, referencias bibliográficas, bibliografía y el glosario de términos.

1 Capítulo 1. Fundamentación teórica

Este primer capítulo contiene la base teórica fundamental para la realización del trabajo de diploma, definiendo los conceptos por lo que se compone un MD, se realizará también un estudio de los distintos procesos de integración de datos. Además profundizar el conocimiento sobre las principales herramientas y metodologías utilizadas, así como software usados en la actualidad o en las que se apoya para la solución del problema que se plantea.

1.1 Sistemas de almacenamiento de la información

Los Sistemas de Almacenamiento de la Información (SAI) surgen a partir de la necesidad que se manifiesta para tener un control, un manejo de los datos, debido al gran cúmulo de información que mantienen las empresas, los negocios y como forma de abordar la necesidad de interpretación, análisis y clasificación de grandes volúmenes de datos en un dominio específico del conocimiento.

Muchas veces el gran volumen de datos es un problema que influye ya que en ocasiones se necesitan realizar consultas y extraer con mayor rapidez y efectividad la información, provocando que los sistemas realizados se fueran desvaneciendo ante la necesidad de elaborar el análisis de los datos. Los SAI son: “conjunto de componentes interrelacionados que recolectan (o recuperan), procesan, almacenan y distribuyen información para apoyar la toma de decisiones y el control en una organización” (Laudon, 2008).

Actualmente existen diversos sistemas de información, entre los más conocidos se encuentran: Los sistemas de ventas y marketing y los sistemas de apoyo a la toma de decisiones, los cuales permiten a los directivos la realización de una correcta toma de decisiones de forma rápida, exclusiva y poco estructurada. Entre los problemas que abordan se definen que para que se pueda llegar a una solución el

procedimiento del problema podría no estar definido con anterioridad.

Los MD, AD, y los almacenes de datos operacionales son algunos de los sistemas de apoyo a la toma de decisiones que más se utilizan en el mundo actual para brindar servicios a gerentes, directores y ejecutivos en una empresa.

1.1.1 Almacenes de datos

La aparición de los AD propició respuestas a grandes problemas existentes. En la actualidad se puede ver y afirmar que contiene una tecnología madura, estable y da solución a las problemáticas que se presentan. Esta es una tecnología como otra cualquiera, que continúa en constante evolución, a pesar de ser clasificada como madura.

Una forma más de adentrarse en este tema de los AD, es haciendo referencia a las principales personalidades que giran sobre este campo, a pesar que existen diferentes criterios en algunos aspectos por parte de estas, pero siempre giran sobre un mismo punto.

Inmon define un AD como un “conjunto de datos orientados por temas, integrados, variantes en el tiempo y no volátiles, que tienen por objetivo dar soporte a la toma de decisiones de los usuarios finales, integrando y depurando información desde uno o más sistemas operacionales. Este maneja un gran volumen de datos históricos donde los usuarios pueden realizar consultas”. (Inmon, 2005)

Según Kimball, conocido como principal promotor del enfoque dimensional para el diseño de AD lo definen como “una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis”. (Kimball, 2002)

El MD es una vista lógica de los datos en bruto provistos por su sistema de operaciones, finanzas hacia el AD con la adición de nuevas dimensiones o

información calculada.

Los MD son subconjuntos de datos de un AD para áreas específicas con el objetivo de responder a un determinado análisis, función o necesidad. Constituye una tecnología de bases de datos que ha tomado gran auge debido al crecimiento y muchas veces subutilización de los grandes bancos de datos históricos almacenados en las organizaciones. Al igual que en un AD, los datos están estructurados en modelos de estrella o copo de nieve y un MD puede ser dependiente o independiente de un almacén. El MD se destaca por una definición de requerimientos más fácil y rápida. También se simplifica el desarrollo de todo el mecanismo de su base de datos y con ello baja enormemente todo el coste del proyecto, así como su duración. Normalmente, un MD resuelve aplicaciones a nivel departamental.

Se puede decir de forma general que un AD no es más que una base de datos orientada al análisis de la información almacenada en ella, es caracterizada por la integración y depuración de la información de diversas fuentes para procesarla.

1.1.1.1 Características de los almacenes de datos

El AD presenta disímiles características para garantizar su desarrollo entre las cuales se encuentran:

Integrado: Los datos almacenados deben integrarse en una estructura consistente, todos los datos deben almacenarse en un mismo formato.

Temático: Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales y se centran en entidades de alto nivel, no en los procesos.

Histórico: El AD se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones. Se utiliza para realizar análisis de tendencias.

Y estos tienen que ser incrementados periódicamente.

No volátil: El AD puede ser leído pero no modificado. Es decir, se incorporan los últimos valores que tomarán las distintas variables contenidas en él, sin ningún tipo de acción sobre los valores que ya existían. (Herrera, 2010)

1.1.1.2 Ventajas y desventajas del almacén de datos

Los AD presentan ventajas como son la facilidad de la información, presentan competitividad en el mercado, permite realizar un análisis más preciso y consistente, esta herramienta ayuda a tener mejores tiempos de respuestas y mejora el proceso de producción. Mejora la entrega de información completa, correcta, consistente, oportuna y accesible. También presenta varios inconvenientes como son la complejidad de la integración, requiere una gran inversión, debido a que su correcta construcción no es sencilla y consume grandes recursos, además, su implementación implica desde la adquisición de herramientas de consulta y análisis, hasta la capacitación de los usuarios.

1.1.2 Mercados de datos

Kimball en conjunto con otros autores definen un MD como: “Un conjunto flexible de datos, idealmente basado en el dato más atómico posible granular para ser extraído de las fuentes operacionales y presentado en un modelo simétrico dimensional que es más resistente cuando se enfrentan con las más inesperadas consultas de los usuarios”.

Se puede decir que los MD están conectados con la arquitectura de los AD en su forma más simple y que representan los datos de un sólo proceso del negocio a la vez. (Wolff, 2000)

Los MD son en esencia un AD, pero de forma más reducida a un departamento específico siendo así la fuente de análisis de la Dirección Justicia, La unión de

todos los mercados de las direcciones es lo que conformaría el Almacén de datos para la administración provincial de Artemisa.

1.1.2.1 Características de los mercados de datos

- Se centran en los requisitos de los usuarios asociados a un departamento o área de negocio concretos.
- Como diferencia con los AD, los mercados no contienen datos operacionales detallados.
- Son más sencillos a la hora de utilizarlos y comprender sus datos, debido a que la cantidad de información que contienen es mucho menor que en los AD.

1.1.2.2 Diferencias entre mercado de datos y almacenes de datos

Como diferencias que presenta el MD es que se centra solamente en los requerimientos de usuarios asociados con un departamento o función de negocio, además de que los MD normalmente no contienen datos operacionales detallados a diferencia de los AD. Debido a que los mercados contienen menos información comparados con los AD, los MD son más fáciles de entender y de navegar.

Al hablar de los MD se ponen de manifiesto algunos conceptos erróneos, como es la comparación con los AD y al final se concreta que son como estos, pero en versión pequeña, y en cierto modo es así.

Como errores que se presentan sobre la implementación y funcionamiento de los mercados se dice que los AD a veces son más simples de implementar, este planteamiento se define como falso ya que la implementación es muy similar, y debe proporcionar las mismas funcionalidades. Se dice que son pequeños conjuntos de datos y en consecuencia tienen menor necesidad de recursos, esto también es falso, debido a que una aplicación corriendo sobre un MD necesita los

mismos recursos que si corriera sobre un almacén.

Las consultas son mucho más rápidas, dado el menor volumen de datos, esta característica que se plantea es errónea porque, el menor volumen de datos se debe a que no se tienen todos los datos de toda la empresa, pero sí se tienen todos los datos de un determinado sector de la empresa, por lo que una consulta sobre dicho sector tarda lo mismo si se hace sobre el MD que si se hace sobre el AD.

En algunos casos se añade tiempo al proceso de actualización, esta información es incierta debido que al actualizar el mercado de datos desde el almacén cuesta menos, ya que los formatos de los datos son o suelen ser idénticos, que actualizar el almacén de datos desde sus fuentes de datos primarios, donde es necesario realizar operaciones de transformación.

1.2 Arquitectura del diseño

La arquitectura de una aplicación de MD se puede definir en términos similar a un AD. Es conveniente que una aplicación de MD se desarrolle considerando que va a ser explotada en una arquitectura de tres capas las cuales son: la capa de presentación la que va a contener los componentes de interfaz de usuario y de despliegue. La segunda sería la capa lógica analítica que contendría las consultas, procesos, y funciones de formateo y por último la capa de datos.

La arquitectura de MD establece el marco de trabajo, estándares y procedimientos para el mercado a un nivel departamental, por el contrario, de los AD que lo hace a un nivel empresarial. Los objetivos que presenta la arquitectura son simples, integrar al mercado las necesidades de información empresarial. Uno de los principales resultados del desarrollo de la arquitectura AD incluyen el modelo de datos fuente, el modelo de datos conceptual, la arquitectura tecnológica, los estándares y procedimientos y el plan de implementación incremental para el AD.

1.3 Modelo de datos del almacén de datos

En cuanto al modelo de datos del AD se pueden encontrar el modelado dimensional, ya que es uno de los modelos más usado y permite un manejo de los datos más adecuado. Está basado en el esquema estrella y copo de nieve que están compuestos por:

Dimensiones: las cuales son un atributo estructural de los cubos y sirven como un mecanismo de selección de datos, organizadas en jerarquías de categorías y niveles que describen los datos de las tablas de hechos

Hechos: las cuales contienen datos de interés que presentan un nivel de granularidad.

1.3.1 Esquema estrella

El esquema estrella como bien indica su nombre presenta forma de estrella, con puntos radiales desde el centro. El centro de la estrella consiste de una o más tablas de hechos, y las puntas de la estrella son las tablas dimensionales. Este modelo entonces, resulta ser asimétrico, pues hay una tabla dominante en el centro con varias conexiones a las otras tablas. Las tablas dimensionales tienen sólo la conexión a la tabla de hecho y ninguna más.

1.3.2 Esquema copo de nieve

La diferencia del esquema copo de nieve comparado con el esquema estrella, está en la estructura de las tablas dimensionales: estas tablas en el esquema copo de nieve están normalizadas. Cada tabla dimensional contiene sólo el nivel que es la clave primaria en la tabla y la llave foránea de su parentesco del nivel más cercano del diagrama.

1.3.3 Tabla de hechos

La tabla de hecho es el esquema dimensional, donde se almacenan las mediciones numéricas del negocio. Estas medidas se hacen sobre la granularidad de la tabla, esta queda determinada por el nivel de detalle que se almacenará en la tabla. Cada medida es tomada de la intersección de las dimensiones que la definen. Idealmente está compuesta por valores numéricos, continuamente evaluados y aditivos.

La clave de la tabla de hechos recibe el nombre de clave compuesta o concatenada, debido a que se forma de la concatenación de las llaves primarias de las tablas dimensionales a las que está unida. Así entonces, se distinguen dos tipos de columnas en una tabla de hechos: columnas de hecho y columnas llaves. Donde la columna de hecho es la que almacena alguna medida de negocio y una columna llaves forma parte de la clave compuesta de la tabla.

1.3.4 Jerarquías

Son grupos de atributos que siguen un orden preestablecido. Una jerarquía implica una organización de niveles dentro de una dimensión, con cada nivel representando el total agregado de los datos del nivel inferior. Las jerarquías definen cómo los datos son sumariados desde los niveles más bajos hacia los más altos. Una dimensión típica soporta una o más jerarquías naturales.

Una jerarquía puede, pero no exige contener todos los valores existentes en la dimensión. Se debe evitar caer en la tentación de convertir en tablas dimensionales separadas cada una de las relaciones muchos a uno presentes en las jerarquías.

1.3.5 Niveles

Un nivel representa un nivel particular de agregación dentro de una dimensión; cada nivel representa la sumariación total de los datos desde el nivel inferior. Como ejemplo de nivel se presenta la dimensión Tiempo que presenta cuatro

niveles: mes, semestre, trimestre año. El nivel mes representa el nivel base, el trimestre es el acumulado de los 3 meses, el nivel semestre representa la sumarización de los totales por trimestre, y el nivel año representa la sumarización de los totales para los semestres. Agregar niveles de sumarización otorga flexibilidad adicional a usuarios para analizar los datos.

1.3.6 Granularidad

Es el nivel de detalle o de desagregación común por el cual se definen los datos que se almacenan en una tabla de hechos. También se puede definir como el nivel más bajo de información que será almacenado en dicha tabla. Influye en la cardinalidad tanto de la tabla como de las dimensiones, a mayor granularidad mayor será el número de registros.

