

Universidad de las Ciencias Informáticas
Facultad 3



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

Título: Mercado de Datos para los departamentos de
Procesos Penales y Gestión de Cuadros y Personal de
Apoyo de la Fiscalía General de la República de Cuba.

Autores: Beatriz Danmara Hidalgo García

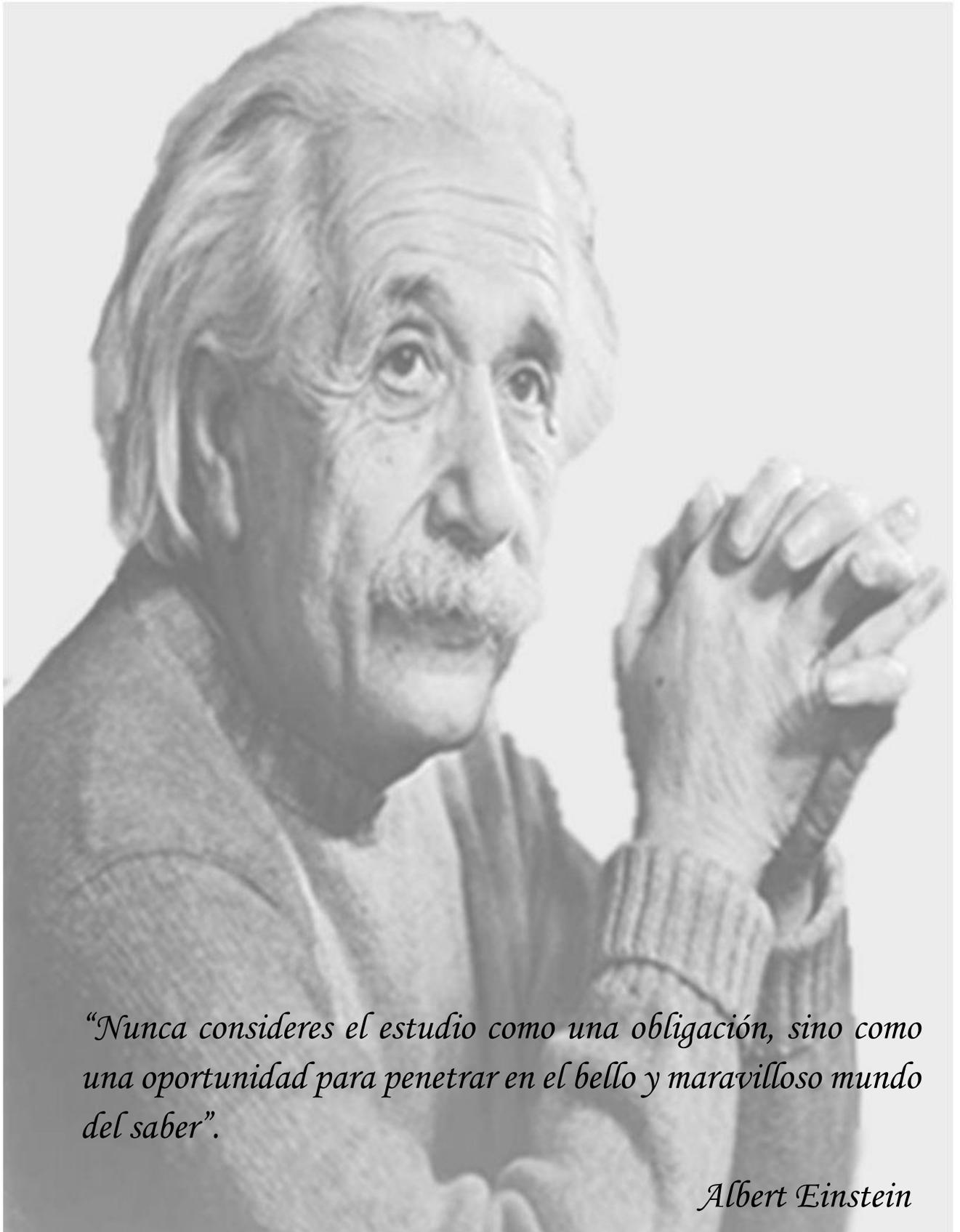
Yasmani Otero Morfa

Tutores: Ing. Manuel Álvarez Alonso

Ing. José Carlos Pupo Acosta

Ing. Abel Andres Írsula Tumbarell

Ciudad de la Habana, mayo 2015



“Nunca consideres el estudio como una obligación, sino como una oportunidad para penetrar en el bello y maravilloso mundo del saber”.

Albert Einstein



DECLARACIÓN JURADA DE AUTORÍA

Declaración jurada de autoría

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

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

Beatriz D. Hidalgo García

Yasmani Otero Morfa

Ing. Manuel Álvarez Alonso

Ing. José Carlos Pupo Acosta

Ing. Abel Andres Irsula Tumbarell



DATOS DE CONTACTO

Autores:

Beatriz D. Hidalgo García

Correo electrónico: bdhidalgo@estudiantes.uci.cu

Yasmani Otero Morfa

Correo electrónico: yomorfa@estudiantes.uci.cu

Tutor:

Ing. Manuel Álvarez Alonso

Correo electrónico: malvareza@uci.cu

Co-tutores:

Ing. José Carlos Pupo Acosta

Correo electrónico: jcpupo@uci.cu

Ing. Abel Andres Irsula Tumbarell

Correo electrónico: aairsula@uci.cu



DEDICATORIA

De Beatriz:

Con todo mi cariño y mi amor para las personas que hicieron todo en la vida para que yo pudiera lograr mis sueños, por motivarme y darme la mano cuando sentía que el camino se terminaba. Gracias por existir.

Mamá y Papá

A tu paciencia y comprensión, preferiste sacrificar tu tiempo para que yo pudiera cumplir con el mío. Por tu bondad y sacrificio me inspiraste a ser mejor para ti, ahora puedo decir que esta tesis lleva mucho de ti, gracias por estar siempre a mi lado.

Yasmani

De Yasmani:

Gracias a esas personas importantes en mi vida, que siempre estuvieron listas para brindarme toda su ayuda, ahora me toca regresar un poquito de todo lo inmenso que me han otorgado. Esta tesis va por ustedes.

Mama y Papá

A mi compañera y amiga por estar ahí en los momentos más duros y difícil sin emitir una sola queja. Gracias por ayudarme a reflexionar en los momentos que lo necesitaba, por estar pendiente de mí. Estoy orgulloso de haber triunfado contigo. Gracias por estar a mi lado.

Beatriz



AGRADECIMIENTOS

De Beatriz y Yasmani

Llegado este momento es muy difícil agradecer a todas las personas que nos han ayudado a cumplir este sueño. Todas son importantes para nosotros. Por esta razón no pretendemos nombrarlas una a una, siempre sé que nos quedaría alguien sin mencionar y no sería justo. Por eso decidimos que fuese así:

A todas aquellas personas que estuvieron ahí cuando los necesitamos y fueron capaces de ayudarnos con todo el amor del mundo. Gracias.



RESUMEN

Resumen

En el presente trabajo se expone la creación de los mercados de datos para los departamentos de Procesos Penales y Gestión de Cuadros y Personal de Apoyo de la Fiscalía General de la República de Cuba (FGR). El desarrollo de estos mercados de datos surge con el objetivo de facilitarle a la FGR una forma más eficiente de poder almacenar la información que se gestiona en los departamentos antes mencionados para su explotación y análisis. La implementación, el diseño y el modelado dimensional de los datos se basan en la metodología de Kimball. La fase más importante del proceso de construcción de los mercados de datos es el proceso de extracción, transformación y carga de la información, desarrollado con la herramienta Pentaho Data Integration. En esta fase se obtienen las necesidades de información de los usuarios del negocio y se almacenan en los mercados de datos correspondientes. Se utilizó para la validación de la solución las pruebas de volumen, carga y estrés para comprobar la capacidad y rendimiento de los mercados de datos en consecuencia con los requisitos definidos por el cliente.

Palabras Claves: *Mercados de datos, modelado dimensional, metodología.*



ÍNDICE

Índice

Introducción.....	1
Capítulo 1: marco teórico	4
Introducción	4
1.1 Sistemas transaccionales	4
1.2 Inteligencia de negocios	4
1.3 Almacén de datos o data warehouse	4
1.4 Comparación entre los sistemas transaccionales y los almacenes de datos	6
1.5 Mercado de datos o data mart	7
1.6 Sistema olap	8
1.7 Arquitectura del almacén de datos	11
1.8 Metodologías de desarrollo para un almacén de datos	12
1.9 Tendencias.....	17
1.10 Herramientas y tecnologías usadas para la realización de los mercados.	18
1.11 Pruebas de rendimiento.....	22
1.12 Conclusiones del capítulo:.....	23
Capítulo 2: diseño e implementación de los mercados de datos	24
Introducción	24
2.1 Arquitectura de los mercados de datos.....	24
2.2 Análisis del estado de las fuentes de datos	24
2.3 Metodología de kimball	25
2.4 Conclusiones del capítulo.....	39
Capítulo 3: pruebas de validación	40
3 Introducción.....	40
3.1 Pruebas de rendimiento.....	40
3.2 Conclusiones del capítulo	45
Conclusiones generales	46
Recomendaciones	47
Bibliografía.....	48
Anexos	¡ERROR! MARCADOR NO DEFINIDO.



Índice de tabla

Tabla 1: Comparación entre el sistema transaccional y el almacén de datos	6
Tabla 2: Comparación entre Inmon y Kimball.....	13
Tabla 3: Temas analíticos departamento Procesos Penales Sumarios	27
Tabla 4: Nivel de granularidad para las tablas de dimensiones tiempo, instancia, indicador y provincia	29
Tabla 5: Matriz de procesos/dimensiones Índice de Peligrosidad Predelictiva.....	32
Tabla 6: Lista de atributos de las tablas de dimensiones.	32
Tabla 7: Lista de atributos de las tablas de dimensiones de Gestión de Cuadros y Personal de Apoyo.....	32
Tabla 8: Correspondencia de datos	36
Tabla 9: Representación de las instancias y su respectiva cantidad de usuarios	41
Tabla 10: Representación del tiempo de respuesta de las consultas	41
Tabla 11: Representación de los campos de la herramienta JMeter	42
Tabla 12: Comparación de los tiempos de respuesta	44
Tabla 13: Representación de la cantidad de usuarios para la prueba de estrés.	44
Tabla 14: Temas analíticos del departamento Gestión de cuadro y personal de Apoyo ¡Error! Marcador no definido.	
Tabla 15: Temas analíticos del departamento Índice de Peligrosidad Predelictiva	¡Error! Marcador no definido.
Tabla 16: Temas analíticos del departamento Procesos Penales Ordinarios ¡Error! Marcador no definido.	
Tabla 17: Granularidad del mercado de datos Gestión de cuadro y Personal de Apoyo ¡Error! Marcador no definido.	
Tabla 18: Matriz de procesos/dimensiones Procesos Penales Ordinarios.	¡Error! Marcador no definido.
Tabla 19: Matriz de procesos/dimensiones Procesos Penales Sumarios.....	¡Error! Marcador no definido.
Tabla 20: Matriz de procesos/dimensiones Gestión de Cuadros y Personal de Apoyo. ¡Error! Marcador no definido.	
Tabla 21: Atributos de las tablas de hechos del mercado de datos para los departamentos Procesos Penales Ordinarios	¡Error! Marcador no definido.
Tabla 22: Atributos de las tablas de hechos del mercado de datos Índice de Peligrosidad Predelictiva	¡Error! Marcador no definido.
Tabla 23: Atributos de las tablas de hechos del mercado Gestión de Cuadros y Personal de Apoyo.. ¡Error! Marcador no definido.	
Tabla 24: Atributos de la tabla de hechos del mercado de datos Procesos Penales Sumarios.....	¡Error! Marcador no definido.