1.3.7 Modelo dimensional

El modelo dimensional ha ganado gran popularidad en el mundo de los AD en los últimos años. Como bien lo indica su nombre, este modelo permite realizar un modelado y representar las estructuras, siendo esto consecuente con las especificaciones y exigencias del negocio. Para la materialización de este modelo se utiliza la propuesta de Ralph Kimball llamada esquema de estrella que consiste a grandes rasgos, en una tabla central denominada tabla de hecho y un conjunto de pequeñas tablas llamadas dimensiones que se relacionan con esta tabla central se les denomina esquema de estrella por su similitud a una estrella natural.

El modelo dimensional divide el mundo de los datos en dos grandes grupos: las medidas y las dimensiones del entorno a estas medidas. Las medidas que generalmente son numéricas, se almacenan en tablas de hechos y las descripciones de los entornos que son textuales, se almacenan en las tablas dimensiones. Las tablas de hechos son tablas primarias en el modelo dimensional y contiene los valores del negocio.

Cada proceso genera un diagrama dimensional, y su representación recibe el nombre de cubo. Cada cubo va a contener un conjunto de dimensiones y hechos. Las dimensiones son los objetos o elementos que, relacionados entre sí, determinan el proceso en el que participan. Los hechos son las medidas numéricas que caracterizan las acciones que ocurren entre los distintos objetos o dimensiones.

El modelo dimensional debe estar preparado para almacenar el estado real de cada dimensión en cada momento, por lo que debe definirse, de acuerdo a las necesidades que se tengan, el comportamiento a seguir ante la ocurrencia de cambios. Para manejar las modificaciones en las dimensiones existen diversas técnicas, siendo una de ellas sobrescribir la información almacenada. Otras más especializadas y frecuentemente usadas para los AD son la de crear un nuevo récord con la información duplicada, excepto por el o los campos que se modifiquen, o agregar un campo en la misma fila, manteniendo un valor anterior y uno actual. Estas técnicas son conocidas como slowly changing dimension (Kimball, 1996).

Una ventaja de utilizar el Modelo Dimensional es que admite la adición de dimensiones y hechos que no se habían previsto, sin que esto implique volver a cargar los datos ya almacenados. Es posible adicionar nuevos hechos en la tabla de hechos siempre que se mantenga el mismo nivel de granularidad, y se pueden adicionar dimensiones si ellas no generan más de un valor para cada fila en la tabla de hecho.

También es posible agregar nuevos atributos a las dimensiones. Esta característica de gran adaptabilidad es muy deseable, pues a medida que los analistas añaden nuevos requerimientos al sistema, se pueden ir incorporando las modificaciones sin que esto implique demasiados cambios.

1.4 Procesamiento analítico en línea

El proceso de análisis se realiza de igual forma lo que varía en uno y otro caso es la metodología de almacenamiento. Permite al usuario seleccionar y extraer la información desde diferentes puntos de vista, se traduce en un conjunto de técnicas y estándares. Tiene como objetivo agilizar el proceso de consulta de grandes volúmenes de datos.

La forma de almacenamiento es crítica para garantizar la velocidad de recuperación de la información, las zonas de ubicación de las agregaciones y el procesamiento de los datos en general.

El modelo ROLAP (Relational Online Analytical Process, en inglés) es usado fundamentalmente sobre información que no se consulta frecuentemente. Por ejemplo información histórica de muchos años de antigüedad. Los datos son acumulados en filas y columnas de forma relacional, y se les presentan a los usuarios en forma de dimensiones del negocio. (Luján, 2006)

EL MOLAP (Multidimensional Online Analytical Process, en inglés) Son herramientas que acceden a datos que no están almacenados en registros de tablas, sino que almacenan los datos en arreglos de varias dimensiones, llamados cubos. Estos cubos utilizan índices para optimizar el acceso a los datos.

El almacenamiento HOLAP (Hybrid Online Analytical Process, por sus siglas en inglés), como su nombre lo indica, es un híbrido entre los métodos ROLAP y MOLAP. Permite almacenar una parte de los datos como en un sistema MOLAP y el resto como en uno ROLAP. El grado de control que el operador de la aplicación tiene sobre este particionamiento varía de unos productos a otros.

El modelo ROLAP es el aceptado para la realización de este MD por lo antes planteado.

1.5 Tecnología y tendencias actuales

Con el actual avance tecnológico las compañías y empresas necesitan cada día más realizar análisis a la información que utilizan así como a los datos históricos. Una de las principales herramienta que se ha encontrado para contrarrestar esta problemática, provoca cambiar la manera de almacenamiento de los datos usando como principal herramienta los AD.

Los AD han tenido un auge considerable y se han visto obligada a desarrollarse debido a la gran competencia entre las grandes empresas, ya que estas necesitan de una herramienta que les permita realizar el análisis de información con mayor rapidez. Constituyen uno de los soportes fundamentales para el proceso de toma de decisiones gerenciales; de ahí la importancia de que la información guardada en ellos sea confiable y con calidad. Uno de los procesos en la construcción de estos y que contribuye a lograr este objetivo es la limpieza de datos, y junto con ella la estandarización de direcciones.

Se puede decir que las empresas que utilizan AD cuentan con mayor aceptación de sus productos. El AD hace lo posible para aprovechar el valor potencial de los recursos de información de la empresa y convertir ese valor potencial en valor verdadero. El cual extiende el alcance de la información para que se acceda directamente en línea, lo que a la vez contribuye en su capacidad para operar con mayor efectividad las tareas rutinarias.

Las corporaciones de hoy se esfuerzan por conducir sus negocios hacia una base internacional. Vemos compañías que surgieron en Estados Unidos y se expandieron a Europa, Asia y África. La expansión del negocio crea la necesidad de acceder a datos corporativos que están ubicados en diferentes puntos geográficos, lo que puede ser un gran problema para el usuario. La solución es creando

versiones más pequeñas del AD, los mercados de datos. Estas versiones se crean usando algún criterio particular, como por ejemplo el lugar geográfico.

Los MD se implementan antes que el AD, como un plan piloto, siendo el conjunto de mercados específicos orientados al AD. Se debe restringir el acceso al mercado y a las bases de datos con las que trabaja a sólo personal autorizado. En la mayoría de las empresas se puede apreciar que cuentan de una u otra manera con diferentes mercados, lo cual esta muy bien, ya que las empresas de hoy se han vuelto inforveras, es decir, tienen la necesidad constante de consumir información para poder sobrevivir.

Sin embargo muchos de estos mercados fueron creados enfocados en los datos y no en las necesidades de información de los usuarios, por lo que algunas empresas aceptan el reto de implementar un MD y muchas de ellas lo implementaron con éxito como son por ejemplo: el Ministerio de Justicia de Lima en Perú, el Cuadro de Mando de la Administración de Justicia en España, que desarrolla un sistema de inteligencia de negocio.

En Cuba también se usa esta tecnología ya que ha estado al tanto de los últimos cambios tecnológicos, aunque faltan pasos por dar en este tema. Las empresas cubanas aspiran a un mejor manejo de sus datos y mejor visión de los mismos, un ejemplo fehaciente de tal desempeño es la Universidad de las Ciencias Informáticas la cual ha diseñado un mercado de datos para la sala situacional de un Sistema de Gestión Penitenciaria tiene como objetivo el soporte de las decisiones estratégicas del Ministerio del Interior y Justicia y de la Dirección General de Custodia y Rehabilitación del Recluso, contribuyendo a garantizar el respeto a los derechos de los internos, su actividad de rehabilitación y reinserción en la sociedad.

Según el estudio realizado del estado actual de los almacenes de datos se determinó que algunos de los sistemas encontrados se asemejan en parte al que

se pretende desarrollar, diferenciando la información que contiene esta dirección, por lo que ninguna de estas soluciones son factibles, y se decide implementar una solución que cubra las necesidades de esta dirección justicia con sus departamentos.

1.6 Metodologías para el desarrollo de un mercado de datos

Para lograr el desarrollo del mercado de datos con la calidad requerida se necesita tener conocimiento previo de lo que es una metodología y los distintos tipos que existen. Se puede decir que una metodología no es más que un conjunto de herramientas y técnicas que se son usadas para la realización de dicho producto. Dentro de las metodologías conocidas para el desarrollo de un MD se encuentran: Metodología de William H. Inmon se basa en el enfoque descendente y propone en primer lugar la construcción del almacén de datos y a partir de este la realización de los mercados, también afirma la creación del repositorio corporativo como fuente de información histórica de carácter persistente y con calidad.

Existen otras metodologías como son DM2 la cual se basa en las necesidades de información a nivel gerencial, donde la información debe ser encarada como patrimonio de la empresa, accesible a quien la necesite. Esta metodología acorta en función razonable el tiempo entre el inicio del análisis y la implantación. Esta rapidez no solo es buena para el cliente sino que también es exigida y necesaria por el propio ambiente que lo rodea. La metodología Hefesto, es sencilla, ordenada, explícita y efectiva.

Después de un estudio sobre las metodologías se define la seleccionada para la construcción del MD de la dirección justicia de la administración provincial de Artemisa es la metodología de Proceso de Desarrollo en la Línea Soluciones de Almacenes de Datos e Inteligencia de Negocio. Fue elaborada a partir del estudio y análisis de la metodología de Kimball y teniendo en cuenta la experiencia de los

recursos humanos en otros proyectos desarrollados. Es una metodología mixta que reúne elementos de varias metodologías de desarrollo de proyectos de integración de datos, ya que cubre todas las fases por las que pasa su construcción, desde el levantamiento de información inicial, la arquitectura, el diseño, la implementación de la herramienta de inteligencia de negocio, prueba, despliegue, soporte, hasta la gestión del proyecto.

1.7 Herramientas para el desarrollo de un mercado de datos

1.7.1 Visual Paradigm

Visual Paradigm en su versión 6,4 es usada para la realización de este mercado. Es una herramienta que se encuentra desarrollada por Visual Paradigm Internacional una de las principales compañías de herramientas CASE.

Tiene como características que es un producto de calidad, soporta aplicaciones Web, presenta generación de código para Java y exporta como HTML. Es fácil de instalar y actualizar y tiene compatibilidad entre ediciones, propicia un conjunto de ayudas para el desarrollo de programas informáticos, desde la planificación, pasando por el análisis y el diseño, hasta la generación del código fuente de los programas y la documentación. Actualmente es muy utilizada internacionalmente debido a su licencia gratuita y comercial.

Esta herramienta de modelado ha sido concebida para un ciclo completo del proceso de desarrollo del software, es de gran utilidad para el trabajo del analista. Fue diseñado para los usuarios interesados en la construcción de sistemas de software.

1.7.2 Power Architect

Herramienta innovadora, pensada específicamente para el almacenamiento y

repositorio de datos. Con ella, el diseñador podrá abrir múltiples conexiones concurrentes a bases de datos, crear y explorar perfiles de datos fuente, arrastrar y soltar esquemas de datos, tablas y columnas dentro del modelo de datos, y confeccionar la base de datos resultante con su plantilla ETL asociada.

Hasta el diseño más complicado, la base de datos más grande o el modelo más inabarcable, podremos gestionarlo y manejarlo con Power Architect. Con esta herramienta se facilita mucho el trabajo a los arquitectos de mercados de datos, permite modelar todos los esquemas, hacer ingeniería inversa, autogenerar metadatos de los ETL. Esta es una herramienta completamente gratis para uso personal. Soporta múltiples plataformas de bases de datos.

1.7.2.1 Versiones que presenta Power Architect.

Versión 0.9.6

Puede incluir actualizaciones no especificadas, mejoras o correcciones de errores.

Versión 1.0.6

Aunque es un producto comercial, la versión community es Open Source y gratuita. Frecuencia de actualización alta. La versión community gratuita es suficiente para un trabajo de modelado básico. Soporta ingeniería directa e inversa de forma cómoda (se pueden agregar fácilmente las tablas hijas) con bastantes opciones gráficas de presentación. Es rápido, amigable y con él se puede generar una documentación mínima decente. De todos los analizados es el que menos opciones de modelado y documentación tiene, aunque cumple su cometido muy dignamente.

Versión 0.9.13

Es una herramienta de modelado visual de datos diseñado para los arquitectos de datos, analistas, diseñadores y otros profesionales. Presenta un diseño rápido de cada aspecto de su modelo de datos utilizando diagramas y una vista jerárquica de

la estructura de su modelo. Su modelo de datos sigue siendo independiente de la plataforma, lo que le permite mantener un esquema de base de datos única que funciona bien con las plataformas de bases de datos.

Power Architect es también muy adecuado para el almacenamiento de datos y el diseño de mercado de datos. Puede abrir bases de datos de múltiples fuentes al mismo tiempo, y luego arrastrar y soltar objetos, tales como esquemas, tablas y columnas.

Proporciona una gran variedad de herramientas para ver y comparar las estructuras de datos y las asignaciones. Por ejemplo, puede comparar la estructura de dos bases de datos, destacando las diferencias y las similitudes y generar las sentencias DDL para sincronizarlos. También puede crear un informe de mapeo visual listado de las tablas de origen utilizados en el modelo de datos, o crear un perfil fácil de leer, resumir los datos contenidos en una base de datos.

Tanto si está creando o manteniendo un modelo de datos, Power Architect proporciona las herramientas para ayudarle a diseñar su modelo en una fracción del tiempo. Esta son características que permiten este programa, para al buen desarrollo de productos en la facultad, ya que esta versión es la seleccionada para el desarrollo del MD.

1.7.3 Pentaho data integration

Pentaho data integration es una de las herramientas de extracción, transformación y carga, la cual permite realizar el diseño gráfico de la transformación ETL. Reúne un conjunto de componentes que permiten modelar y ejecutar transformaciones sobre flujos de datos. Puede funcionar sobre varias plataformas como son la Microsoft Windows, Linux, Solaris etc.

Esta herramienta brinda soporte para metadatos. Incorpora operaciones de transformación, así como funciones que permiten operar con los campos en el flujo de datos, renombrando, calculando campos en función de otros, correlacionando valores y realizando búsquedas auxiliares en bases de datos. ofrece, soporte para operaciones de dimensiones lentamente cambiantes, permite ejecutar código javascript dentro de las transformaciones e incorpora un evaluador de expresiones regulares. Entre las tareas que se pueden incorporar están la copia y eliminación de ficheros, así como su transportación usando FTP, además se pueden descompactar ficheros, esta herramienta es fácil de usar.

1.7.4 Pentaho BI server 3.8.0

Provee el soporte y la infraestructura necesaria para crear soluciones de inteligencia empresarial a problemas de negocios. El marco proporciona los servicios básicos, incluidos autenticación, registro, auditoría, servicios web y motor de reglas. La plataforma también incluye un motor de solución que integra reportes, análisis y componentes de minería de datos. Su diseño modular y la arquitectura basada en plug-in permite a toda, o parte de la plataforma, estar inmersa en aplicaciones de terceros por los usuarios finales, así como fabricantes de equipos originales. (Vega, 2008) Por estas características esenciales es que se decide usar esta herramienta.

1.7.5 Mondrian schema workbench 3.2.1

Es una interfaz de diseño que le permite crear y probar esquemas de cubos OLAP visualmente. Los modelos de esquemas XML de metadatos se crean en una estructura específica utilizada por el motor de mondrian. La estructura de estos modelos se pueden considerar de forma de cubos, que utilizan hechos existentes y tablas de dimensiones que se encuentran en su gestor de base de datos. (Luján,

2006) Esta herramienta es la seleccionada para el desarrollo de los cubos OLAP del MD de la dirección de justicia de la administración provincial de Artemisa.

Conclusiones

A partir de la investigación de diversas soluciones actuales se realizó un estudio de las tendencias, tecnologías y herramientas actuales para el diseño e implementación del mercado de datos. Con este estudio se determinó utilizar como herramienta CASE Visual Paradigm, para realizar el diseño del MD, PostgreSQL como sistema gestor de base de datos, pertenecientes a la familia de software libre. Además se definió como metodología a utilizar: Kimball y se puntualizaron los principales componentes de la arquitectura de un MD.

2 Capítulo 2: Análisis y diseño del mercado de datos de la dirección de justicia de la administración provincial de Artemisa

El propósito central de este capítulo es presentar los artefactos que ayudan a construir la propuesta, se plasman las necesidades de información, los requisitos funcionales y no funcionales, además de las reglas del negocio identificadas en el levantamiento de información. Se muestra el diseño de diagramas de casos de uso del sistema, el modelo de datos, los subsistemas de integración y visualización.

2.1 Definición del negocio

La dirección justicia la cual cuenta con 3 departamentos y una unidad de prestación de servicios, tiene la misión de ejecutar y asistir al Gobierno del territorio en la implementación de la política jurídica aprobada para su esfera de atención; controlar el cumplimiento de las normas legales y las disposiciones metodológicas que correspondan y cumplir a su nivel con las funciones y tareas del organismo central que le son atribuidas.

Actualmente la dirección de justicia de la administración provincial de Artemisa recibe los datos desde los centros informantes en ficheros de texto, documentos impresos o son captados de forma manual, esta información se almacena durante años por lo que se hace difícil su consulta y al contar con grandes volúmenes de información generada, impide la agilidad y rapidez necesaria durante sus procesos de trabajo.

Esto implica la pérdida de información, atraso y lentitud en las entregas a la alta gerencia, imposibilitando la correcta toma de decisiones. Además no se cuenta con una herramienta de reportes para los directivos lo que trae consigo el duplicado de la información, la ausencia de un mecanismo de control de la misma, accesos no autorizados o no definidos por la administración y que los datos con que se trabajen

no sean certeros.

2.2 Temas de análisis

Para permitir una mejor organización de la propuesta se definieron varios temas de análisis los cuales son:

- Análisis de los indicadores Cuadro.
- Análisis de los indicadores Potencial Científico.
- Análisis de los indicadores Planteamientos Ante el Delegado.
- Análisis de los indicadores del Registro Civil dentro del Departamento Registral.
- Análisis de los indicadores del Registro Propiedad dentro del Departamento Registral
- Análisis de los indicadores de Solicitud penal expedientes revisión, dentro del Departamento Asistencia Jurídica.
- Análisis de los indicadores de la Consultoría Jurídica dentro del Departamento Asistencia Jurídica.
- Análisis de los indicadores de la Actividad Notarial dentro del Departamento Notaría.
- Análisis de los indicadores del Registro Mercantil, dentro del Departamento Unidad de Prestación de Servicios.
- Análisis de los indicadores del Registro de Asociaciones, dentro del Departamento Unidad de Prestación de Servicios.
- Análisis de los indicadores de Secretaría Interna, dentro del Departamento Unidad de Prestación de Servicios.

2.3 Reglas del negocio

Se definen las reglas de negocio identificadas en el levantamiento de información y en el análisis de las fuentes. Dichas reglas son entrada fundamental para los

procesos de diseño del almacén, extracción, transformación y carga e inteligencia de negocio.

- La información está dividida por departamentos, por lo que existe un gran flujo de información.
- Los ficheros .xls guardan la información mensual, semestral, trimestral y anualmente.
- En caso de presentar algún error en la información el centro de información lo soluciona, de no ser así, se acerca a la dirección justicia y en caso de que esta no pueda darle solución se ve al municipio que brindó la información. Además el sistema lanzará una alerta.
- Los datos nulos se les pone 0.
- En el caso de los asentamientos se define como una jerarquía de municipio.
- Debe de existir nomenclador de país.
- Debe de existir nomenclador organismo.
- Los códigos de los nomencladores deben de aparecer en los clasificadores de nomencladores.

2.4 Especificación de los requisitos de software

En el proceso de análisis fueron identificados los requerimientos necesarios para satisfacer las necesidades del cliente y lograr la realización de un producto con calidad.

2.4.1 Requisitos informativos

Los Requisitos Informativos (RI), no son más que la información que debe estar disponible, la entrada fundamental para todo el proceso de inteligencia del negocio y los futuros reportes bases. Los reportes que actualmente se efectúan con dicha información son:

RI1: Obtener la cantidad de aprobados, cantidad de cubiertos y cantidad de

vacantes por Semestre, DPA, Organismo y Cargo.

RI2: Obtener cantidad de PCC, cantidad de UJC, Otras por Semestre, DPA y Organismo.

RI3: Obtener cantidad de categoría, cantidad de mujeres por Semestre, DPA, Organismo y Categoría.

RI4: Obtener cantidad por Trimestre, DPA y Ind_Planteamiento.

RI5: Obtener cantidad de hospitales, cantidad en el registro dentro del término y cantidad en el registro fuera del término por Mes, DPA, Organismo, e Inscripciones de nacimiento.

RI6: Obtener cantidad de Inscritos, cantidad concertados y cantidad retroactivos por Mes, DPA, Organismo e Inscripciones de Matrimonio.

RI7: Obtener la cantidad, con lugar, con lugar en parte y sin lugar por Mes, DPA, Organismo e Ind_Expediente.

RI8: Obtener la cantidad por Mes, DPA, Organismo e Ind_Actos.

RI9: Obtener la cantidad por Mes, DPA, Organismo e Ind_Búsquedas.

RI10: Obtener la cantidad por Mes, DPA, Organismo e Inscripciones de Defunción.

RI11: Obtener cantidad, cantidad de provenientes de propio registro, cantidad de proveniente de otros registros y cantidad de provenientes de tribunales por Mes, DPA, Organismo e Ind_Notas Marginales.

RI12: Obtener cantidad en extracto, cantidad literal y cantidad exentas por Mes, DPA, Organismo y Certificaciones Expedidas.

RI13: Obtener cantidad por Mes, DPA, Organismo e Ind_Solicitud de certificaciones.

RI14: Obtener cantidad nacional, cantidad del exterior y total por Mes y Registro.

RI15: Obtener certificaciones expedidas, muestras comprobadas y porciento por Mes y DPA.

RI16: Obtener cantidad de certificaciones expedidas y total general por Mes, DPA y Certificaciones Expedidas.

RI17: Obtener la cantidad de nacidos fuera de provincia, cantidad de nacidos en provincia, cantidad de tarjetas entrantes, cantidad de tarjetas no entrantes, porciento, cantidad de causas de no entrega por el artículo 48, cantidad de causas de no entrega por atlas, cantidad de causas de no entrega por número de carné de identidad, cantidad de causas de no entrega por otras, reconocimiento y subsanación por Trimestre y DPA.

RI18: Obtener cantidad de total de libros de nacimiento, cantidad de total de libros de matrimonios, cantidad de total de libros de defunciones, cantidad de total de libros de ciudadanía, cantidad de total de libros, cantidad de encuadernados, cantidad de reconstruidos, cantidad transcriptos, cantidad de libros en blanco de nacimiento, cantidad de libros en blanco de matrimonios, cantidad de libros en blanco de defunciones, cantidad total de libros en blanco, cantidad de libros pendiente de encuadernado, cantidad de libros pendiente de transcribir, cantidad de libros pendiente de reconstruir por Trimestre y DPA.

RI19: Obtener cantidad notas enviadas cambio nombre, cantidad notas enviadas matrimonio, cantidad notas recibidas matrimonio, cantidad notas recibidas nacimiento, cantidad notas enviadas nacimiento, cantidad notas enviadas defunción, cantidad notas recibidas defunción, cantidad notas enviadas reconocimientos, cantidad notas recibidas totales, cantidad notas recibidas reconocimientos, cantidad notas enviadas otros, cantidad notas recibidas otros, cantidad notas enviadas totales, cantidad notas recibidas subsanaciones error, cantidad notas enviadas subsanaciones error, cantidad notas recibidas cambio nombre por Trimestre y DPA.

RI20: Obtener cantidad de matrimonios, cantidad de expedientes con modificaciones o adiciones, cantidad de subsanación de errores, cantidad de

inscripciones fuera de término, total, cantidad de declaraciones juradas, cantidad de búsquedas en provincia, cantidad de búsquedas fuera de provincia, total búsquedas, cantidad de notas marginales, cantidad de certificaciones extracto, cantidad de certificaciones literal, total certificaciones, cantidad de llamada en provincia, cantidad de llamada fuera de provincia, total llamadas, total actos, e ingreso de sello por Mes y Palacio.

RI21: Obtener total radicadas en el año, inscripciones radicadas suspendidas, inscripciones radicadas denegadas, total inscritas, inscripciones radicadas canceladas, inscripciones no estatales por documentos notariales, inscripciones radicadas tramites, inscripciones no estatales escritura obra nueva, inscripciones no estatales documentos administrativos, inscripciones no estatales resolución judicial, inscripciones no estatales total particulares, inscripciones no estatales inscrita segunda vez mas, inscripciones no estatales inscritas primera vez, inscritas estatales subordinación nacional, inscritas estatales subordinación local, inscritas estatales solares yermos, inscritas estatales obras nuevas, inscritas estatales aprobados inversión extranjera, inscritas estatales contrato de arrendamiento, inscritas estatales edificio multifamiliar, inscritas estatales finca rustica, total inscripciones estatales, inscritas estatales primera vez, inscritas estatales segunda vez o mas, radicadas pendientes inscribir anno anterior, anno, total manzanas ordenadas por Mes y DPA.