Índice de Figuras

Figura 1: Componentes de un sistema de BI.....	4
Figura 2: Representación de la integración de un almacén de datos.	5
Figura 3: Representación de la tematicidad del almacén de datos.....	5
Figura 4: Representación de los datos históricos.....	6
Figura 5: Representación de la no volatilidad del almacén de datos.	6
Figura 6: Cubo OLAP.....	8
Figura 7: Esquema en estrella.	10
Figura 8: Esquema en copo de nieve.	11
Figura 9: Esquema constelación de hechos.	11
Figura 10: Arquitectura de un almacén de datos.....	12
Figura 11: Arquitectura de los mercados de datos.	24
Figura 12: Ejemplo del uso del patrón de llave subrogada.	33
Figura 13: Modelo físico de datos para el departamento de Índice de Peligrosidad Predelictiva.	34
Figura 14: Modelo físico de datos para el departamento de Procesos Penales Sumario.	35
Figura 15: Fragmento del modelo físico para el departamento de Índice de Peligrosidad Predelictiva.	36
Figura 16: transformación para la dimensión instancia.	37
Figura 17: Fragmento de transformación para el hecho denuncias recibidas en la fiscalía.	37
Figura 18: Trabajo para el departamento Procesos Penales Sumario.	38
Figura 19: Esquema multidimensional para el mercado de datos del departamento Procesos Penales Sumarios	38
Figura 20: Fragmento de visualización de los datos del cubo Procesos Penales Sumarios para el mercado de datos del departamento Procesos Penales Sumarios.....	39
Figura 21: Prueba de volumen para el mercado de datos para el departamento Procesos Penales Sumarios.	40
Figura 22: Resultados del plan de pruebas simulando para la Fiscalía Provincial.	43
Figura 23: Resultados del plan de pruebas simulando para la Fiscalía General.....	43



ÍNDICE

Figura 24: Resultados del plan de prueba simulado.	44
Figura 25: Diseño físico para el mercado de datos del departamento Procesos Penales Ordinarios... ¡Error! Marcador no definido.	
Figura 26: Diseño físico para el mercado de datos del departamento Gestión de cuadro y Personal de Apoyo..... ¡Error! Marcador no definido.	
Figura 27: Conexión a la base de datos. ¡Error! Marcador no definido.	
Figura 28: Transformación realizada para la dimensión indicador..... ¡Error! Marcador no definido.	
Figura 29: Transformación realizada para la dimensión provincia. ¡Error! Marcador no definido.	
Figura 30: Transformación realizada para la dimensión tiempo..... ¡Error! Marcador no definido.	
Figura 31: Transformación realizada para las tablas de hechos..... ¡Error! Marcador no definido.	
Figura 32: Fragmento del trabajo realizado para el mercado de datos..... ¡Error! Marcador no definido.	
Figura 33: Transformación realizada para las tablas de hechos..... ¡Error! Marcador no definido.	
Figura 34: Fragmento del trabajo realizado para el mercado de datos..... ¡Error! Marcador no definido.	
Figura 35: Transformación realizada para la dimensión cargo. ¡Error! Marcador no definido.	
Figura 36: Transformación realizada para la dimensión provincia. ¡Error! Marcador no definido.	
Figura 37: Transformación realizada para las tablas de hechos..... ¡Error! Marcador no definido.	
Figura 38: Fragmento del trabajo realizado para el mercado de datos..... ¡Error! Marcador no definido.	
Figura 39: Conexión del Workbench..... ¡Error! Marcador no definido.	
Figura 40: Esquema multidimensional para el mercado de datos de Gestión de cuadro y Personal de Apoyo..... ¡Error! Marcador no definido.	
Figura 41: Esquema multidimensional para el mercado de Procesos Penales Ordinarios. ¡Error! Marcador no definido.	
Figura 42: Esquema multidimensional para el mercado de Índice de Peligrosidad Predelictiva..... ¡Error! Marcador no definido.	
Figura 43: Procesos Penales Ordinarios. ¡Error! Marcador no definido.	
Figura 44: Índice de Peligrosidad Predelictiva..... ¡Error! Marcador no definido.	
Figura 45: Gestión de Cuadros y Personal de Apoyo. ¡Error! Marcador no definido.	
Figura 46: Departamento Procesos Penales Ordinarios. ¡Error! Marcador no definido.	
Figura 47: Departamento Índice de Peligrosidad Predelictiva..... ¡Error! Marcador no definido.	
Figura 48: Departamento Gestión de Cuadros y Personal de Apoyo. ¡Error! Marcador no definido.	
Figura 49: Departamento de Índice de Peligrosidad Predelictiva..... ¡Error! Marcador no definido.	
Figura 50: Departamento de Procesos Penales Ordinarios. ¡Error! Marcador no definido.	
Figura 51: Departamento de Procesos Penales Sumarios. ¡Error! Marcador no definido.	
Figura 52: Departamento de Índice de Peligrosidad Predelictiva..... ¡Error! Marcador no definido.	
Figura 53: Departamento de Procesos Penales Ordinarios. ¡Error! Marcador no definido.	
Figura 54: Departamento de Procesos Penales Sumarios. ¡Error! Marcador no definido.	



ÍNDICE

Figura 55: Departamento de Índice de Peligrosidad Predelictiva.....**¡Error! Marcador no definido.**
Figura 56: Departamento de Procesos Penales Ordinarios.**¡Error! Marcador no definido.**
Figura 57: Departamento de Procesos Penales Sumarios.**¡Error! Marcador no definido.**



INTRODUCCIÓN

Introducción

Actualmente en las actividades diarias de cualquier organización, se generan datos como producto secundario, que son el resultado de todas las transacciones que se realizan. Es muy común, que los mismos se almacenen y administren a través de sistemas transaccionales en bases de datos relacionales. La idea central que se persigue en las organizaciones es que dejen de ser solo simples datos para convertirse en información que enriquezca la toma de decisiones. Precisamente, la inteligencia de negocios (Business Intelligence - BI), entrará a jugar un rol protagónico debido a que proporcionará las herramientas necesarias para analizar la información contenida en esos sistemas transaccionales. Para desarrollar la BI en las empresas u organizaciones, es necesario gestionar datos guardados en distintos formatos, fuentes, para luego depurarlos e integrarlos, además de almacenarlos en un solo destino o base de datos que permita su posterior análisis y exploración, es imperativo y de vital importancia contar con una herramienta que satisfaga todas estas necesidades. Esta recibe el nombre de Almacén de datos y Mercado de datos, siendo el segundo un subconjunto del primero.

Cuba no está al margen de los adelantos tecnológicos del mundo y ha volcado su esfuerzo para que las empresas sean cada día más eficientes y competitivas. La Universidad de las Ciencias Informáticas (UCI) es un centro de altos estudios que tiene como misión servir de soporte a la industria cubana de la informática y del mismo modo contribuye a que la meta de informatización del país sea alcanzada. Nuestra universidad cuenta con diversos centros de desarrollo enfocados a soluciones informáticas entre los que está presente el Centro de Gobierno Electrónico (CEGEL). En el mismo se encuentra el Sistema de Informatización de la Gestión de la Fiscalía (SIGEF), el cual se encarga de informatizar los procesos de la Fiscalía General de la República de Cuba (FGR), que es el órgano encargado de vigilar por el estricto cumplimiento de la constitución, las disposiciones legales y las leyes.

Entre los módulos del SIGEF se encuentran, Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva y Gestión de Cuadros y Personal de Apoyo. El sistema transaccional existente no facilita la búsqueda de información histórica con agilidad, ni la realización de comparaciones con la totalidad de los datos significativos en un determinado intervalo en el tiempo, debido a la forma en que está organizada dicha información; además provoca que los reportes estadísticos que hoy se obtienen no cumplan con los tiempos de respuestas esperados por los directivos de la FGR.

Dada la situación anteriormente descrita, se plantea el siguiente **problema de la investigación** ¿Cómo organizar la información existente en el sistema transaccional de los departamentos Procesos Penales y Gestión de Cuadros y Personal de Apoyo de la Fiscalía General de la República de Cuba, para disminuir los tiempos de acceso de los directivos de estas áreas a los datos contenidos en los reportes estadísticos?



INTRODUCCIÓN

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

El **objetivo general** de este trabajo es desarrollar los mercados de datos que organicen la información existente en el sistema transaccional de los departamentos Procesos Penales y Gestión de Cuadros y Personal de Apoyo de la Fiscalía General de la República de Cuba, para disminuir los tiempos de acceso de los directivos de estas áreas a los datos contenidos en los reportes estadísticos.

Teniendo como referencia el objetivo general y el problema se plantea como **idea a defender**: Con el desarrollo de los mercados de datos que organicen la información existente en el sistema transaccional de los departamentos Procesos Penales y Gestión de Cuadros y Personal de Apoyo de la Fiscalía General de la República de Cuba, se disminuirán los tiempos de acceso de los directivos de estas áreas a los datos en los reportes estadísticos.

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

- Definir el marco teórico referencial de la investigación a partir del análisis de las principales características de los mercados de datos.
- Identificar las necesidades de información de los departamentos Procesos Penales y Gestión de Cuadros y Personal de Apoyo de la Fiscalía General de la República de Cuba, a partir de los procesos implicados en los mercados de datos.
- Realizar el análisis, diseño e implementación de los mercados de datos definidos para facilitar el análisis de la información contenida.
- Validar los mercados de datos obtenidos a través de pruebas de rendimiento.

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

- Análisis de información significativa relacionada con las características y comportamiento de los Mercados de Datos.
- Selección de la metodología, herramientas y tecnologías a emplear en el desarrollo de los Mercados de Datos.
- Diseño de los Mercados de Datos de los departamentos procesos Penales y Gestión de cuadro y personal de Apoyo, para facilitar el análisis de información contenida.
- Implementación de los Mercados de Datos diseñados para los departamentos procesos Penales y Gestión de cuadro y personal de Apoyo.
- Definición de los tipos de pruebas a ejecutar para validar los mercados de datos.
- Ejecución de las pruebas de volumen, carga y estrés para validar el Mercado de Datos elaborado.



INTRODUCCIÓN

Métodos de investigación científica a utilizar:

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

Métodos teóricos:

Analítico-Sintético: Posibilita el análisis y procesamiento de la información recopilada durante la investigación sobre los Mercados de Datos, metodologías, tecnologías y herramientas de desarrollo.

Histórico Lógico: Permite el estudio de los antecedentes y tendencias actuales de los almacenes de datos, metodologías, tecnologías y herramientas de desarrollo.

Modelación: Facilita realizar la modelación del ciclo de vida del Mercado de Datos para los módulos Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva y Gestión de Cuadros y Personal de Apoyo de SIGEF.

Métodos empíricos:

Medición: Posibilita realizar pruebas al Mercado de Datos para verificar el correcto funcionamiento de las funcionalidades implementadas.

Experimento: Mediante este método se verifica la validez de la propuesta de solución, al comprobar la disminución real de los tiempos de acceso a los datos.

Con el propósito de dar cumplimiento a los objetivos trazados, el presente documento estará estructurado en tres capítulos que a continuación se exponen:

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

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

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



CAPÍTULO 1: MARCO TEÓRICO

Capítulo 1: Marco teórico

Introducción

En el presente capítulo se abordarán temas teóricos relacionados con los almacenes de datos y los mercados de datos. Se podrá conocer acerca de algunas metodologías usadas para su desarrollo, además de las herramientas utilizadas para su realización.

1.1 Sistemas transaccionales

OLTP es la sigla en inglés de Procesamiento de Transacciones En Línea (On-Line Transaction Processing). Es un tipo de sistema de información diseñado para recolectar, almacenar, modificar y recuperar todo tipo de información que es generada por las transacciones en una organización. Suelen ser los primeros sistemas de información que se implementan en la empresa. Son optimizados para almacenar grandes volúmenes de datos, pero no para analizar los mismos (Alegsa, 2009).

1.2 Inteligencia de negocios

Se denomina inteligencia empresarial, inteligencia de negocios o BI (del inglés Business Intelligence) al conjunto de estrategias y tecnologías que ayudan a convertir los datos en información de calidad, y dicha información en conocimiento que permita una toma de decisiones más acertadas y ayude a mejorar la competitividad (Ramos, 2011). En la siguiente figura se muestran los componentes de un sistema de BI.

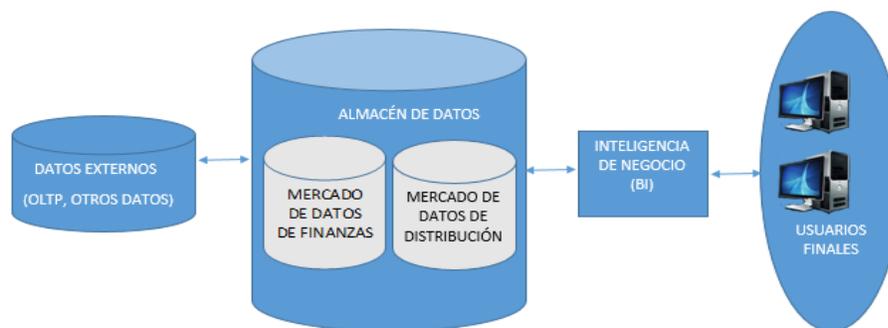


Figura 1: Componentes de un sistema de BI.

1.3 Almacén de datos

Un almacén de datos es una base de datos corporativa en la que se integra información depurada de las diversas fuentes que hay en la organización. Dicha información debe ser homogénea y fiable, se almacena de forma que permita su análisis desde muy diversas perspectivas, y que a su vez dé unos tiempos de respuesta óptimos. Para ello la información se encuentra altamente desnormalizada (Ramos, 2011).



CAPÍTULO 1: MARCO TEÓRICO

Características del almacén de datos

Un almacén de datos se caracteriza por ser (Ramos, 2011):

Integrado: Los datos se deben integrar en una estructura consistente, debiendo eliminarse las inconsistencias existentes entre los diversos sistemas operacionales. La información se estructura en diversos niveles de detalle para adecuarse a las necesidades de consulta de los usuarios. Algunas de las inconsistencias más comunes son: en nomenclatura, en unidades de medida, en formatos de fechas, múltiples tablas con información similar. En la siguiente figura se muestran un ejemplo de esta característica.

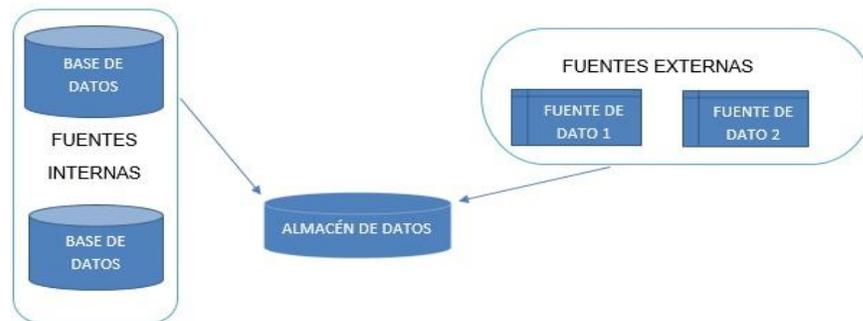


Figura 2: Representación de la integración de un almacén de datos.

Orientado a temas: Los datos están organizados por temas para facilitar el entendimiento por parte de los usuarios, de forma que todos los datos relativos a un mismo elemento de la vida real queden unidos entre sí. Por ejemplo, todos los datos de un cliente pueden estar consolidados en una misma tabla, todos los datos de los productos en otra, y así sucesivamente. En la siguiente figura se muestran un ejemplo de esta característica.

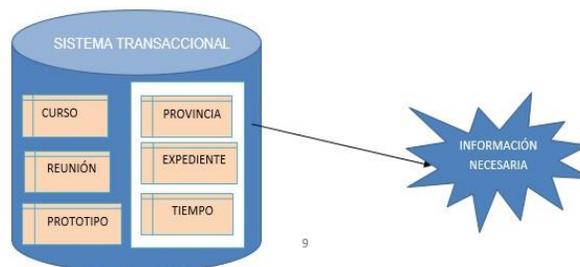


Figura 3: Representación de la tematicidad del almacén de datos.

Histórico: Los datos que existen en un almacén de datos pueden ir variando a lo largo del tiempo, deben quedar reflejados de forma que al ser consultados reflejen estos cambios y no se altere la



CAPÍTULO 1: MARCO TEÓRICO

realidad que había en el momento en que se almacenaron. En la siguiente figura (Ramos, 2011) se muestran un ejemplo de esta característica.

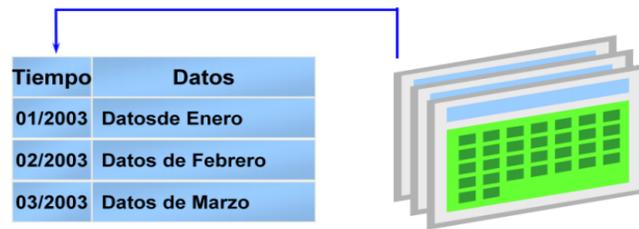


Figura 4: Representación de los datos históricos.

No volátil: La información contenida en un almacén de datos, una vez introducida, debe ser de sólo lectura, nunca se modifica ni se elimina, y ha de ser permanente y mantenerse para futuras consultas. En la siguiente figura se muestran un ejemplo de esta característica.

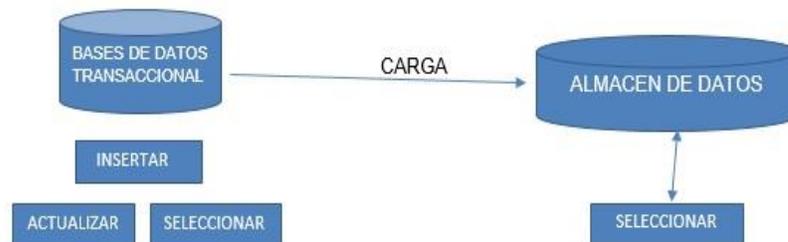


Figura 5: Representación de la no volatilidad del almacén de datos.

1.4 Comparación entre los sistemas transaccionales y los mercados de datos.

Los sistemas transaccionales y los mercados de datos se diferencian principalmente en las siguientes características que se muestran en la tabla a continuación (Trujillo, 2009).

Tabla 1: Comparación entre el sistema transaccional y mercados de datos

SISTEMA TRANSACCIONAL	ALMACÉN DE DATOS
Predomina la actualización.	Predomina la consulta.
Importancia del dato actual.	Importancia del dato histórico.
Estructura relacional.	Estructura multidimensional.
La actividad más importante es tipo operativa, día a día.	La actividad más importante consiste en el análisis de la información
Usuarios de perfiles medios o bajos.	Usuarios de perfiles altos.
Bases de datos medianas (100Mb-1Gb).	Bases de datos grandes (100Gb-1Tb).



CAPÍTULO 1: MARCO TEÓRICO

Datos en general normalizados.	Datos en general desnormalizados.
--------------------------------	-----------------------------------

En las organizaciones se utilizan comúnmente bases de datos transaccionales, las cuales son muy factibles cuando se necesita modificar, actualizar e insertar información, sin embargo cuando se realizan consultas para seleccionar grandes volúmenes de datos se afecta negativamente los tiempos de respuesta a dichas consultas. A continuación se presentan otros inconvenientes de los sistemas transaccionales (Nima, 2015):

- Cuando el sistema está respondiendo a consultas analíticas complejas puede ser lento o impredecible, lo que causa un servicio poco eficiente a los usuarios del proceso analítico en línea.
- Los datos que se modifican con frecuencia interfieren en la coherencia de la información analítica.

Para ofrecer una solución a estos inconvenientes se han diseñado los mercados de datos. Estos son capaces de brindar una solución debido a que (Nima, 2015):

- Contienen datos transformados que son válidos, coherentes, consolidados y con el formato adecuado para realizar el análisis.
- Proporcionan datos estables que representan el historial de la empresa.
- Organizan los datos buscando la eficiencia de las consultas analíticas.

Se decide utilizar los mercados de datos para gestionar la información de los departamentos de Procesos Penales y Gestión de Cuadros y Personal de Apoyo de la Fiscalía General de la República de Cuba, dado que permiten consultar grandes cantidades de datos de manera eficiente y almacenar información histórica de estos departamentos.

1.5 Mercado de datos

En muchas ocasiones los términos mercados de datos y almacén de datos son utilizados indistintamente. La diferencia es solamente en cuanto al alcance. Mientras que un almacén de datos es un sistema centralizado con datos globales de la empresa y de todos sus procesos operacionales, un mercado de datos es un subconjunto temático de datos, orientado a un proceso o un área de negocio específica propia de la organización (Ramos, 2011).

Características de un mercado de datos (Kimball, 2004)

- Posee usuarios limitados, debido a que solo utilizan el sistema las personas con conocimientos del negocio en esa área donde se encuentran.
- Se especializa en el almacenamiento de los datos de un área de negocio específica.



CAPÍTULO 1: MARCO TEÓRICO

- Posee función de apoyo debido a que ayuda a la consulta, el análisis de información y a los reportes de la organización.
- Son fácilmente entendibles y navegables, porque emplea términos propios de la organización para facilitar su entendimiento, además de presentar gráficas interactivas.

Ventajas de un Mercado de datos

- Los mercados de datos son específicos para una necesidad de datos seleccionados, lo que trae consigo proveer la capacidad de analizar y explotar las diferentes áreas de trabajo y de realizar un análisis inmediato de las mismas, enfatizando el cómodo acceso a una información relevante.
- Elimina la producción y el procesamiento de datos que no son utilizados ni necesarios por eso se simplifica el desarrollo de todo el mecanismo de su base de datos y por ende baja substancialmente todo el coste del proyecto, así como su duración.
- Transforma datos orientados a las aplicaciones en información orientada a la toma de decisiones, permitiendo así que las mismas logren tener la calidad esperada y puedan brindar beneficios grandes y tangibles.
- Provee una interfaz de consulta que permite al usuario la posibilidad de análisis de la información.
- Provee de la cantidad de información necesaria a la mayor cantidad de personas posibles dentro del departamento (Kimball, 2004).

1.6 Sistema OLAP

OLAP es el acrónimo en inglés de procesamiento analítico en línea (On-Line Analytical Processing) (Frahling, 2015). Hace uso de bases de datos multidimensionales que tienen como objetivo agilizar la consulta de grandes cantidades de datos. En un modelo de datos OLAP, la información es vista como cubos, los cuales consisten de categorías descriptivas (dimensiones) y valores cuantitativos (medidas). OLAP es la tecnología utilizada para la creación de los almacenes de datos (Nima, 2015). En la siguiente figura (Kimball, 2004) se muestran un ejemplo de un cubo OLAP.

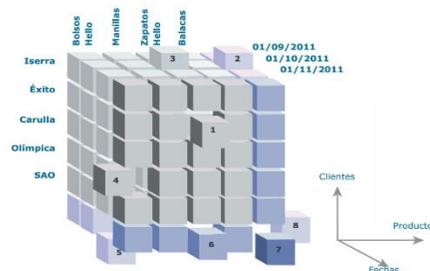


Figura 6: Cubo OLAP.



CAPÍTULO 1: MARCO TEÓRICO

Tipos de sistemas OLAP

Tradicionalmente, según la estructura de datos utilizada este sistema se clasifica según las siguientes categorías (Ibarra, 2006):

ROLAP es el acrónimo inglés de Relacional Online Analytical Processing es decir, procesamiento analítico relacional en línea. Es una implementación OLAP que almacena los datos en una base de datos relacional. Típicamente, los datos son detallados, evitando las agregaciones y las tablas se encuentran normalizadas. Estos datos se calculan en tiempo de consulta por lo que el tiempo de respuesta sería mayor pero sin usar enormes cantidades de almacenamiento en disco.

MOLAP es el acrónimo inglés de Multidimensional Online Analytical Processing, es decir, procesamiento analítico multidimensional en línea. Es una implementación OLAP que almacena los datos en una base de datos multidimensional. Para optimizar los tiempos de respuesta, el resumen de la información es usualmente calculado por adelantado.

HOLAP: (Hybrid OLAP). Almacena algunos datos en un motor relacional y otros en una base de datos multidimensional.

Luego de conocidos los diferentes modelos de almacenamiento de OLAP y atendiendo a sus características se decide utilizar ROLAP, dado que las bases de datos que se construirán como parte de los mercados de datos están soportadas por el gestor de base de datos PostgreSQL, el cual es un sistema gestor de base de datos relacional.

Como se mencionó anteriormente los cubos OLAP se estructuran en tablas de hechos y tablas de dimensiones, a continuación se explica en qué consisten:

Tablas de hechos

Las tablas de hechos contienen datos cuantitativos. Representan datos instantáneos en el tiempo, que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de dimensiones. Los datos presentes en las tablas de hechos constituyen el volumen de almacén de datos, y pueden estar compuestos por millones de registros dependiendo de su granularidad y antigüedad de la organización. (Dario, 2009).

Tablas dimensiones

Las tablas de dimensiones por su parte definen como están los datos organizados lógicamente y proveen el medio para analizar el contexto del negocio. Contienen datos cualitativos. Representan los aspectos de interés, mediante los cuales los usuarios podrán filtrar y manipular la información almacenada en la tabla de hechos (Dario, 2009).



CAPÍTULO 1: MARCO TEÓRICO

Luego de saber en qué consisten las tablas de hechos y dimensiones es importante conocer lo que se puede lograr al involucrar ambos términos, en ese caso se estaría hablando de un análisis multidimensional.