RI22: Obtener cantidad y total por Trimestre, DPA, Organismo, y Actividad del Registro de la Propiedad.

RI23: Obtener cantidad de inscripciones y total por Mes, DPA y Organismos

RI24: Obtener el plan total, el real total, el plan del 1er trimestre, el real del 1er trimestre, el plan del 2do trimestre, el real del 2do trimestre, el plan del 3er trimestre, el real del 3er trimestre, el plan del 4to trimestre, el real del 4to trimestre por Mes y Entidad.

RI25: Obtener el plan total, el real total, el plan del 1er trimestre, el real del 1er trimestre, el plan del 2do trimestre, el real del 2do trimestre, el plan del 3er trimestre, el real del 3er trimestre, el plan del 4to trimestre, el real del 4to trimestre por Mes y DPA.

RI26: Obtener cantidad de revisiones penales por Trimestre, DPA, Organismo y Revisión Penal.

RI27: Obtener cantidad por Trimestre, DPA, Organismo e Indicadores Generales.

RI28: Obtener cantidad por Trimestre, DPA, Organismo y Contrato de Servicios.

RI29: Obtener cantidad por Trimestre, DPA, Organismo y Asuntos tramitados.

RI30: Obtener cantidad por Trimestre, DPA, Organismo y Situación Financiera.

RI31: Obtener cantidad de Asesores propios por Trimestre, DPA, Organismo y Organismos.

RI32: Obtener la cantidad de ccs fortalecidas, cantidad de ccs no fortalecidas, cantidad total contratados, cantidad iguala tradicional, cantidad eventual, cantidad nueva modalidad, cantidad contratadas por unidades de bufetes, cantidad con asesores propios, porciento atendido, sin servicios jurídicos, cantidad total entidades territorio, por Trimestre, DPA, Organismo y Sector.

RI33: Obtener cantidad de ingresos por Semestre, DPA, Organismo e Ingreso.

RI34: Obtener cantidad de salario promedio devengado por Semestre, DPA, Organismo y Salario.

RI35: Obtener cantidad de indicadores de la actividad notarial por Mes, DPA, Organismo y Documento.

RI36: Obtener importe en MN, importe en CUC por Mes, DPA, Organismo, y Recaudación.

RI37: Obtener cantidad pendiente, cantidad pendiente del trimestre anterior, cantidad solicitados, cantidad de entregados en término, cantidad de entregados

fuera de término, cantidad pendientes en término, cantidad pendientes fuera de término, cantidad desistidos y cantidad consultadas por Trimestre y DPA.

RI38: Obtener primera inscripción, reactivación, extinción, certificación, deposito_balance, cambio directivos, director general, cambio objeto social, cambio denominación, cambio domicilio, otros y suspensión por Trimestre, DPA y Organismo.

RI39: Obtener las aprobadas, las aprobadas canceladas y las aprobadas que funcionan por Trimestre y Denominaciones Religiosas.

RI40: Obtener cantidad por Mes, DPA, Organismo, e Ind_Planteamientos de la población.

RI41: Obtener cantidad y cantidad de mujeres por Semestre, DPA, Organismo y Secciones.

RI42: Obtener cantidad de plantillas aprobadas, cantidad de plantillas cubiertas por hombres, cantidad de plantillas cubiertas por mujeres, cantidad de plantillas vacantes, promedio trabajadores, salario devengado y salario por Trimestre, DPA, Organismo y Categoría Ocupacional.

2.4.2 Requisitos funcionales.

Los requerimientos funcionales son capacidades o condiciones que el sistema debe cumplir. En la realización de los casos de uso del negocio, se obtienen las actividades que serán objeto de automatización. Estas actividades no son exactamente los requerimientos funcionales, pero sí son el punto de partida para identificar qué debe hacer el sistema. Los requerimientos funcionales no alteran la funcionalidad del producto, esto quiere decir que los mismo, se mantienen invariables sin importarle con que propiedades o cualidades se relacionen.

Los requisitos funcionales de la solución del sistema están soportados por:

- Asegurar la limpieza de los datos.
- Asegurar la carga de los datos desde distintas fuentes.
- Autenticar usuario.
- Realizar los reportes
- Realizar vistas de análisis
- Asegurar la extracción de los datos

2.4.3 Requisitos no funcionales

Los requerimientos no funcionales son propiedades o cualidades que el producto debe cumplir, es el que hace al producto atractivo, usable, rápido o confiable. A continuación se especifican algunos de los requisitos no funcionales propuestos para el mercado de datos.

Para consultar la información de manera generalizada consultar el documento Especificación de Requisitos.

➤ **Usabilidad**

- Cumplir con las pautas de diseño de la interfaces.
- Mostrar los mensajes, títulos y demás textos que aparezcan en la interfaz del sistema en idioma español.

➤ **Fiabilidad**

- Garantizar la persistencia de la información.

➤ **Eficiencia**

- Para garantizar la persistencia de la información se realizará un respaldo total de los datos del mercado de datos con una frecuencia estimada.

➤ **Restricciones de diseño**

- Utilizar el lenguaje de programación definido durante la investigación.

- Lograr que los elementos definidos en el mercado tengan una estructura homogénea.
- Las estructuras del mercado de datos se nombrarán de una manera estándar teniendo en cuenta el tipo de estructura que se maneje.

➤ **Requisitos para la documentación de usuarios y ayuda del sistema**

- Confeccionar manual de usuario
- El sistema debe estar acompañado de un documento que guiará la ejecución del usuario teniendo en cuenta cada funcionalidad.

➤ **Requisitos para la documentación de usuarios en línea y ayuda del sistema.**

- Se dispondrá de una guía de ayuda sobre componentes del sistema, además de una gran información sobre herramientas ETL, presentadas en el sitio de la comunidad de Pentaho.

➤ **Componentes Comprados**

- La herramienta utilizada están licenciada bajo GPL, y la misma es Open Source y totalmente gratis. Es ampliamente reutilizable y extensible.

➤ **Interfaz**

- Los requerimientos de interfaz de usuario se centran en la presentación de la información de cara al cliente.
- Los reportes estadísticos deben contar con una interfaz simple que facilite la interacción usuario-aplicación.
- Las interfaces de salida no serán cargadas con información innecesaria.

➤ **Interfaces de usuario**

- Los reportes estadísticos deben contar con una interfaz simple que facilite la interacción usuario-aplicación.

- Las interfaces de salida no serán cargadas con información innecesaria.

➤ **Interfaces de hardware**

- En el proceso de integración es necesario la utilización de una memoria mínima de 1 GigaByte (GB) para el proceso de transformación.
- Se debe contar de un área de almacenamiento intermedio de 80 GB.
- Para la visualización y la inteligencia de negocio se necesita una memoria de 1 GB.

➤ **Interfaces de software**

- Instalar en las estaciones de trabajo el software necesario para el correcto funcionamiento del sistema
- Debe existir un navegador asociado al sistema operativo que se escoja para lograr que las interfaces web de las tablas de salida puedan visualizarse.

➤ **Interfaces de Comunicación**

- La comunicación entre la base de datos de integración y el almacén de datos es a través del protocolo TCP/IP.
- El sistema necesita estar conectado directamente a un dispositivo de red.

➤ **Requisitos de Licencia**

- Las herramientas a utilizar se encuentran bajo licencia GPL.

2.5 Casos de uso del sistema

Los casos de uso ayudan a describir qué es lo que se debe hacer, están enfocados al punto de vista del usuario. Es decir, describen un uso y cómo este interactúa con el usuario. En este documento se encuentra parte de la documentación para mayor información consultar el modelo de casos de uso.

2.5.1 Diagrama de Casos de Uso

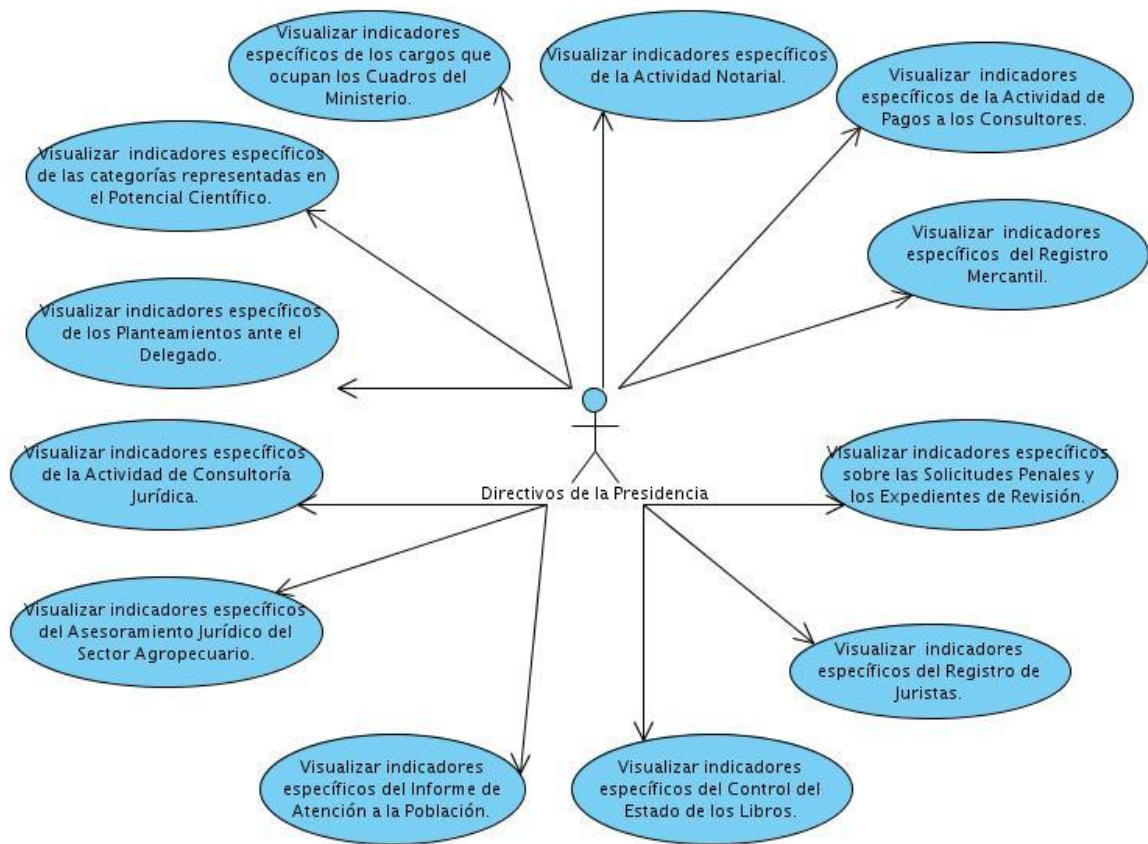


Ilustración 1: Diagrama de Caso de Uso

Diagrama de Casos de Uso Funcionales

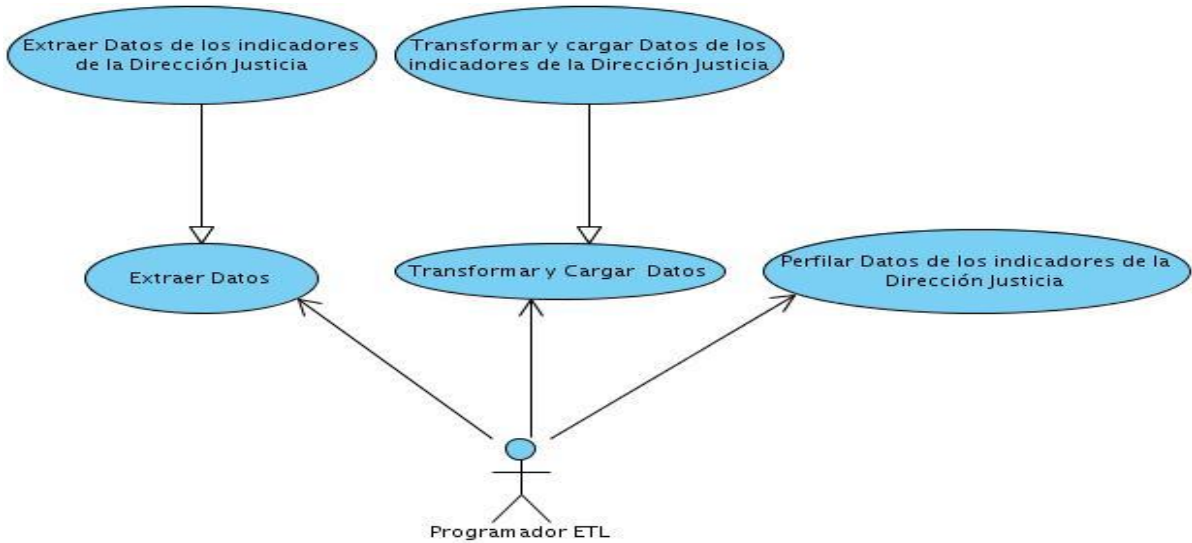


Ilustración 2 : Diagrama Caso de Uso Funcional

Especificación de los Casos de Uso

2.5.2 Casos de Uso de Información

Caso de Uso:	Visualizar indicadores específicos de los cargos que ocupan los Cuadros del Ministerio.
Actores:	Rol o Grupo (Usuario).
Resumen:	
Precondiciones:	Compleitud del almacén. Carga de los datos.
Referencias	
Prioridad	
Flujo Normal de Eventos	
Sección	
Acción del Actor	Respuesta del Sistema
Entra al sistema.	Muestra los reportes referentes a los indicadores de los cargos.
Selecciona reporte (Visualizar	Muestra tipo de reporte (Visualizar

indicadores específicos de los cargos que ocupan los Cuadros del Ministerio.)	indicadores específicos de los cargos que ocupan los Cuadros del Ministerio.)
Prototipo de Interfaz	
Poscondiciones	Disponibilidad de opciones (cruces de variables) de reportes relacionados con el hecho Cuadros del Ministerio

Tabla 1: Visualizar indicadores específicos de los cargos que ocupan los Cuadros del Ministerio.

2.5.3 Casos de Uso Funcionales

Caso de Uso:	Extraer datos de los indicadores de la dirección justicia.	
Actores:	Programador ETL.	
Resumen:		
Precondiciones:	Disponibilidad de las fuentes.	
Referencias		
Prioridad		
Flujo Normal de Eventos		
Sección		
Acción del Actor	Respuesta del Sistema	
1. Realiza la conexión a los indicadores de la dirección justicia	1.1 Responde a la solicitud de conexión.	
2. Selecciona estructuras o archivos a extraer.		
3. Realiza la extracción.		
Prototipo de Interfaz		
Flujos Alternos		
Acción del Actor	Respuesta del Sistema	
	1.1 No responde a solicitud de conexión.	

Notifica el error a administrador de los indicadores de la dirección justicia	
2. Si hay control de cambios, verifica si hay modificaciones. En caso de que no haya, va al paso 3 de flujos normal. En caso de que haya, va al paso 2 del flujo normal.	
Prototipo de Interfaz	
Poscondiciones	Datos disponibles para transformar.

Tabla 2: Extraer datos de los indicadores de la dirección justicia.

2.6 Diseño de la estructura física de datos

El modelo físico del mercado de datos dirección justicia cuenta con 42 tablas de hecho y 37 dimensiones con 4 generales y 33 específicas. Para el modelado de todas las tablas se deben seguir los estándares y codificaciones definidas por el departamento y el cliente. Para el modelado del MD se siguió el esquema en estrella.

2.7 Matriz bus

Es una matriz que especifica la relación entre los hechos y las dimensiones, donde las columnas contienen todas las dimensiones y las filas las tablas de hechos, la intersección de una fila con una columna especifica si hay relación entre una tabla de hechos y una dimensión.

H/D	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1												x								
2				x								x								
3											x	x								

D-18:dim_ind_notas_marginales

D-19:dim_ind_planteamiento

D-20:dim_ind_planteamientos_poblacion

D-21:dim_ind_solicitud_certificaciones

D-22:dim_indicadores_generales

D-23:dim_ingresos

D-24:dim_inscripciones_defuncion

D-25:dim_inscripciones_ind_expediente

D-26:dim_inscripciones_matrimonio

D-27:dim_inscripciones_nacimiento

D-28:dim_recaudacion

D-29:dim_registro

D-30:dim_revision_penal

D-31:dim_salario

D-32:dim_secciones

D-33:dim_sector

D-34:dim_situacion_financiera

D-35:dim_dpa

D-36:dim_temporal_mes

D-37:dim_organismo1

Hechos (H)-X:

H-1: hech_actividad_consultoria_juridica_asesores_propios

H-2: hech_actividad_consultoria_juridica_asuntos_tramitados

H-3: hech_actividad_consultoria_juridica_contrato_servicios

H-4: hech_actividad_consultoria_juridica_indicadores_generales

H-5: hech_actividad_consultoria_juridica_situacion_financiera

H-6: hech_actividad_notarial

H-7: hech_actividad_pago_consultores_ingreso

H-8: hech_actividad_pago_consultores_salario
H-9: hech_antecedentes_penales
H-10: hech_asesoramiento_juridico_sector_agropecuario
H-11: hech_asociaciones
H-12: hech_atencion_poblacion
H-13: hech_control_certificaciones_expedidas_provincia
H-14: hech_control_notas_enviadas_recibidas
H-15: hech_control_ordenamiento_inmobiliario_entidad
H-16: hech_control_ordenamiento_inmobiliario_plan_municipio
H-17: hech_cuadro1
H-18: hech_ind_municipio_tarjeta_menor
H-19: hech_indicador_municipio_control_libros
H-20: hech_inscripciones_practicadas_organismo
H-21: hech_ordenamiento_inmobiliario_municipio
H-22: hech_planteamiento_ante_delegado
H-23: hech_potencial_cientifico
H-24: hech_recursos_humanos
H-25: hech_registro_civil_certificaciones_expedidas
H-26: hech_registro_civil_ind_actos
H-27: hech_registro_civil_ind_busquedas
H-28: hech_registro_civil_ind_expediente
H-29: hech_registro_civil_ind_notas_marginales
H-30: hech_registro_civil_ind_solicitud_certificaciones
H-31: hech_registro_civil_inscripciones_defuncion
H-32: hech_registro_civil_inscripciones_matrimonio
H-33: hech_registro_civil_inscripciones_nacimiento
H-34: hech_registro_del_estado_civil
H-35: hech_registro_juristas

H-36: hech_registro_mercantil

H-37: hech_registros_propiedad

H-38: hech_solicitud_penal_expedientes_revision

H-39: hech_muestreo_certificaciones_expedidas

H-40: hech_organizaciones_politicas

2.8 Modelo de datos dimensional

Un conjunto de dimensiones, medidas, y hechos identificados de la dirección justicia conforman el modelo dimensional. A continuación se muestra un ejemplo de las tablas descriptivas de dimensiones relacionadas con el hecho y las medidas correspondientes a la tabla de hecho cuadro. En la Ilustración 3, se muestra una vista del modelo de datos, en el cual se ve la relación que existe entre las dimensiones y la tabla de hechos, así como una pequeña descripción.

Las restantes descripciones se encuentran en el documento Especificación del Modelo de datos Dimensional.

2.8.1 Tabla descriptiva de la Dimensión

Nombre del atributo	Descripción	Tipo de Dato	Política de actualización
pk_dim_cargo1_id	Llave primaria de la dimensión.	Integer(10)	trimestral
cargo1_nombre	Almacena el nombre del cargo de cada cuadro.	Varchar(255)	trimestral

Tabla 4: Descripción de la dimensión cargo.

Nombre del atributo	Descripción	Tipo de Dato	Política de actualización
dim_dpa_id	Llave primaria de la dimensión.	Integer(10)	

municipio_codigo	Almacena el código del municipio	Integer(10)	
municipio_nombre	Almacena el nombre del municipio	Varchar(30)	
municipio_descripcion	Almacena el código de la provincia	Varchar(45)	
acentamiento	Almacena el nombre de la provincia	Varchar(45)	

Tabla 5: Descripción de la dimensión DPA

Nombre del atributo	Descripción	Tipo de Dato	Política de actualización
dim_organismo1_id	Llave primaria de la dimensión.	Integer(10)	
organismo1_nombre	Almacena el nombre del organismo.	Varchar(255)	

Tabla 6: Descripción de la dimensión organismo

Nombre del atributo	Descripción	Tipo de Dato	Política de actualización
dim_temporal_semes tre_id	Llave primaria de la dimensión.	Integer(10)	
anno_numero	Almacena el número oficial del año de la información.	Integer (10)	
anno_codigo	Almacena el código oficial del año de la información.	Integer(10)	
anno_nombre	Almacena el nombre oficial del año.	Varchar(25)	

semestre_numero	Almacena el número oficial del semestre de la información.	Integer (10)	
semestre_codigo	Almacena el código oficial del semestre de la información.	Integer (10)	
semestre_nombre	Almacena el nombre del semestre de la información	Varchar(25)	

Tabla 7: Descripción de la dimensión temporal semestre

2.8.2 Medidas presente en la tabla de hecho:

hech_cuadro1

- cantidad_cubierto
- cantidad_aprobado
- cantidad_vacante

hech_asociaciones

- aprobadas_canceladas
- aprobadas_funcionan
- aprobadas

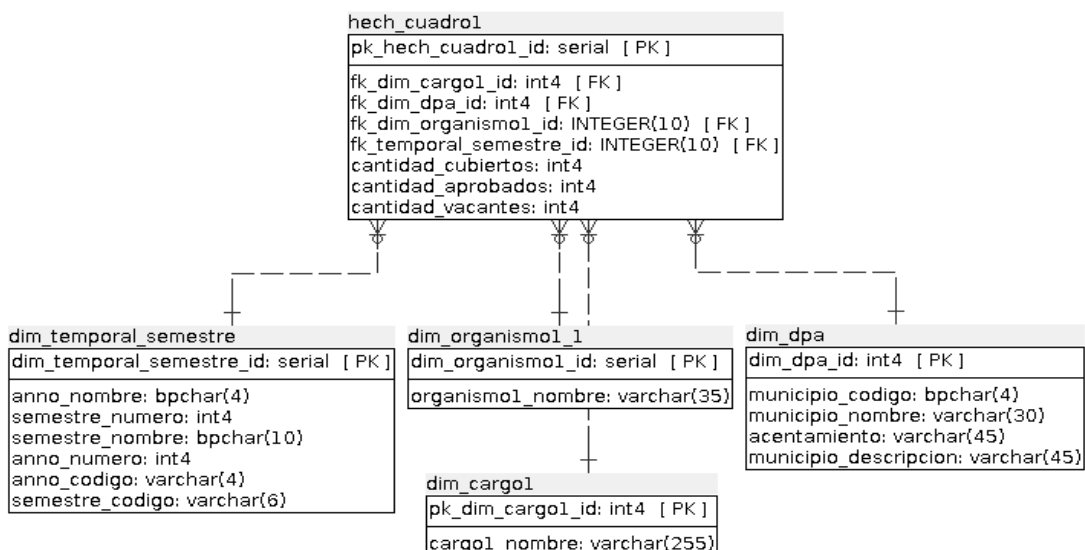


Ilustración 3 : hech_cuadro1

Las tablas de hechos son las que almacenan las medidas numéricas. En esta tabla es donde va a residir, toda la información existente, que servirá como fuente principal para la realización de las estructuras que soporten los reportes más comunes, encontrándose en su totalidad de características en el documento Especificación del Modelo de datos Dimensional. Anteriormente se encuentra la ilustración 3 presentando un ejemplo descriptivo.

2.9 Política de respaldo y recuperación

Las políticas de seguridad y respaldo que se utilizarán en el MD justicia es sencilla, pero a la vez sólida, por ello se miden 3 puntos esenciales:

- **Periodicidad de las salvas:** Mensualmente se realizan las salvas de toda la información contenida en la base de datos. La organización lo tiene definido de esta manera, certificando en todo momento la existencia de una copia escrita de la información que está presente en el servidor.
- **Tablas involucradas:** Las tablas que se involucran en la realización son las

tablas de hecho y las tablas de dimensiones apoyan el proceso de ETL.

- **Backups existentes:** Actualmente no existen backups en esta área.

Conclusiones

Al concluir este capítulo quedó definido el negocio y sus reglas, los temas de análisis, la necesidad de los usuarios así como los requisitos; se definieron las dimensiones, métricas y hechos asociados, se diseñó el modelo de caso de uso del sistema donde se representa la relación de los casos de usos y los actores de la aplicación, realizándose después el modelado de datos siendo estos necesarios para llevar a cabo la propuesta de solución.

3 Capítulo 3: Implementación del mercado de datos de la dirección de justicia de la Administración Provincial de Artemisa

En este capítulo se realizarán todas las transformaciones necesarias como extraer, transformar y cargar, además de la implementación de los visualización de los datos, así como almacenamiento de los datos en las tablas correspondientes de acuerdo al modelo de datos definido en el capítulo anterior.