Análisis multidimensional

El análisis multidimensional provee diferentes perspectivas de los datos a través de las diferentes dimensiones. Ejemplos de dimensiones pueden ser: tiempo, ubicación, producto. En general, los usuarios observan la información contenida en la tabla de hechos por alguna de las dimensiones, por ejemplo “ver todos los clientes (en este caso los datos de clientes representarían los hechos) por región (la región sería la dimensión)”. Esta dimensión es geográfica, y el usuario puede solicitar detalles adicionales una vez realizada la consulta. Por ejemplo, “ver todos los clientes nuevos por distrito” (Aguilera, 2010).

Para lograr realizar el análisis multidimensional, las bases de datos multidimensionales implican tres posibles modelos de datos, los mismos se utilizarán dependiendo de la información referente al negocio.

Tipos de modelos de las base de datos multidimensionales

Para representar la información de las bases de datos multidimensionales existen tres posibles tipos de modelado de los datos.

Esquema en estrella: Es la arquitectura de almacén de datos más simple. Este diseño consta de una sola tabla de hechos central rodeada de tablas de dimensiones y juntos forman una estructura que permite implementar mecanismos básicos para poder utilizarla con una herramienta de consultas OLAP (Bernabeu, 2010).

En la siguiente figura (Bernabeu, 2010) se muestra un esquema de estrella con una sola tabla de hechos relacionada con cuatro tablas de dimensiones. El esquema de estrella puede tener cualquier número de tablas de dimensiones.

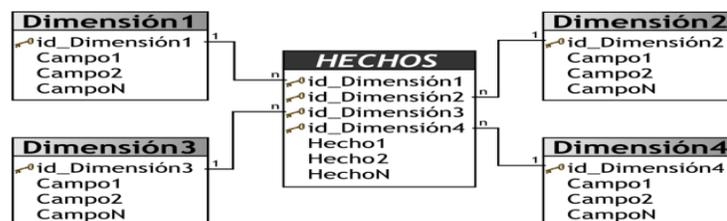


Figura 7: Esquema en estrella.

Esquema en copo de nieve: Es una variedad más compleja del esquema en estrella. Este esquema es usado cuando las tablas de dimensiones están muy grandes o complejas y es muy difícil representar los datos en esquema en estrella. Este esquema consta de una tabla de hechos que está relacionada



CAPÍTULO 1: MARCO TEÓRICO

con muchas tablas de dimensiones, que pueden estar relacionadas a su vez a otras tablas de dimensiones (Bernabeu, 2010).

En la siguiente figura (Bernabeu, 2010) se muestra un esquema en copo de nieve con la tabla de hechos, relacionada con seis dimensiones.

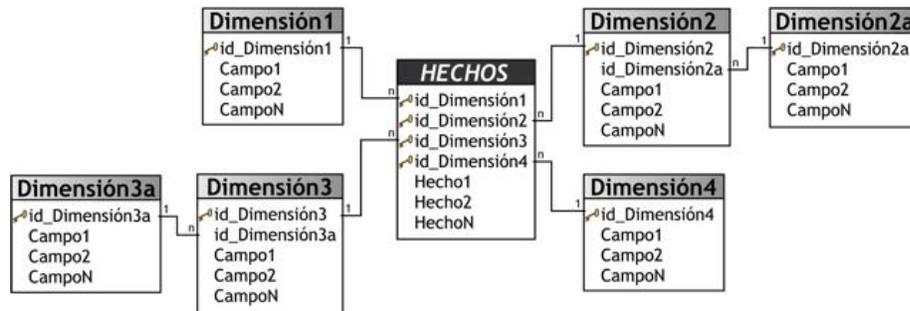


Figura 8: Esquema en copo de nieve.

Esquema de constelación de hechos: Un esquema de constelación es una combinación de un esquema de estrella y un esquema de copo de nieve. Este esquema es más complejo que las otras arquitecturas debido a que contiene múltiples tablas de hechos. Con esta solución las tablas de dimensiones pueden estar compartidas entre más de una tabla de hechos.

En la siguiente figura (Bernabeu, 2010) se muestra un esquema en constelación de hechos con dos tablas de hechos relacionadas con las dimensiones correspondientes.

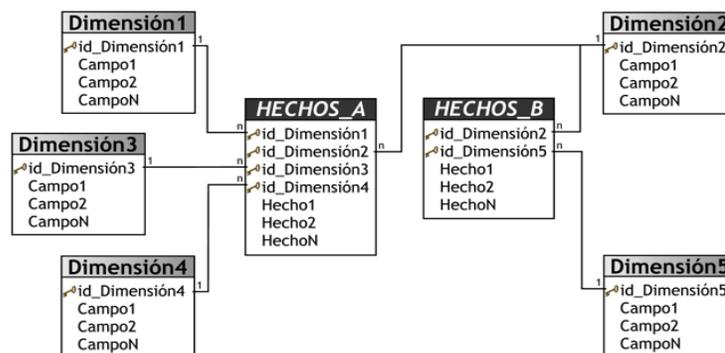


Figura 9: Esquema constelación de hechos.

1.7 Arquitectura del almacén de datos (Kimball, 2004)

La arquitectura del almacén de datos juega un papel fundamental en la construcción del mismo. A continuación se definirán los elementos primordiales que se estarán relacionando y que ejecutan una función específica en todo el proceso, para que al final el almacén logre el propósito por el cual fue creado. Su estructura básica incluye:



CAPÍTULO 1: MARCO TEÓRICO

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

- En el **gestor de carga** los datos son integrados, transformados y limpiados, para luego ser cargados en el almacén de datos básicamente es el proceso de extracción, transformación y carga (ETL) por sus siglas en inglés. Se extraen datos de las diversas fuentes que se requieran, y son transformados para resolver posibles problemas de inconsistencias entre los mismos y finalmente, después de haberlos depurado se procede a su carga en el depósito de datos.
- En el **gestor del almacén de datos** la información del almacén de datos se estructura en cubos multidimensionales, los cuales preparan esta información para responder a consultas dinámicas. Realiza copias de resguardo incremental o total de los datos del almacén de datos.
- El **gestor de consultas** realiza las operaciones necesarias para soportar los procesos de gestión y ejecución de consultas relacionales, tales como join (unión) y agregaciones y de consultas propias del análisis de datos. Recibe las consultas del usuario, las aplica a las tablas correspondientes y devuelve los resultados obtenidos.
- Las **herramientas de consulta y análisis** son sistemas que permiten al usuario realizar la exploración de datos del almacén de datos. Básicamente constituyen el nexo entre el depósito de datos y los usuarios. Los usuarios acceden al almacén de datos utilizando diversas herramientas de consulta, exploración, análisis y reportes.
- Los **usuarios** que posee el almacén de datos son aquellos que se encargan de tomar decisiones y de planificar las actividades del negocio, es por ello que se hace tanto énfasis en la integración y limpieza de datos, para poder conseguir que la información posea toda la calidad posible.

En la siguiente figura se muestran la arquitectura de un almacén de datos.



Figura 10: Arquitectura de un almacén de datos.

1.8 Metodologías de desarrollo para un almacén de datos

Entre las metodologías conocidas para el desarrollo de los almacenes de datos se encuentran: la de Bill Inmon y la de Ralph Kimball (George, 2012). A continuación en la siguiente tabla se dan a conocer las particularidades de cada una.



CAPÍTULO 1: MARCO TEÓRICO

Tabla 2: Comparación entre Inmon y Kimball

METODOLOGÍA DE INMON	METODOLOGÍA DE KIMBALL
Conocida como una metodología Top-Down. Debido a que propone construir primero el almacén de datos y luego los mercados de datos.	Conocida con metodología Bottom-Up. Debido a que propone construir primeramente los mercados de datos que conformarán el almacén de datos.
El almacén de datos alimentará a los distintos mercados de datos de la organización.	El almacén de datos será la unión de todos los mercados de datos.
Se centrará en toda la información de la empresa por lo que demorará para brindar información que sea de utilidad para la misma.	Permite que se pueda obtener información útil en cuestión de días debido a que solo brindará información departamental.
Es útil para cuando los datos proceden de todos los departamentos de la organización y representan una mayor complejidad.	Es útil para cuando los datos proceden de un número limitado de fuentes.
Es usada cuando se dispone de equipos grandes de especialistas.	Es usada cuando no se dispone de gran experiencia en el tema. Equipos pequeños generalistas.

Teniendo en cuenta la información referente a ambas metodologías se decide escoger la propuesta por Kimball por las siguientes razones:

- Esta metodología posee una clara exposición de las fases y actividades propias de cada una, así como un gran número de ejemplos documentados en los cuales apoyarse dado que se dispone de poca experiencia.
- Existen otros mercados de datos que ya han sido desarrollados con esta metodología para otros departamentos de la FGR.
- Solamente se dispone de la información referente a los departamentos Procesos Penales y Gestión de cuadro y personal de apoyo.
- Los departamentos Procesos penales y Gestión de cuadro y personal de apoyo manejan información importante para la fiscalía, haciendo uso de esta metodología se podrá obtener la implementación de los mercados de datos a corto plazo.

Metodología de Ralph Kimball



CAPÍTULO 1: MARCO TEÓRICO

La metodología se basa en lo que Kimball denomina ciclo de vida dimensional del negocio. Este ciclo de vida está basado en cuatro principios básicos (Kimball, 2004):

- **Centrarse en el negocio:** Hay que concentrarse en la identificación de los Requisitos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio.
- **Construir una infraestructura de información adecuada:** Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de Requisitos de negocio identificados en la empresa.
- **Realizar entregas en incrementos significativos:** Crear el almacén de datos en incrementos entregables. Hay que usar el valor de negocio de cada elemento identificado para determinar el orden de aplicación de los incrementos.
- **Ofrecer la solución completa:** Proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Para comenzar, esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible. También se deberá entregar herramientas de consulta, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación.

La construcción de un almacén de datos es sumamente compleja, y Kimball propone una metodología que permite simplificar esa complejidad. A continuación se describe cada una de las tareas que propone Kimball para el desarrollo de los mercados de datos (Kimball, 2004).

Planificación

En este proceso se determina el propósito del almacén de datos, sus objetivos específicos y el alcance del mismo, los principales riesgos y una aproximación inicial a las necesidades de información.

Análisis de requisitos

La definición de los requisitos es en gran medida un proceso de entrevistar al personal de negocio y técnico. Se debe aprender tanto como se pueda sobre el negocio, los competidores, la industria y los clientes del mismo. A partir de las entrevistas, se identifican los temas analíticos y procesos de negocio. Los temas analíticos agrupan requisitos comunes en un tema común.

Modelado Dimensional

La creación de un modelo dimensional es un proceso dinámico y altamente iterativo. El proceso consiste en varios pasos:

- **Elegir el proceso de negocio:** Consiste en elegir el área a modelar. Esta es una decisión de la dirección, y depende fundamentalmente del análisis de requisitos de información y de los temas analíticos anotados en la etapa anterior.



CAPÍTULO 1: MARCO TEÓRICO

- **Elegir las dimensiones:** Las dimensiones surgen naturalmente de las discusiones del equipo, y facilitadas por la elección del nivel de granularidad y de la matriz de procesos/dimensiones. Las tablas de dimensiones tienen un conjunto de atributos (generalmente textuales) que brindan una perspectiva o forma de análisis sobre una medida en una tabla hechos.
- **Establecer el nivel de granularidad:** La granularidad significa especificar el nivel de detalle máximo de los datos. La elección de la granularidad depende de los requisitos de información del negocio. La sugerencia general es comenzar a diseñar el almacén de datos al mayor nivel de detalle posible, debido a que se podría luego realizar agrupamientos al nivel deseado.
- **Identificar las tablas de hechos y medidas:** El último paso consiste en identificar las medidas que surgen de los procesos de negocios. Las medidas habitualmente se vinculan con el nivel de granularidad y se encuentran en tablas que se denominan tablas de hechos. Cada tabla de hechos tiene como atributos una o más medidas de un proceso organizacional, de acuerdo a los requisitos de información. Un registro contiene una medida expresada en números, sobre la cual se desea realizar una operación de agregación, en función de una o más dimensiones.
- **Matriz de proceso/dimensiones:** Por otra parte, a partir del análisis de los requisitos de información se puede construir una herramienta de la metodología denominada matriz de procesos/dimensiones. Esta matriz tiene en sus filas los procesos de negocios identificados, y en las columnas, las dimensiones identificadas. Cada X en la intersección de las filas y columnas significa que en el proceso de negocio de la fila seleccionada se identifican las dimensiones propuestas.
- **Identificación de atributos de dimensiones y tablas de hechos:** La segunda parte de la sesión inicial de diseño consiste en completar cada tabla con una lista de atributos bien formada. Esta lista o grilla se forma colocando en las filas los atributos de la tabla, y en las columnas la siguiente información: características relacionadas con la futura tabla dimensional del almacén de datos, por ejemplo tipo de datos.
- **Implementar el modelo dimensional detallado:** Este proceso consiste simplemente en completar la información incompleta de los pasos anteriores. El objetivo en general es identificar todos los atributos útiles y sus ubicaciones, definiciones y reglas de negocios asociadas que especifican cómo se cargan estos datos.
- **Prueba del modelo:** Si el modelo ya está estable, lo que se hace habitualmente es probarlo contra los requisitos de información del negocio.