3.1 Implementación de los subsistemas de integración

Es el proceso de extracción es donde se obtiene toda la información de las distintas fuentes tanto internas como externas. Se cargan los datos de los archivos excel y así adaptarlos al modelo relacional que se ha establecido.

Se seleccionan los campos necesarios, luego estos datos se transforman, limpian y estandarizan porque los mismos pueden ser incoherentes, tener errores o estar incompletos. Con esto se busca obtener datos precisos, completos, consistentes, interpretables y lo más accesibles posibles. Luego como paso final se realiza la carga de datos hacia el MD, organizados y actualizados, teniendo como salida la tabla correspondiente en la base de datos.

A continuación se muestra un ejemplo de las transformaciones realizadas teniendo como entrada un excel.

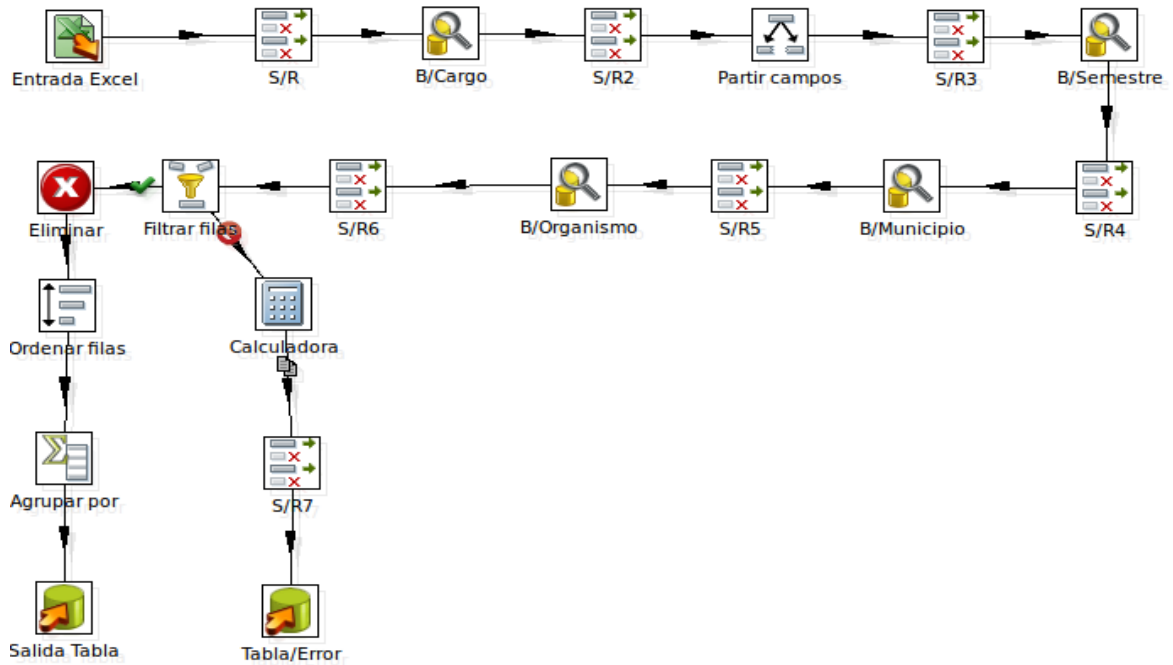


Ilustración 4: Extracción, transformación y carga del hecho Cuadro, entrada en excel.

Como primer paso para realizar la transformación es la extracción de los datos fuentes, en este caso la entrada se hace a través de un excel, obteniendo y comparando valores que a su vez van a ser usados para la realización de los cálculos correspondientes para llenar las medidas de la tabla de hecho registro del estado civil, validando en la salida los datos incorrectos y pasando a una tabla de error en caso de que exista.

A continuación se muestra un ejemplo de las transformaciones realizadas teniendo como entrada una tabla.

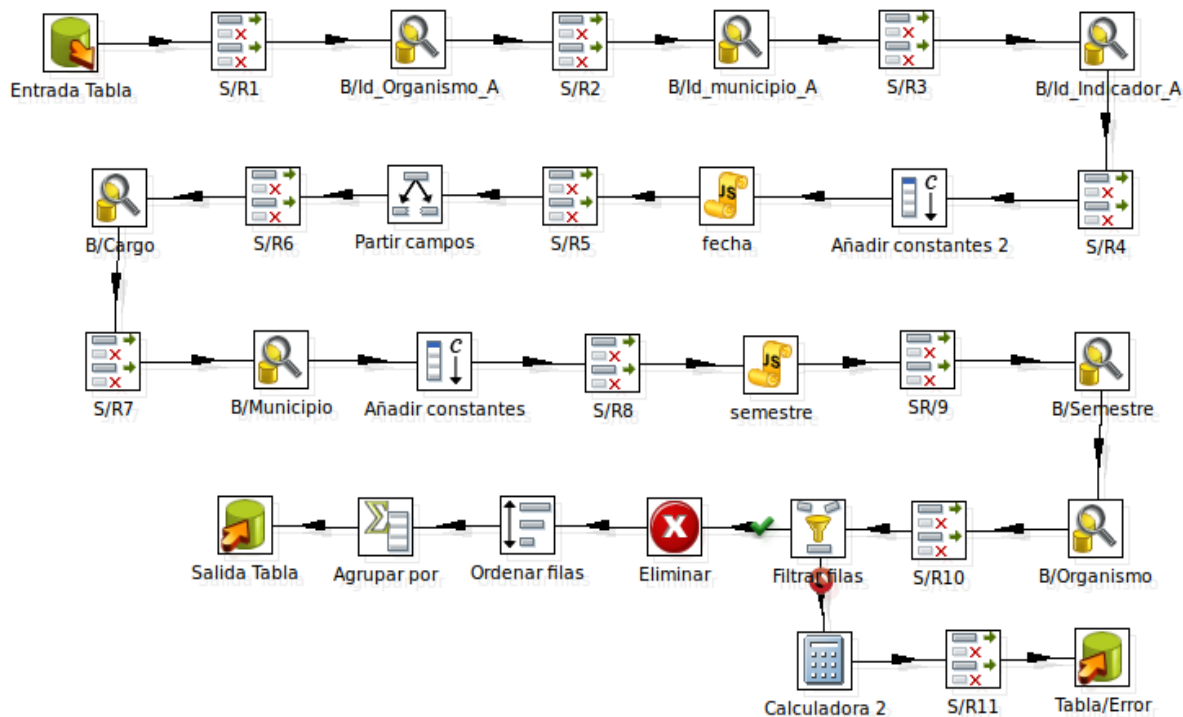


Ilustración 5: Extracción, transformación y carga del hecho Cuadro, entrada en tabla.

Los datos de la tabla son obtenidos a través del campo, entrada tabla luego se, realiza una secuencia de pasos continuos para validar que no contengan valores nulos y que los id buscados sean válidos. Se utiliza el componente añadir constantes para añadir las constantes que no vienen directamente de la fuente y luego se realiza la validación java script para la suma de los campos que pasarán a la tabla mediante la salida de la tabla validando que los datos en una misma fecha y en este caso de un mismo cargo no se dupliquen si se vuelven a insertar. Los datos correctos van a la tabla de hecho pero si existe algún error manda la información a la tabla de error, permitiendo conocer al usuario el campo donde se encuentra dicho error.

3.2 Implementación de los trabajos

Un trabajo o mayormente conocido como Job (por sus siglas en inglés) consisten en la realización de un conjunto de tareas para realizar determinadas acciones. Se pueden realizar un grupo de transformaciones dependiendo de la secuencia a seguir, una transformación no se empieza a ejecutar si la anterior no ha terminado. Para la correcta realización de un Job, se debe tener bien definido cuales son las dimensiones estáticas y cuales no, pues en un Job sólo se cargan las dimensiones que pueden tomar valores nuevos o cambiar los que tenían anteriormente.

Debido a que la información origen es muy extensa, se crearon tres trabajos, por área temporal uno que almacena los mensuales, otro los semestrales y por último los trimestrales y una vez estando allí, los datos son preparados y llevados al lugar destino, que es el MD.

En las siguientes ilustraciones se muestran los trabajos que se realizaron en este proceso.

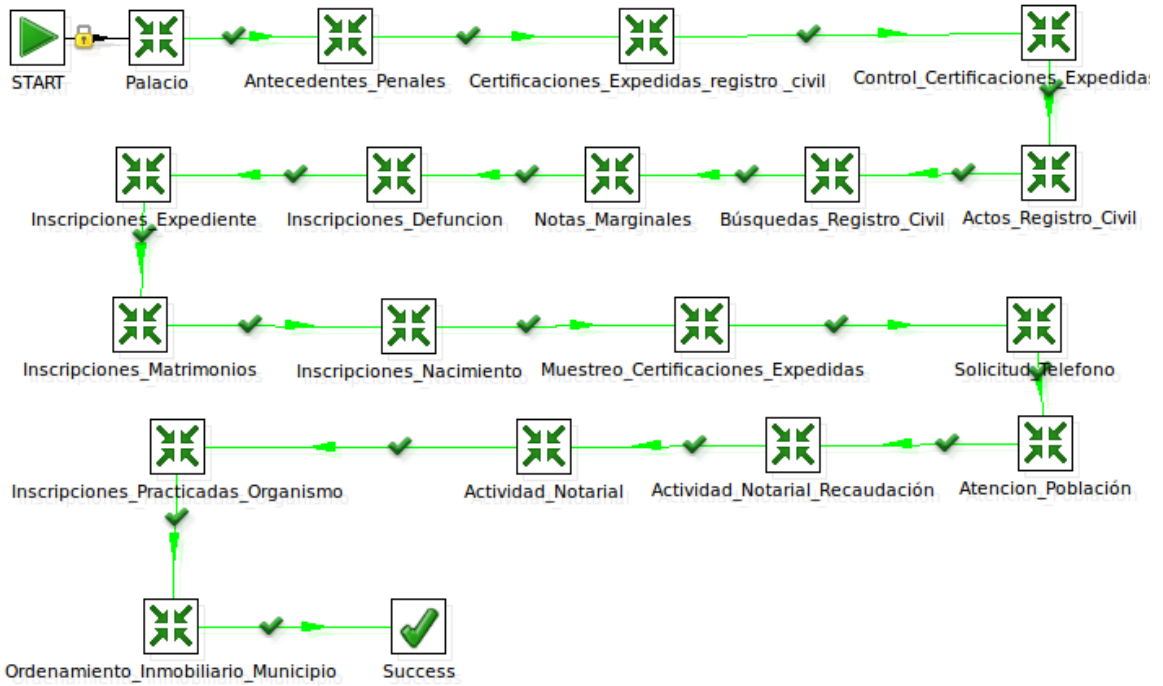


Ilustración 6: Trabajo Mensual

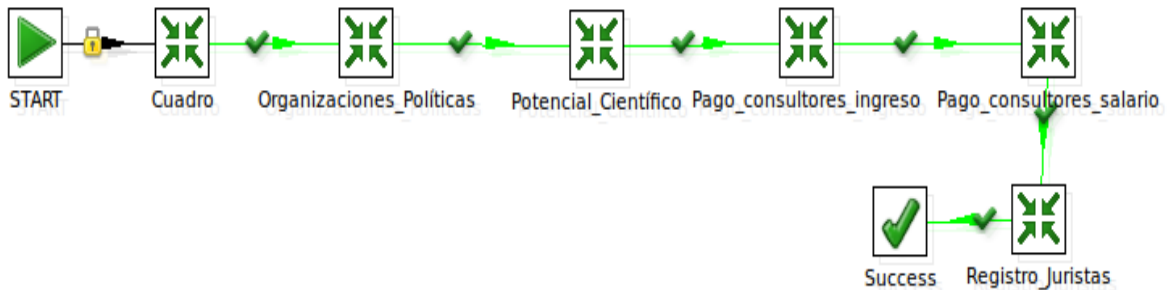


Ilustración 7: Trabajo Semestral

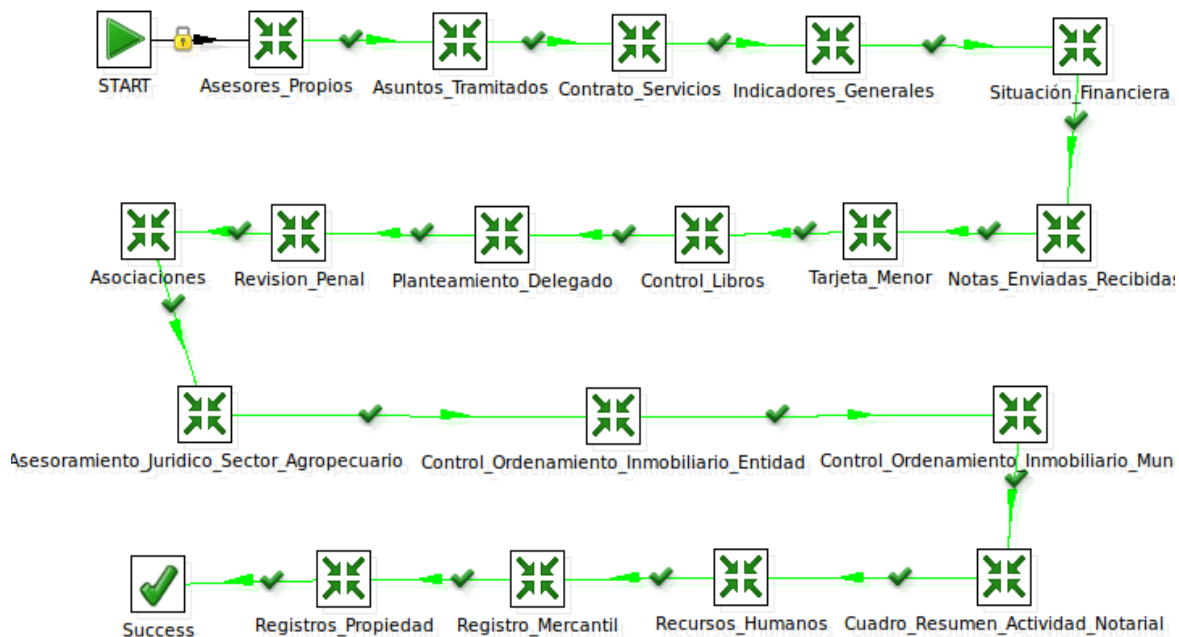


Ilustración 8: Trabajo Trimestral

3.3 Implementación de los subsistemas de visualización

Es el proceso de inteligencia de negocio encargado de mostrar el resultado final del ciclo de desarrollo de un mercado de datos a través de la confección de los cubos y reportes de información que pueden ser de varios tipos.

3.3.1 Diseño de cubos OLAP

El diseño de los cubos multidimensionales se realizan mediante la herramienta Schema Workbench, definiendo tantos cubos como tablas de hechos, que abarcarán todas las dimensiones descritas anteriormente. Por cada tabla de hecho se define un cubo que carga la información.

En la presente investigación se diseñaron 42 cubos multidimensionales, especificando en los cubos las dimensiones y características que se corresponden

con estas tablas. A continuación se muestra en la ilustración 11 un ejemplo de un cubo dimensional, el cual muestra la información de la tabla de hecho Cuadro.

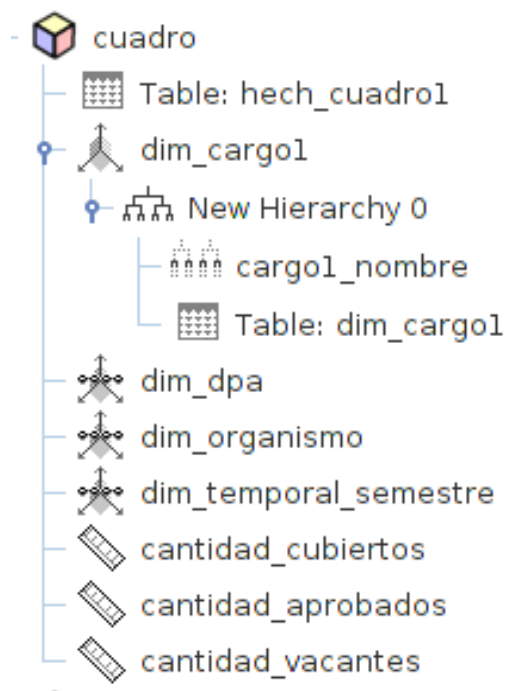


Ilustración 9: Diseño de los Cubos OLAP

3.3.2 Arquitectura de información

A continuación se muestra en la ilustración 12 la estructura organizativa que conforma el mapa de navegación correspondiente al mercado de datos de la dirección justicia para la visualización de los datos.

La cual está compuesta por un Nivel Principal, denominado MD_Dirección_Justicia, la cual contiene 11 Área de Análisis, cada una con su Libro de Trabajo y dentro las vistas asociadas.

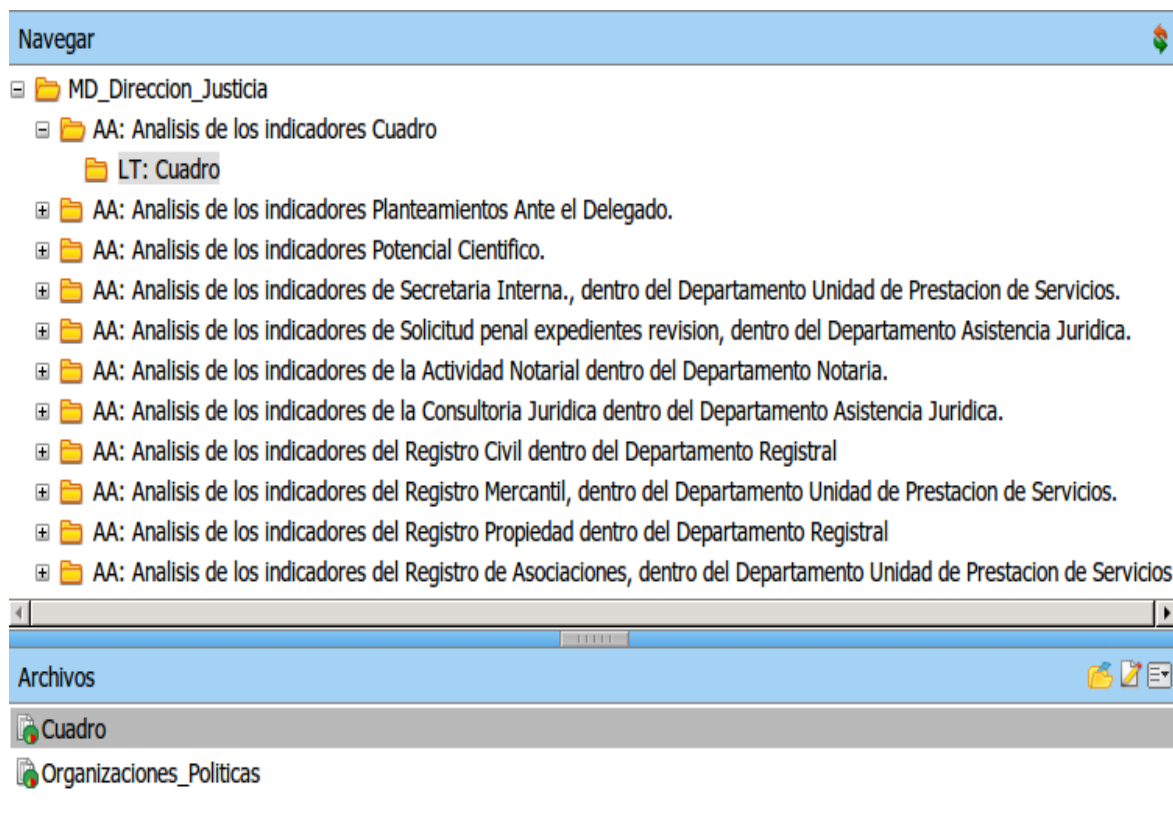


Ilustración 10: Mapa de Navegación

3.3.3 Vistas de análisis

Las vistas de análisis no son más que las vistas de cierta información que permiten al usuario realizar el análisis a los indicadores presentes en ella, se cargan en el Pentaho BI-Server tras haber realizado la publicación de los cubos OLAP, se presenta en la ilustración 13 un ejemplo de vista de análisis correspondiente al control de plantilla de cuadros del sistema del ministerio. En la ilustración 14 se presenta un gráfico generado a partir de la vista de análisis correspondiente al control de plantilla de cuadros del sistema del ministerio representando a través de

Capítulo 3. Implementación del mercado de datos

los colores rojo, azul y verde la cantidad de cubiertos, de aprobados, y vacantes por las dimensiones especificadas.

				Medidas		
Municipios	Semestre	Organismos	Cargos	● Cubiertos	● Aprobados	● Vacantes
▢ Todos los Municipios	▢ Años	▢ Todos los Organismos	▢ Todos los Cargos	11	9	9
Artemisa	▢ Años	▢ Todos los Organismos	▢ Todos los Cargos	11	9	9
		▢ 2011	▢ Todos los Organismos	▢ Todos los Cargos	11	9
	S1	▢ Todos los Organismos	▢ Todos los Cargos	11	9	9
		▢ MINJUS	▢ Todos los Cargos	11	9	9
		Ministerio de Justicia	▢ Todos los Cargos	11	9	9
			Directores Municipales	4	2	2
			Jefes de Departamentos	0	0	0
Subdirectores	7		7	7		

Ilustración 11: Vista de análisis correspondiente al Control de plantilla de Cuadros del sistema del ministerio.

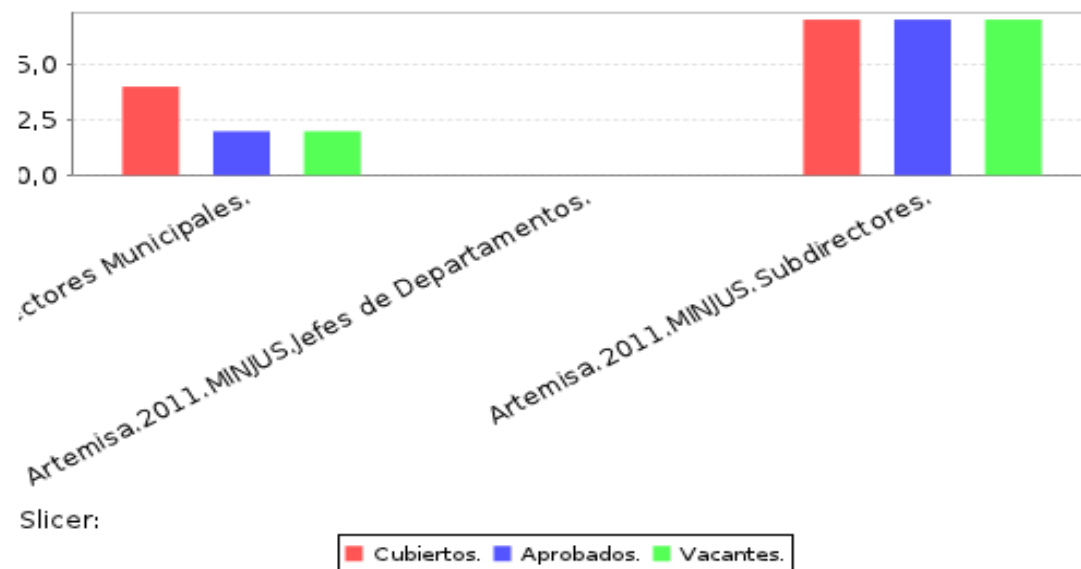


Ilustración 12: Gráfico correspondiente al control de plantilla de cuadros del sistema del ministerio

Conclusión

Con la realización de este capítulo quedaron definidos los elementos para el desarrollo del MD. Se realizó la implementación del modelo de datos físico definiéndose dos esquemas: dimensiones que recoge un conjunto de dimensiones general y mart_Justicia con la información característica de esta dirección, para una mejor estructuración de la información. Se realizó el proceso de ETL, extrayendo los datos de diferentes fuentes, para luego realizar la carga de los datos a la base de datos. Luego de tener los datos cargados se transforman en información valiosa para el cliente a través de la visualización de los datos logrando así una mejor perspectiva de análisis de los datos.

4 Capítulo 4: Validación del mercado de datos de la dirección de justicia de la administración provincial de Artemisa

En el presente capítulo se aplican diferentes tipos de pruebas no automatizadas, tanto a la aplicación como a la documentación para validar el funcionamiento del mercado de datos. Durante el transcurso de este capítulo se validan las entradas de datos, verificando que los resultados obtenidos se correspondan con los resultados esperados.

4.1 Calidad de software

La calidad es un elemento esencial en todos los aspectos de la vida incluyendo el desarrollo de software. El interés por la calidad crece de forma continua, a medida que los clientes se vuelven más selectivos y comienzan a rechazar productos poco fiables o que realmente no dan respuesta a sus necesidades.

Por tanto los desarrolladores de software necesitan que los procesos de desarrollo y los productos terminados sean evaluados para tener una garantía de que el software tiene alta calidad. Por ese motivo surgen modelos, técnicas, patrones y normas de calidad que establecen cómo hacer un software con calidad desde los procesos que lo inician hasta el producto terminado.