CAPÍTULO 1: MARCO TEÓRICO

- **Revisión y validación del modelo:** Una vez que se tiene confianza plena en el modelo, se ingresa en esta etapa, lo cual implica revisar el modelo con diferentes audiencias, cada una con diferentes conocimientos técnicos y del negocio.

Diseño físico

En esta parte, se contestarán las siguientes preguntas:

¿Cómo se puede determinar cuán grande será el almacén de datos?

¿Cuáles son los factores de uso que llevarán a una configuración más grande y más compleja?

¿Cómo se debe configurar el sistema?

¿Cuánta memoria y servidores se necesitan?

¿Qué tipo de almacenamiento y procesadores?

¿Cómo instalar el software en los servidores de desarrollo, prueba y producción?

¿Qué necesitan instalar los diferentes miembros del equipo de construcción del almacén de datos en sus estaciones de trabajo?

¿Cómo convertir el modelo de datos lógico en un modelo de datos físicos en la base de datos relacional?

¿Cómo conseguir un plan de indexación inicial?

¿Debe usarse la partición en las tablas relacionales?

Diseño del sistema de Extracción, Transformación y Carga (ETL)

El sistema de Extracción, Transformación y Carga por sus siglas en inglés (ETL) es la base sobre la cual se alimenta el almacén de datos. ETL es el proceso que organiza el flujo de los datos entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un almacén de datos, reformatearlos, limpiarlos y cargarlos en otra base de datos, mercado de datos o almacén de datos para analizar.

La primera parte del proceso ETL consiste en localizar y extraer los datos desde los sistemas de origen. Cada sistema separado puede usar una organización diferente de los datos o formatos distintos. La extracción convierte los datos a un formato homogéneo y consolidado para iniciar el proceso de transformación.

La fase de transformación de un proceso de ETL aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados, como pueden ser: filtrado, limpieza, depuración, homogeneización, agrupación de la información y seleccionar campos específicos que se consideren necesarios para el almacén de datos. Por último la fase de carga que es la que se encarga de organizar y actualizar los datos en los mercados de datos.



CAPÍTULO 1: MARCO TEÓRICO

Especificación de aplicaciones de inteligencia de negocios

Las aplicaciones de BI son la cara visible de la inteligencia de negocios: los informes y aplicaciones de análisis proporcionan información útil a los usuarios. Las aplicaciones de BI incluyen un amplio espectro de tipos de informes y herramientas de análisis, que van desde informes simples de formato fijo a sofisticadas aplicaciones analíticas que usan complejos algoritmos e información del dominio.

1.9 Tendencias

En el mundo, empresas que manejan un considerable número de información de sus clientes, operaciones comerciales, mercadeo y productos; utilizan los almacenes de datos para regular y tomar decisiones oportunas en el entorno empresarial. Los sectores donde principalmente se ha implantado almacenes de datos son las empresas de telecomunicaciones, empresas de transporte aerolíneas, transporte de cargas, transporte de pasajeros, empresas de fabricación de bienes de consumo masivo, empresas aseguradoras (Reyes, 2013).

Tendencias actuales de los almacenes de datos en el mundo

A nivel mundial las empresas están cada vez más interesadas en manejar de forma óptima el gran volumen de información que es generado en su accionar diario, debido a las ventajas competitivas que puede proporcionarles.

Tales son los casos de:

- La oficina nacional de estadística (ONE) en la República Dominicana pone a disposición de la ciudadanía el sub-portal del almacén de datos del sistema estadístico nacional (SEN), valiosa herramienta que concentra las estadísticas, publicaciones y bases de datos relevantes, así como indicadores para la toma de decisiones y la elaboración de políticas públicas (Pablo Tactuk , 2013).
- Almacén de datos para la prestación del servicio público de información estadística, Ricardo Luján Salazar, instituto nacional de estadística, geografía e informática de México. El Instituto nacional de estadística, geografía e informática (INEGI) tiene la responsabilidad de coordinar los sistemas nacionales estadístico y de información geográfica de México (Salazar, 2009).

Tendencias actuales de los almacenes de datos en Cuba

Cada día se hace más evidente en el mundo la brecha entre las naciones desarrolladas y subdesarrolladas, debido a los obstáculos para acceder a las tecnologías de punta. A pesar de ello, la isla insiste en desarrollar una base tecnológica endógena con el apoyo de sus profesionales en labor conjunta de todas las empresas. Para el país; anticiparse al desarrollo tecnológico es un reto, es por ello que son varias las empresas que pueden contar con un almacén de datos para ayudar a mantener



CAPÍTULO 1: MARCO TEÓRICO

un historial de su información y apoyar a la toma de decisiones de las mismas, entre estas se encuentran:

- Mercados de datos para el análisis estadístico de la información. Su desarrollo facilita la toma de decisiones en el área de Cooperación intencional de la Universidad de las Ciencias Informáticas, identificando y prediciendo tendencias futuras a partir de datos acumulados. Se obtuvo como resultados un mercado de datos poblado y funcional, con una capa de inteligencia de negocio que facilita el estudio de la información (Pompa, 2014).
- La empresa de proyectos de arquitectura e ingeniería (EMPAI) de Matanzas desarrolló el diseño e implementación de una herramienta que permita gestionar y organizar homogéneamente la información relevante actual e histórica sobre los indicadores de eficiencia de la gestión contable de la empresa, utilizando las posibilidades que brindan los almacenes de datos (Hidalgo-Gato, 2010).

1.10 Herramientas y tecnologías usadas para la realización de los mercados.

Para realizar el modelado del negocio, diseño e implementación y prueba del sistema se hace necesario el uso de herramientas y tecnologías que permitan el desarrollo de una solución robusta, escalable y eficiente. En este epígrafe se muestra un análisis de las herramientas a ser usadas para la realización del almacén de datos.

1.10.1 Herramientas de modelado:

Visual Paradigm en su versión 8.0 (Sierra, 2007): es una herramienta CASE que soporta el modelado mediante UML y proporciona asistencia a los analistas, ingenieros de software y desarrolladores, durante todos los pasos del ciclo de vida de desarrollo de un Software.

Las ventajas que proporciona Visual Paradigm son:

- Dibujo: Facilita el modelado de UML, debido a que proporciona herramientas específicas para ello. Esto también permite la estandarización de la documentación, debido a que la misma se ajusta al estándar soportado por la herramienta.
- Corrección sintáctica: Controla que el modelado con UML sea correcto.
- Coherencia entre diagramas: Al disponer de un repositorio común, es posible visualizar el mismo elemento en varios diagramas, evitando duplicidades.
- Integración con otras aplicaciones: Permite integrarse con otras aplicaciones, como herramientas ofimáticas, lo cual aumenta la productividad.
- Trabajo multiusuario: Permite el trabajo en grupo, proporcionando herramientas de compartición de trabajo.



CAPÍTULO 1: MARCO TEÓRICO

- Reutilización: Facilita la reutilización, debido a que se dispone de una herramienta centralizada donde se encuentran los modelos utilizados para otros proyectos.
- Generación de código: Permite generar código de forma automática, reduciendo los tiempos de desarrollo y evitando errores en la codificación del software.

Se seleccionó el Visual Paradigm 8.0 como herramienta de modelado, teniendo en cuenta la posibilidad que brinda de modelar los mercados de datos y partiendo de estos generar el código para crearlos en el gestor PostgreSQL. Además, en su selección se tuvo en cuenta la existencia de abundante documentación y un amplio uso en el entorno donde se desarrolló la solución, elementos que hacen más simple el proceso de aprendizaje del equipo de desarrollo.

Gestor y administrador de base de datos

PostgreSQL en su versión 9.1: es un sistema de gestión de bases de datos objeto-relacional, distribuido bajo licencia y con código fuente disponible libremente. Es el sistema de gestión de bases de datos de código abierto más potente del mercado. Es multiplataforma. Soporta distintos tipos de datos, además del soporte para los tipos base, también soporta datos de tipo fecha, monetarios, etc. También permite la creación de tipos propios. Utiliza un modelo cliente/servidor y usa multiprocesos en vez de multihilos para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto y el sistema continuará ejecutándose. Permite el trabajo con grandes volúmenes de datos y admite una alta concurrencia de usuarios accediendo al sistema (Martinez, 2013).

Administrador de bases de datos

PgAdmin III en su versión 1.14.1: es una aplicación gráfica para gestionar el gestor de bases de datos PostgreSQL, siendo la más completa y popular con licencia open source (código abierto). Está escrita en C++ usando la librería gráfica multiplataforma wxWidgets, lo que permite que se pueda usar en Linux, FreeBSD, Solaris, Mac OS X y Windows. Es capaz de gestionar versiones a partir de la PostgreSQL 7.3 ejecutándose en cualquier plataforma.

PgAdmin III está diseñado para responder a las necesidades de todos los usuarios, desde escribir consultas SQL simples hasta desarrollar bases de datos complejas. La interfaz gráfica soporta todas las características de PostgreSQL y facilita enormemente la administración. La aplicación también incluye un editor SQL con resaltado de sintaxis, un editor de código de la parte del servidor, un agente para lanzar scripts programados, y mucho más. La conexión al servidor puede hacerse mediante conexión TCP/IP o Unix Domain Sockets (en plataformas *nix) (Ventura, 2014).

Se seleccionó PostgreSQL en su versión 9.1 como gestor de base de datos debido a que permite trabajar con grandes volúmenes de información y sus límites de capacidad de almacenamiento por



CAPÍTULO 1: MARCO TEÓRICO

tablas columnas y como gestor están acorde con las necesidades que impone el desarrollo que se pretende ejecutar. Otros de los elementos de peso en su definición fueron: primero que es una aplicación de código abierto, por lo que va en línea con las políticas de informatización del país y segundo que es el mismo Sistema Gestor de Base de Datos (SGBD) y versión que se utilizó en la implementación de los mercados de datos ya existentes para la FGR lo que contribuye a garantizar la compatibilidad entre los mismos.

1.10.2 Plataforma Pentaho Open Source Business Intelligence

Pentaho es una herramienta de Inteligencia de negocio para la gestión y toma de decisiones empresariales. Es una plataforma compuesta de diferentes programas que satisfacen los requisitos de Inteligencia de negocio. Ofreciendo soluciones para la gestión y análisis de la información, incluyendo el análisis multidimensional OLAP, presentación de informes, minería de datos y creación de cuadros de mando para el usuario. La plataforma ha sido desarrollada bajo el lenguaje de programación Java y tiene un ambiente de implementación también basado en Java, haciendo así que Pentaho sea una solución muy flexible al cubrir una alta gama de necesidades empresariales (Martínez, 2009).

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

Pentaho Bi Server-ce en su versión 4.8.0: es una aplicación 100% Java que permite gestionar todos los recursos de inteligencia de negocio. Cuenta con una interfaz de usuario de BI donde se encuentran disponibles todos los informes, vistas OLAP y cuadros de mando. Así mismo como accesos a una consola de administración que permitirá gestionar y supervisar tanto la aplicación como los usuarios, informes que consulta cada usuario, cuando se han consultado y el rendimiento de la aplicación. La gestión de la configuración, tanto de la instalación inicial como del mantenimiento está muy bien resuelta. Se integra con la mayoría de entornos y se puede comunicar con otras aplicaciones vía servicio web. Integra todos los recursos informacionales en una única plataforma de explotación. Proporciona mucha libertad al usuario y los desarrolladores para crear contenidos nuevos (Martínez, 2009).

Pentaho Data Integration en su versión 4.2.1: Muchas organizaciones tienen información disponible en aplicaciones y base de datos separados. Pentaho Data Integration abre, limpia e integra esta valiosa información y la pone en manos del usuario. Provee una consistencia, una sola versión de todos los recursos de información, que es uno de los más grandes desafíos para las organizaciones hoy en día. Pentaho Data Integration permite una poderosa ETL (Extracción, Transformación y Carga). Las características básicas de esta herramienta son:

Entorno gráfico de desarrollo.

- Uso de tecnologías estándar: Java, XML, JavaScript.