La calidad de software se pudiera definir como la capacidad que tiene un software para adherirse a un conjunto de requisitos, especificaciones, estándares, normas o procedimientos que deben cumplirse durante el proceso de desarrollo, y que permitan verificar que el producto cumple con los requisitos explícitos establecidos y las expectativas del cliente.

La importancia de la calidad del software radica en que el software es un producto mental, no es restringido por las leyes de la física o por los límites de los procesos de fabricación, es algo abstracto y su calidad también lo es. Sirve de plataforma para desarrollar en la organización, una serie de actividades, procesos y

procedimientos, encaminados a lograr que las características del producto cumplan con los requisitos del cliente u organización, en pocas palabras que sean de calidad. (Lamancha, 2006)

4.1.1 Elementos de calidad de software

Mantenibilidad: Capacidad del producto de software de ser modificado. Las modificaciones pueden incluir las correcciones, mejoras o adaptaciones del software a cambios en el ambiente, así como en los requisitos y las especificaciones funcionales.

Funcionalidad: Es la capacidad del software para proporcionar funciones que satisfacen las necesidades declaradas e implícitas cuando el software se usa bajo las condiciones especificadas.

Portabilidad: Capacidad de producto de software de ser transferido de un ambiente a otro.

Confiabilidad: La capacidad del producto de software para mantener un nivel de ejecución especificado cuando se usa bajo las condiciones especificadas.

Eficiencia: Capacidad del producto de software para proporcionar una ejecución o desempeño apropiado, en relación con la cantidad de recursos utilizados usados, bajo condiciones establecidas.

Usabilidad: Capacidad del producto de software de ser comprendido, aprendido, utilizado y de ser atractivo para el usuario, cuando se utilice bajo las condiciones especificadas.

Estas características se miden mediante la aplicación de diversas pruebas que son aplicadas a medida que el producto es construido.

4.2 Pruebas de software

Las pruebas del software constituyen un elemento fundamental que permiten verificar y revelar la calidad de un producto con el objetivo de verificar si el sistema

cumple con lo esperado por el cliente. Para determinar el nivel de calidad se deben efectuar pruebas que permitan comprobar el grado de cumplimiento de las especificaciones iniciales del sistema. Para la validación del MD se pueden aplicar los diferentes tipos de pruebas que se especifican a continuación:

Prueba unitaria: Es el proceso de probar los componentes individuales de la solución. El propósito es identificar diferencias entre la especificación de los artefactos y el comportamiento real de cada módulo.

Prueba de integración: Es el proceso en el cual los componentes son agregados para crear componentes más grandes. Es la prueba realizada para mostrar que aunque los componentes hayan pasado satisfactoriamente las pruebas de unidad, la integración de los componentes es incorrecta.

Prueba de sistema: Se refiere al comportamiento del sistema integrado. Durante la etapa de las pruebas unitarias y de integración deben haberse identificado la mayoría de las no conformidades. La prueba de sistema se aplica generalmente para probar los requerimientos no funcionales de la solución.

Pruebas de aceptación: Se realizan para probar que el sistema cumpla con los requerimientos especificados por el cliente.

4.3 Validación y prueba

Para el desarrollo de un producto con la calidad y aceptación requerida y para que este no sea entregado al cliente con errores es necesario realizarle al producto un conjunto de pruebas que se realizan a través del diseño de los casos de prueba.

Para el mercado de datos Dirección Justicia se diseñaron 22 casos de prueba, uno por cada caso de uso, a continuación se muestra el diseño de caso de prueba correspondiente al caso de uso “Visualizar indicadores específicos de los cuadros del Ministerio.” (Ver expediente de proyecto correspondiente al MD dirección

justicia).

Escenario	Descripción	Perfiles de Análisis	Indicadores a medir	Respuesta del sistema	Flujo central	Resultado de la Prueba
EC 1.1: Cuadro	Permite visualizar tipos de reportes con las variables presentes en el mismo.	Cargos Organismo Municipio Semestre	cantidad_cu biertos cantidad_ap robados cantidad_van cantes	Se muestran los reportes referentes a los indicadores de los cargos que ocupan los Cuadros del Ministerio	<ul style="list-style-type: none"> ➤ Se abre la aplicación. ➤ Se autentifica. ➤ Se entra al sistema. ➤ En el menú lateral del navegador se selecciona el Departamento Justicia ➤ Se selecciona el área de análisis Análisis de los indicadores Cuadro. ➤ Se selecciona el libro de trabajo dentro la vista de análisis. ➤ En el área de trabajo se visualiza la tabla correspondiente a la vista. 	Satisfactorio

Tabla 8 : Caso de prueba basado en caso de uso

Conclusiones

En este capítulo se abordaron los temas referentes a la calidad del software, los elementos que la componen y los diferentes tipos de prueba. Se realizaron los casos de pruebas aplicados al mercado de datos dirección justicia, con un total de 22 casos de prueba arrojando resultados satisfactorios, evidenciando un correcto funcionamiento del MD.

Conclusiones generales

Con la elaboración de este trabajo se logró como resultado el mercado de datos para la dirección justicia de la administración provincial de artemisa, cumpliendo con los objetivos planteados y arribando a las siguientes conclusiones:

- Los MD constituyen una robusta solución para el almacenamiento de los datos relacionados con los indicadores de la dirección justicia.
- Se realizó el modelo de datos, identificando 37 dimensiones y 42 hechos para la integración de los datos, permitiendo el almacenamiento con altos índices de seguridad de los datos.
- Se logró implementar las estructuras dimensionales utilizadas para la solución, y con el uso de las mismas se agilizará el proceso de toma de decisiones para la dirección justicia.
- Las 22 pruebas realizadas a través de los casos de pruebas basado en casos de uso permitieron validar la solución propuesta obteniendo resultados satisfactorios en cada una de ellas.

Recomendaciones

- Realizar el despliegue del mercado de datos en la dirección justicia de la administración provincial de Artemisa.
- Una vez realizado el despliegue atender a la explotación de la aplicación.

Referencias bibliográfica

1. Herrera, Cristhian. Todo lo que querias saber sobre Data Warehouse (I). [En línea] 3 de Febrero de 2010. <http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=datawarehouse#2.4.Character AdsticascdelDatawarehouse|outline>.
2. Inmon, William H. Building the Data Warehouse. EUA : Wiley Publishing Inc, 2005.
3. Kimball. The Data Warehouse Toolkit. EUA : Wiley Publishing Inc.2002.
4. Lamancha, Beatriz Pérez. Tesis de Maestría en Informática: Proceso de Testing Funcional Independiente. Montevideo, Uruguay: s.n., 2006.
5. Laudon Kenneth, Laundon Jane. Sistemas de información gerencial Google Libros .ed. Última actualización: 2008.().Disponible: en <http://books.google.com.cu/books>.
6. Luján Mora, S. Diseño de almacenes de datos con UML, 2004. 27. p.
7. Vega Torres, L. La inteligencia de negocio. Su implementación mediante la pltaforma Pentaho., CITMATEL, 2008. 22.
8. Velasco JJ. IBM construye el mayor almacén de datos del mundo. Última actualización: 2011. (). Disponible en: <http://alt1040.com/2011/08/ibm-construye-mayor-almacen-datos-mundo>.
9. Wolff, Carmen. [2000] Implementando un DataWarehouse. Revista Ingeniería Informática: revista electrónica del DIICC.

Bibliografías

- Análisis de información. Disponible en: <http://www.iic.uam.es/es/soluciones-y-servicios/analisis-de-informacion>.
- Business Intelligence - DataMart Disponible en: <http://www.swgreenhouse.com/Productos/Hi-Spins/DataMart.html>.
- Data warehouse, la clave en el desarrollo de la industria del Retail | GestioPolis. Disponible en: <http://www.gestiopolis.com/canales6/ger/desarrollo-industrial-data-warehouse.htm>.
- Date J. C.. Introducción a los sistemas de bases. Google Libros Disponible en: <http://books.google.com/cu/books?almacenesdedatosoperacionales>.
- Esquema copo de nieve (snowflake schema). Disponible en: http://etl-tools.info/es/bi/almacenedatos_esquema-copo-de-nieve.htm.
- ETL-Tools.Info.."Esquema de constelación de hechos," (2006), Disponible en: http://etl-tools.info/es/bi/almacenedatos_esquema-constelacion.htm.
- Evelia, C.C.M. "DataWarehouse".(2009) Disponible en: <http://hp.fciencias.unam.mx/~alg/bd/dwh.pdf>.
- González O Darián. Diseño e Implementación de un Almacén de datos Operacionales para la Corporación CIMEX. Trabajo de Diploma, Universidad de las Ciencias Informáticas, 2009.
- Herrera, Cristhian. Todo lo que querías saber sobre Data Warehouse (I). [En línea] 3 de Febrero de 2010.

[http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=datawarehouseuse#2.4.Character AdsticascadelDatawarehouse|outline](http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=datawarehouseuse#2.4.Character%20AdsticascadelDatawarehouse|outline).

- Inmon, William H. Building the Data Warehouse. EUA : Wiley Publishing Inc, 2005.
- Kimball. The Data Warehouse Toolkit. EUA : Wiley Publishing Inc.2002.
- Lamancha, Beatriz Pérez. Tesis de Maestría en Informática: Proceso de Testing Funcional Independiente. Montevideo, Uruguay: s.n., 2006.
- Laudon Kenneth, Laudon Jane. Sistemas de información gerencial Google Libros .ed. Última actualización: 2008.(.).Disponible: en <http://books.google.com/cu/books>.
- Luján Mora, S. Diseño de almacenes de datos con UML, 2004. 27. p.
- Mercados De datos Data Mart. Disponible en:<http://www.mitecnologico.com/Main/MercadosDedatosDataMart>.
- Pentaho Data Integration. Disponible en: <http://www.inqbation.com/pentaho-data-integration/>
- Query | Define Query at Dictionary.com. Última actualización: 2011b. Disponible en: <http://dictionary.reference.com/browse/query>.
- Sarduy Dominguez, Yanetsys. El análisis de información y las investigaciones cuantitativa y cualitativa. Última actualización: 2007. Disponible en: http://bvs.sld.cu/revistas/spu/vol33_3_07/spu20207.htm.
- SQL Power Architect herramienta de modelado de datos. Última actualización: 2010. Disponible en:

<http://www.tuinformaticafacil.com/herramientas-desarrollo/sql-powerarchitect-herramienta-de-modelado-de-datos>.

- Vega Torres, L. La inteligencia de negocio. Su implementación mediante la plataforma Pentaho., CITMATEL, 2008. 22.
- Velasco JJ. IBM construye el mayor almacén de datos del mundo. Última actualización: 2011. (). Disponible en: <http://alt1040.com/2011/08/ibm-construye-mayor-almacen-datos-mundo>.
- Wolff, Carmen. [2000] Implementando un DataWarehouse. Revista Ingeniería Informática: revista electrónica del DIICC.

Glosario de términos

MD: Datamart(por sus siglas en ingles) Mercado de Datos.

AD: Datawarehouse(por sus siglas en ingles) Almacén de Datos.

APA: Administración Provincial de Artemisa.

TIC: Tecnología de la Información y las Comunicaciones.

SAI: Sistemas de Almacenamiento de la Información.

OLAP: On Line Analytical Processing(por sus siglas en ingles) Procesamiento Analítico en Línea.

ROLAP: Relational On Line Analytical Processing(por sus siglas en ingles) Procesamiento Analítico Relacional en Línea.

MOLAP: Multidimensional On Line Analytical Processing(por sus siglas en ingles) Procesamiento Analítico Multidimensional en Línea.

ETL: Extraction Transformation and Load (por sus siglas en ingles) Extracción Transformación y Carga.

BI: Business Intelligence(por sus siglas en ingles) Inteligencia del Negocio.

Ralph Kimball: Conocido innovador, escritor, educador y consultor en el campo de Almacenes de Datos. En la actualidad posee más de 100 artículos sobre inteligencia empresarial.

William H. Inmon: Conocido como “El padre de la tecnología de Almacenes de Datos”. Ha escrito más de 650 artículos sobre construcción, uso y mantenimiento de almacenes de datos. Posee la autoría de más de 46 libros de temas relacionados a tecnologías de base de datos.

DDL: (Lenguaje de Definición de Datos) lenguaje proporcionado por el sistema de gestión de base de datos que permite a los usuarios de la misma llevar a cabo las tareas de definición de las estructuras que almacenarán los datos así como de los procedimientos o funciones que permitan consultarlos.