CAPÍTULO 1: MARCO TEÓRICO

- Fácil de instalar y configurar.
- Multiplataforma: windows, macintosh, Linux.
- Basado en dos tipos de objetos: Transformaciones (colección de pasos en un proceso ETL) y trabajos (colección de transformaciones).

Incluye otras herramientas como:

Spoon: para diseñar transformaciones ETL usando un entorno gráfico.

Pan: para ejecutar transformaciones diseñadas con Spoon.

Kitchen: para ejecutar trabajos por línea de comandos (Martínez, 2009).

Mondrian Schema Workbench en su versión 3.2.0: La herramienta permite diseñar una interfaz gráfica capaz de crear y probar esquemas de cubos Mondrian OLAP visualmente. La plataforma de inteligencia de negocio de Pentaho incrusta el motor de consulta Mondrian, como parte de su arquitectura. Mondrian se encarga de recibir consultas multidimensionales y devolver los datos de un cubo. Los archivos XML de metadatos son generados por esta herramienta. Se los podría crear en cualquier editor de texto, pero lo conveniente es que la herramienta genere el XML necesario para que el motor OLAP Mondrian procese el archivo (Méndez, 2013).

Se utilizará para el desarrollo de la solución la suite de Pentaho, debido a que con sus distintas herramientas cubre el flujo de desarrollo propuesto, permite extraer datos desde la fuente de origen que se dispone, por su compatibilidad con PostgreSQL y transformarlos según las necesidades del usuario. Además fue importante en su selección su condición de herramienta de código abierto, elemento que lo alinea con las políticas que para el tema define el estado cubano.

1.10.3 Herramientas utilizadas para realizar las pruebas

Datanamic Data Generator para PostgreSQL en su versión 5.0: Permite poner a prueba su nueva aplicación de base de datos PostgreSQL. Para probar el rendimiento de una aplicación eficaz, debe tener suficientes datos de prueba para exponer posibles problemas de rendimiento de base de datos. Tiene características que ayudarán a generar datos de prueba significativos para su base de datos rápidamente y útiles para la ejecución de pruebas a bases de datos PostgreSQL entre las que se pueden mencionar:

- Acceso directo a bases de datos PostgreSQL (con controlador nativo).
- Rellenar la base de datos directamente o generar scripts de inserción.
- Soporte para PostgreSQL versión 7,8 y 9.
- Incorpora en los generadores más de 40 idiomas.
- Carga datos aleatorios de fuentes tanto externas como internas (Lekstraat, 2011).



CAPÍTULO 1: MARCO TEÓRICO

Apache JMeter 2.9: Clasificada como una herramienta de generación de carga, es utilizada para medir el rendimiento del servidor de aplicaciones. Para dimensionar adecuadamente la memoria del servidor y optimizar los recursos de la máquina (física o virtual) (Sanchez, 2013).

Se seleccionaron las herramientas Datanamic Data Generator y Apache JMeter atendiendo a que las características anteriormente descritas cubren las necesidades que impone el desarrollo propuesto. Otro elemento importante es que existe una amplia documentación disponible de las mismas y son frecuentemente usadas en la universidad, por lo que el aprendizaje que conlleva su utilización es menos costoso en tiempo y esfuerzo para el equipo de desarrollo.

1.11 Pruebas de rendimiento

Las pruebas de rendimiento consisten en someter al sistema a altas cargas de trabajo mediante la simulación de la actividad de los usuarios reales en el sistema. Esta simulación tendrá que ser lo más fidedigna posible para que los resultados sean relevantes y al analizar la información obtenida se logren alcanzar conclusiones acertadas y precisas. La simulación se lleva a cabo utilizando las funcionalidades al alcance de los usuarios, aquellas que sean más usadas, que demanden más recursos, que necesiten más tiempo para ejecutarse, o que, por cualquier motivo, sean interesantes para poner a prueba las capacidades del sistema (Zapata, 2011).

Hay tres tipos de pruebas de rendimiento que se pueden realizar para los almacenes de datos: carga volumen y estrés. La idea de partida, inyectar carga de trabajo en el sistema, es la misma para las tres, la diferencia estriba en el desarrollo de la prueba, su duración y el objetivo que tiene cada una (Zapata, 2011).

Pruebas de carga

Es una prueba que somete al sistema a una carga de trabajo concreta y estable durante un tiempo relativamente corto el cual puede ser de una hora. El nivel de carga debe ser alto y continuo para verificar que el sistema soporta esa carga sin pérdidas de servicio y con un tiempo de respuesta estable y sin degradaciones. Se suelen hacer con varios niveles de carga (100%, 150% y 200%) para conocer el comportamiento del sistema si la carga real llegara a esos niveles (Zapata, 2011).

Prueba de volumen

Este tipo de prueba tiene por objetivo verificar si el sistema es estable durante un largo periodo de tiempo. Básicamente es como una prueba de carga con una duración superior, por ejemplo 24 horas. La idea es encontrar errores acumulativos, es decir, errores que pasan una prueba de esfuerzo por que producen un daño muy pequeño y que a la larga van a terminar deteriorando o colapsando el rendimiento del sistema (Zapata, 2011).



CAPÍTULO 1: MARCO TEÓRICO

Prueba de estrés

Es un tipo de prueba que va buscando el punto de ruptura del sistema, es decir, a qué nivel de carga de trabajo se pierde el servicio y qué pasa en los niveles anteriores. El objetivo de este tipo de pruebas es conocer el límite de carga de trabajo al que se puede llevar a un sistema. Este límite en sí mismo no dice nada, pero hace posible que se puedan tomar medidas antes de llegar a ese nivel (Zapata, 2011).

1.12 Conclusiones del capítulo:

- Se realizó un estudio de las principales metodologías existentes para el desarrollo de los mercados de datos, el cual permitió identificar la metodología de Kimball como la más adecuada para la realización de los mismos, debido a que solo se dispone de la información referente a los departamentos Procesos Penales y Gestión de cuadro y personal de Apoyo.
- Se conocieron las distintas implementaciones OLAP y según las características de cada una se decidió utilizar ROLAP, debido a que utiliza una base de datos relacional soportada por el gestor de base de datos PostgreSQL seleccionado para almacenar la información.
- Se identificaron las herramientas para el desarrollo de los mercados de datos. Teniendo en cuenta sus características y funcionalidades se puede asegurar que son adecuadas para implementar de manera eficiente, la solución propuesta para resolver los problemas que tienen hoy en día las fiscalías del país.



CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DE LOS MERCADOS DE DATOS

Capítulo 2: Diseño e implementación de los mercados de datos

Introducción

En el presente capítulo se llevarán a cabo los pasos descritos por la metodología escogida para la realización de los mercados de datos. Se define la arquitectura que tendrán los mercados de datos. Se analizará el estado actual de las fuentes de datos, la base de datos del SIGEF. Se realizará un estudio acerca de las necesidades de información de los usuarios para identificar los requisitos de información del negocio, lo que será crucial para realizar el modelado y diseño de los mercados de datos. Se identificarán los hechos, medidas y dimensiones para cada uno de los departamentos, teniendo en cuenta las necesidades de información de los usuarios. Posteriormente se realizará la implementación de los mercados de datos para los departamentos Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva y Gestión de Cuadros y Personal de Apoyo.

2.1 Arquitectura de los mercados de datos

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



Figura 11: Arquitectura de los mercados de datos.

La arquitectura mostrada en la figura anterior está formada por varios elementos que interactúan entre sí y que cumplen una función específica que se resumen de la siguiente manera:

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

2.2 Análisis del estado de las fuentes de datos

Para gestionar los datos correspondientes al SIGEF se cuenta con una base de datos relacional y como gestor de base de datos se utiliza PostgreSQL en la versión 9.1.



CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DE LOS MERCADOS DE DATOS

El departamento Índice de Peligrosidad Predelictiva tiene un total de 11 tablas de las cuales 3 son tablas nomencladoras y 8 son tablas de datos, el departamento Procesos Penales Ordinarios tiene un total de 49 tablas de las cuales 9 son tablas nomencladoras y 40 son tablas de datos, el departamento Procesos Penales Sumarios tiene un total de 26 tablas de las cuales 3 son tablas nomencladoras y 23 son tablas de datos y el departamento Gestión de Cuadros y Personal de Apoyo tiene un total de 29 tablas de las cuales 13 son tablas nomencladoras y 16 son tablas de datos. El Esquema base que contiene información importante para los cuatro mercados de datos tiene un total de 203 tablas de las cuales 78 son tablas nomencladoras y 125 son tablas de datos.

2.3 Metodología de Kimball

Seguidamente se le da cumplimiento a los pasos propuestos por la metodología de Kimball para el desarrollo de los mercados de datos de los departamentos Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva y Gestión de Cuadros y Personal de Apoyo.

Planificación

Alcance: Se desarrollarán cuatro mercados de datos para cada uno de los siguientes departamentos: Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva y Gestión de Cuadros y Personal de Apoyo.

Análisis de Requisitos

En esta etapa se identifican las necesidades de información de los usuarios en el desarrollo de la solución. De esta forma se puede conocer: ¿Qué buscan? ¿Qué necesitan? Será de vital importancia dar respuestas a estas preguntas, pues brindarán la posibilidad de que los resultados obtenidos estén en correspondencia con las necesidades. Por esta razón se procede a realizar un análisis de los reportes perteneciente a los departamentos Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva y Gestión de Cuadros y Personal de Apoyo en los cuales se identificaron un total de 48 requisitos de información. A continuación se muestran los requisitos de información del departamento Índice de Peligrosidad Predelictiva. El resto de los Requisitos de información pertenecientes a los departamentos Procesos Penales Sumarios, Procesos Penales Ordinarios y Gestión de cuadro y personal de apoyo se encuentran en el anexo 1.

Requisitos de información para el mercado de datos del departamento Índice de Peligrosidad Predelictiva.

RF1: Total de Expedientes de Índice de Peligrosidad Predelictiva presentados al fiscal según su tipo en un rango de tiempo.

RF2: Total de Expedientes de Índice de Peligrosidad Predelictiva promovidos al tribunal según su tipo en un rango de tiempo.



CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DE LOS MERCADOS DE DATOS

RF3: Total de Expedientes de Índice de Peligrosidad Predelictiva con comparecencias celebradas según su tipo en un rango de tiempo.

RF4: Total de Expedientes de Índice de Peligrosidad Predelictiva con medidas de detención solicitadas por el fiscal según su tipo en un rango de tiempo.

RF5: Total de Expedientes de Índice de Peligrosidad Predelictiva con apelaciones celebradas según su tipo en un rango de tiempo.

RF6: Total de Expedientes de Índice de Peligrosidad Predelictiva devueltos a la PNR según su tipo en un rango de tiempo.

RF7: Total de Expedientes de Índice de Peligrosidad Predelictiva devueltos a la fiscalía según su tipo en un rango de tiempo.

Requisitos no Funcionales (Blanco, 2012)

Los requisitos no funcionales son propiedades o cualidades que el sistema debe cumplir para lograr el correcto funcionamiento de la solución. A continuación se muestran los requisitos no funcionales identificados:

Usabilidad: Los mercados de datos deben garantizarle al usuario el rápido acceso a los datos contenidos en las distintas áreas de análisis. De esta manera el usuario podrá analizar de manera rápida la información que solicita.

Eficiencia: En cuanto a los tiempos de respuesta para acceder a la información contenida en los mercados de datos, serán como promedio de diez segundos.

Para un trabajo óptimo de los mercados de datos se desglosan los recursos tanto de software como de hardware necesarios.

- **Servidores**

Memoria RAM 4 GB o superior.

Disco duro de 1 TB o superior.

Intel(R) Core(TM) i3-2120 CPU @ 3.30 GHz 3.30 GHz o superior.

Tarjeta de red Ethernet.

- **Software PC Servidor**

Deber tener instalado Postgres 9.1

Confiabilidad:

- **Disponibilidad del sistema**

El sistema estará disponible durante el horario laboral, efectuándose en períodos de tiempos definidos el proceso de actualización de la información del servidor local de base de datos con el Centro de Datos.



CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DE LOS MERCADOS DE DATOS

- **Exactitud de las respuestas**

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

Soporte:

- **Escalabilidad**

Por su naturaleza los almacenes de datos desarrollados con la metodología de Kimball son escalables pues están compuestos por mercados de datos que se van adicionado según las necesidades o momento del desarrollo, este fue un elemento que se tuvo en cuenta en la definición de la metodología. Los mercados de datos desarrollados son una muestra de esta característica y a su vez los mismos al centrarse en un negocio específico de una sola área, facilitan su entendimiento y con ello la posibilidad de crecimiento.

Seguridad:

- **Control de acceso basado en roles**

Para garantizar la seguridad se define el acceso a los mercados de datos por roles, por lo que se decide utilizar el modelo RBAC (control de acceso basado en roles). Se le asignan permisos a los roles sobre la base de datos y se asocian los usuarios correspondientes a cada rol. Así se asegura un control de acceso basado en roles donde un usuario dispone exactamente de la cantidad de privilegios necesaria para acceder a la información almacenada en los mercados de datos.

- **Configuración segura del pg_hba.conf**

El fichero pg_hba.conf de postgres SQL se utiliza para definir cómo, dónde y desde qué sitio un usuario puede acceder a la base de datos, de una buena configuración o no, dependerá su seguridad. Este fichero se configura con los parámetros correspondientes a las diferentes instancias de la fiscalía para que puedan acceder a la base de datos.

Tablas de temas analíticos

Para determinar la información clave de las Fiscalías a partir de los requisitos anteriores en los departamentos Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva y Gestión de Cuadros y Personal de Apoyo, se reunieron a los analistas del SIGEF y se seleccionaron los procesos de negocio importantes y los temas analíticos de dichos departamentos que a continuación se muestran en las siguientes tablas.

Tabla 3: Temas analíticos departamento Procesos Penales Sumarios

Tema Analítico	Proceso de Negocio	Requerimiento pedido o inferido	Comentario
-----------------------	---------------------------	--	-------------------



CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DE LOS MERCADOS DE DATOS

Procesos Penales Sumarios	Denuncias recibidas en la Fiscalía en el período.	Análisis de las reclamaciones recibidas, remitidas y devueltas en la fiscalía.	Por provincias y municipios de la fiscalía, períodos de tiempo, devoluciones y acciones tomadas.
---------------------------	---	--	--

Las demás tablas correspondientes a los mercados de datos Procesos Penales Ordinarios, Índice de Peligrosidad Predelictiva, Gestión de cuadro y personal de apoyo se encuentran en el anexo 2

Modelado Dimensional

El modelado dimensional, es la forma en que se transforman los datos y son convertidos en información útil para los usuarios del negocio. El objetivo final es que estos puedan encontrar de manera intuitiva y rápida la información que necesitan. Para la creación del modelo se definirán los procesos del negocio a los cuales se les realizarán los mercados de datos. Se identificarán las dimensiones y se le establecerá el nivel de granularidad. Posteriormente se describirán las tablas dimensiones y hechos.

Elegir el proceso de negocio

Las áreas de proceso a modelar serán los procesos de negocio que pertenecen a los departamentos Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva y Gestión de Cuadros y Personal de Apoyo.

Identificación de las dimensiones

Las dimensiones se obtuvieron mediante el análisis de los reportes de los departamentos Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva y Gestión de Cuadros y Personal de Apoyo y de la base de datos del SIGEF.

Dimensiones comunes para los mercados de datos de los departamentos Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva.

Dimensión tiempo: Representa el espacio de tiempo en el que se está analizando cada acción realizada en la fiscalía, en este caso se analizarán los documentos diarios y en rangos de fechas.

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

Dimensión indicador: Define el tipo de documento y los indicadores por los que se desea analizar.

Dimensión provincia: Contiene todas las provincias de Cuba.

Dimensiones para el mercado de datos del departamento Gestión de Cuadros y Personal de Apoyo.

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



CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DE LOS MERCADOS DE DATOS

Dimensión provincia: Contiene todas las provincias de Cuba.

Dimensión cargo: Define los cargos que se pueden ocupar en la fiscalía.

Establecimiento de granularidad

La granularidad representa el nivel de detalle en el que se desea almacenar la información sobre el negocio. Mientras mayor sea el nivel de detalle de los datos, se tendrán mayores posibilidades de análisis (Avella, 2011).

A continuación se establece la granularidad para las dimensiones comunes de los mercados de datos realizado a los departamentos de Índice de Peligrosidad Predelictiva, Procesos Penales Ordinarios y Procesos Penales Sumarios, especificando el nivel de detalle de los datos. En la tabla se muestra la descripción de los datos asociados a las dimensiones, donde se explica el tipo de dato y una breve descripción de los valores que pueden tomar estos.

Tabla 4: Nivel de granularidad para las tablas de dimensiones tiempo, instancia, indicador y provincia

Nombre de la dimensión.	Nombre del dato.	Tipo de dato.	Valor nulo.	Descripción.
dim_tiempo	id_tiempo	numeric(19,0)	No	Identificador de la tabla dim_tiempo.
	dia	numeric(2, 0)	No	Día en que se realiza una acción.
	mes	varchar(255)	No	Mes en que se realiza una acción.
	anno	numeric(4, 0)	No	Año en que se realiza una acción.
	fecha	date	No	Fecha en que se realiza el proceso.
dim_instancia	id_instancia	numeric(19,0)	No	Identificador de la tabla dim_instancia.
	descripcion	varchar(255)	No	Define las diferentes instancias de la fiscalía en el país.
dim_indicador	id_indicador	numeric(19,0)	No	Identificador de la tabla dim_indicador.
	nombre_reporte	varchar(255)	No	Define el tipo de expediente.



CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DE LOS MERCADOS DE DATOS

	especificacion	varchar(255)	No	Representa el indicador por el cual se va a analizar los expedientes.
dim_provincia	id_provincia	numeric(19, 0)	No	Identificador de la tabla dim_provincia.
	provincia	varchar(255)	No	Contiene el nombre de las provincias de Cuba.
	codigo_provincia	int4	No	Contiene el código de las provincias.

En el anexo 3 se muestra el nivel de granularidad del mercado de datos realizado para el departamento de Gestión de Cuadros y Personal de Apoyo.

Identificar las tablas de hechos

Los hechos proporcionan información sobre el historial de las operaciones del negocio. Cada tabla de hechos posee un nombre que la identifica y contiene las claves de todas las tablas de dimensiones con las que se relaciona. Estas tablas se obtuvieron mediante el análisis de los reportes de los departamentos Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva y Gestión de Cuadros y Personal de Apoyo.

Hechos para el mercado de datos Índice de Peligrosidad Predelictiva:

Expedientes de índice de peligrosidad predelictiva devueltos a la fiscalía: Almacena todos los datos cuantitativos relacionados con los expedientes de índice de peligrosidad predelictiva, devueltos a la fiscalía en las diferentes instancias de la fiscalía según sus indicadores en un período de tiempo.

Expedientes de índice de peligrosidad predelictiva devueltos a la PNR: Almacena todos los datos cuantitativos relacionados con los expedientes de índice de peligrosidad predelictiva devueltos a la PNR en las diferentes instancias de la fiscalía según sus indicadores en un período de tiempo.

Expedientes de índice de peligrosidad predelictiva presentados al fiscal: Almacena todos los datos cuantitativos relacionados con los expedientes de índice de peligrosidad predelictiva presentados al fiscal en las diferentes instancias de la fiscalía según sus indicadores en un período de tiempo.

Expedientes de índice de peligrosidad predelictiva promovidos al tribunal: Almacena todos los datos cuantitativos relacionados con los expedientes de índice de peligrosidad predelictiva promovidos al tribunal en las diferentes instancias de la fiscalía según sus indicadores en un período de tiempo.

Expedientes de índice de peligrosidad predelictiva con comparencias celebradas: Almacena todos los datos cuantitativos relacionados con los expedientes de índice de peligrosidad predelictiva



CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DE LOS MERCADOS DE DATOS

con comparecencias celebradas en las diferentes instancias de la fiscalía según sus indicadores en un período de tiempo.

Expedientes de índice de peligrosidad predelictiva con apelaciones celebradas: Almacena todos los datos cuantitativos relacionados con los expedientes de índice de peligrosidad predelictiva con apelaciones celebradas en las diferentes instancias de la fiscalía según sus indicadores en un período de tiempo.

Expedientes de índice de peligrosidad predelictiva con medidas de detención solicitadas por el fiscal: Almacena todos los datos cuantitativos relacionados con los expedientes de índice de peligrosidad predelictiva con medidas de detención solicitadas por el fiscal en las diferentes instancias de la fiscalía según sus indicadores en un período de tiempo.

Los hechos correspondientes a los mercados de datos Procesos Penales Sumario, Procesos Penales Ordinarios y Gestión de cuadro y personal de apoyo se encuentran en el anexo 4.

Tabla de matriz procesos/dimensiones

Por cada proceso de negocio identificado para cada departamento existe una tabla de hechos. A partir del análisis de las tablas anteriores se puede construir la herramienta de la metodología denominada matriz de procesos (hechos)/dimensiones que se muestra en la siguiente tabla.

Esta matriz tiene en sus filas los hechos correspondientes a los procesos de negocios identificados, y en las columnas, las dimensiones identificadas. Cada X en la intersección de las filas y columnas significa la relación que existe entre los hechos y las dimensiones.

Hechos	Dimensiones			
	Tiempo	Instancia	Indicador	Provincia
Expedientes de Índice de Peligrosidad Predelictiva devueltos a la fiscalía.	X	X	X	X
Expedientes de Índice de Peligrosidad Predelictiva devueltos a la PNR.	X	X	X	X
Expedientes de Índice de Peligrosidad Predelictiva presentados al fiscal.	X	X	X	X
Expedientes de Índice de Peligrosidad Predelictiva promovidos al tribunal.	X	X	X	X
Expedientes de Índice de Peligrosidad Predelictiva con comparecencias celebradas.	X	X	X	X
Expedientes de Índice de Peligrosidad Predelictiva con apelaciones celebradas.	X	X	X	X



CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DE LOS MERCADOS DE DATOS

Expedientes de Índice de Peligrosidad Predelictiva con medidas de detención solicitadas por el fiscal.	X	X	X	X
--	---	---	---	---

Tabla 5: Matriz de procesos/dimensiones Índice de Peligrosidad Predelictiva.

En el anexo 5 se muestra las matrices de procesos/dimensiones para los restantes mercados de datos.

Identificación de atributos de dimensiones y tablas de hechos

En la tabla 6 se muestran los atributos de las tablas de dimensiones comunes para los mercados de datos de los departamentos Procesos Penales Ordinarios, Procesos Penales Sumarios, Índice de Peligrosidad Predelictiva.

Tabla 6: Lista de atributos de las tablas de dimensiones.

Nombre de la dimensión	Atributos
dim_tiempo	id_tiempo, dia, mes, anno, fecha
dim_instancia	id_instancia, descripcion
dim_indicador	id_indicador, nombre_reporte, especificacion
dim_provincia	id_provincia, provincia, codigo_provincia

Para el departamento Gestión de Cuadros y Personal de Apoyo los atributos de las tablas de dimensiones son los siguientes:

Tabla 7: Lista de atributos de las tablas de dimensiones de Gestión de Cuadros y Personal de Apoyo.

Nombre de la dimensión	Atributos
dim_instancia	id_instancia, descripcion
dim_cargo	id_cargo, descripcion
dim_provincia	id_provincia, provincia, codigo_provincia

En el anexo 6 se muestra las tablas de hechos con sus respectivos atributos de los mercados de datos.

Patrones de diseño de bases de datos para los mercados de datos.

En la actualidad las bases de datos suelen ser muy grandes; el trabajo con los patrones de diseño hacen que el trabajo sea más fácil, además asegura un resultado correcto. Estos patrones son utilizados mayoritariamente en los almacenes de datos. Son la guía para el desarrollo de soluciones a problemas comunes. Brindan una solución ya probada y documentada a problemas de desarrollo de software que están sujetos a contextos similares (Tedeschi, 2014).

Patrón de llaves subrogadas

Este patrón es muy utilizado pues se decide generar una llave primara única para cada tabla de dimensión en vez de usar un atributo identificador en el contexto dado. Se definen artificialmente y son



CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DE LOS MERCADOS DE DATOS

de tipo numérico debido a que están demostrados que no se repiten o con una probabilidad extremadamente baja. No tienen relación directa con ningún dato y no poseen ningún significado en especial. Esta clave debe ser el único campo que sea clave principal de cada tabla de dimensión. Una forma de implementación sería, a través de la utilización de herramientas ETL (Blaha, 2010).

A continuación se muestra en la figura un fragmento de una tabla de hechos del mercado de datos para el departamento de Gestión de Cuadros y Personal de Apoyo donde se evidencia la utilización del patrón de llave subrogada.

	id_asignaturas_asistente_fiscal [PK] numeric(19,0)	id_provincia numeric(19,0)	id_instancia numeric(19,0)	asig_vencidas integer	asig_pendientes integer
1	1	1	183	1	2
2	2	15	118	1	2
3	3	15	122	1	2
4	4	2	151	1	2
5	5	16	173	1	2

Figura 12: Ejemplo del uso del patrón de llave subrogada.

Patrón de constelación de hechos y patrón en estrella

Para el modelado de los mercados de datos de los departamentos Gestión de Cuadros y Personal de Apoyo, Procesos Penales Ordinarios, Índice de Peligrosidad Predelictiva se utiliza el patrón constelación de hechos, debido a que estos contienen más de una tabla de hechos. En el modelado del mercado de datos del departamento Procesos Penales Sumarios se selecciona el esquema en estrella debido a que la información se almacena en una sola tabla de hechos. En la figura 13 se muestra el modelo de datos físico diseñado para el departamento Índice de Peligrosidad Predelictiva, donde se evidencia el uso del patrón constelación de hechos y en la figura 14 se muestra el modelo de datos físico diseñado para el departamento Procesos Penales Sumarios, donde se evidencia el uso del patrón esquema en estrella.

Estándar de codificación

Cuando de trabajo en grupo se trata los códigos se vuelven muy variables debido a que todos programan de forma distinta, cada persona lo hace a su manera. Es innegable que en cualquier proyecto es sumamente importante que sus integrantes establezcan acuerdos en la forma de hacer el trabajo; más aún cuando son muchos integrantes trabajando en un mismo tema. Los estándares de codificación sirven como punto de referencia para los desarrolladores, y ayuda a mantener un estilo único de programación y ayudan a mejorar el proceso de codificación haciéndolo más eficiente en especial cuando de trabajo en grupo se trata.

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



CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DE LOS MERCADOS DE DATOS

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

Diseño físico de los mercados de datos

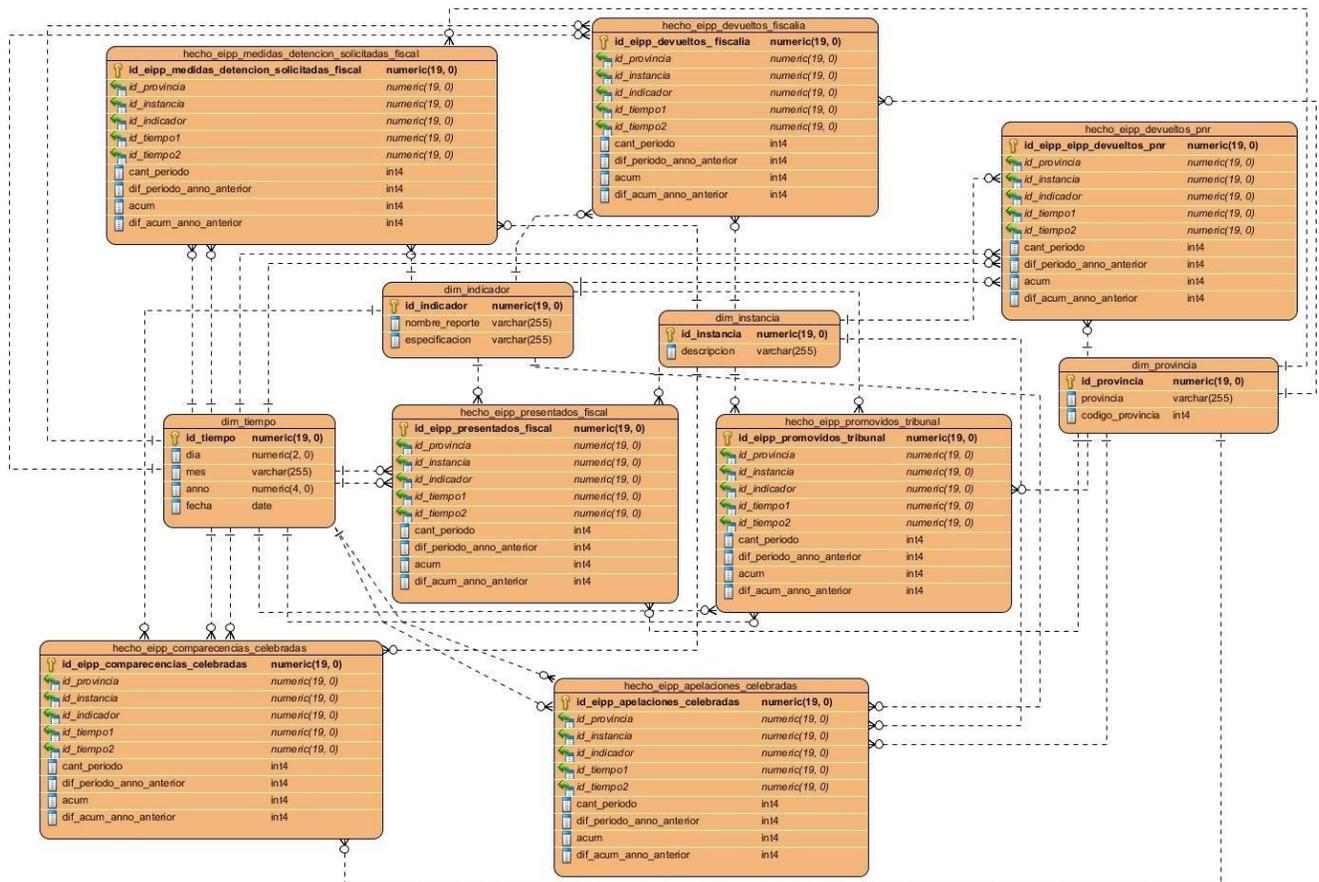


Figura 13: Modelo físico de datos para el departamento de Índice de Peligrosidad Predelictiva.



CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DE LOS MERCADOS DE DATOS

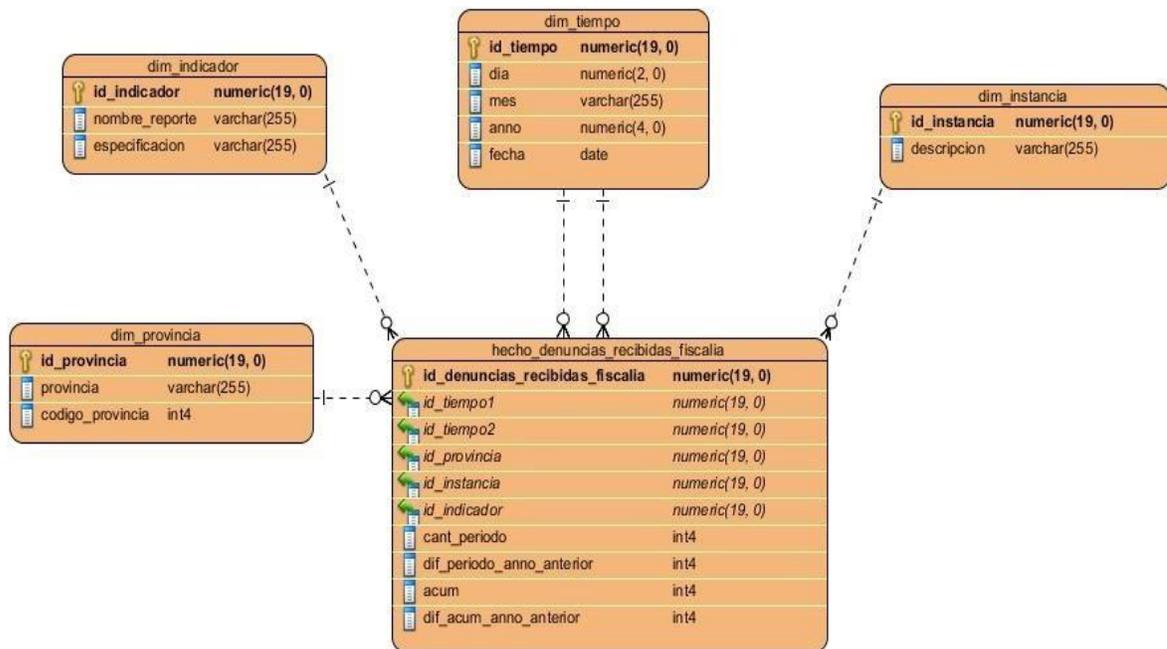


Figura 14: Modelo físico de datos para el departamento de Procesos Penales Sumario. Los modelos físicos para los demás mercados de datos se encuentran en el anexo 5.

Prueba y revisión del modelo

Las pruebas son los procesos que se le realizan al producto para verificar su correcto funcionamiento y calidad. Se le realizan pruebas al modelo haciendo preguntas relacionadas con el negocio para comprobar si la respuesta que devuelve es correcta y muestra la información que se necesita. Un ejemplo de estas preguntas sería:

¿Cuántos Expedientes de Índice de Peligrosidad Predelictiva se han presentado al fiscal por tipo de narcomanía en el periodo del '01-01-2014' al '15-01-2014' en la fiscalía municipal de Matanza?

¿Cuántos Expedientes de Índice de Peligrosidad Predelictiva se han promovido al tribunal por tipo de embriaguez habitual con respecto al año anterior en el periodo del '01-01-2014' al '15-01-2014' en la fiscalía municipal de Matanza?



CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DE LOS MERCADOS DE DATOS

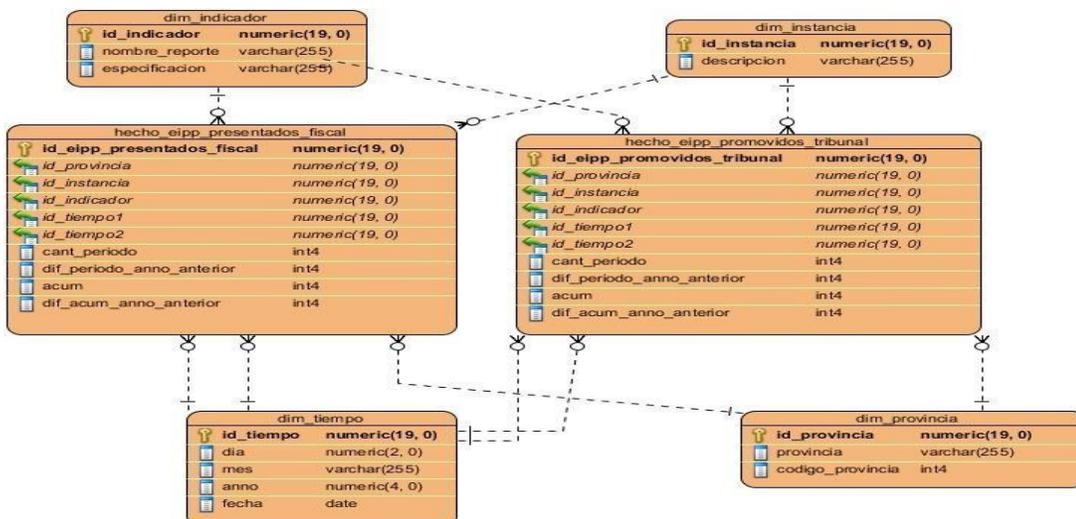


Figura 15: Fragmento del modelo físico para el departamento de Índice de Peligrosidad Predelictiva.

Como se puede observar en la figura la tabla hecho_eipp_presentados_fiscal contiene la medida cant_periodo que almacena la cantidad de expedientes. Esta tabla de hecho tiene relación con la tabla dim_indicador, la cual contiene los tipos de expedientes (en este caso narcomanía), con la tabla dim_tiempo que almacena los rangos de días, con la tabla dim_provincia que almacena las provincias del país y con la tabla dim_instancia que contiene las fiscalías del país.

La tabla hecho_eipp_promovidos_tribunal contiene la medida dif_periodo_anno_anterior que almacena la diferencia de expedientes entre el año actual y el anterior en un rango de fecha. Esta tabla de hecho tiene relación con la tabla dim_indicador, la cual contiene los tipos de expedientes (en este caso embriaguez habitual), con la tabla dim_tiempo, que almacena todas las fechas, con la tabla dim_provincia que almacena las provincias del país y con la tabla dim_instancia que contiene las fiscalías del país.

Correspondencia de datos

Con el fin de evitar la pérdida de información y lograr una mejor organización en el trabajo se realiza una correspondencia entre las tablas de dimensiones y las fuentes de datos. En la siguiente tabla se indica de donde se extrajeron los datos de la base de datos de SIGEF y en las tablas de dimensiones que se insertaron.

Tabla 8: Correspondencia de datos

Tabla de dimensión	Fuente de dato
dim_indicador	reportes.rpt_config
dim_instancia	reportes.v_dpa
dim_provincia	reportes.v_dpa



dim_cargo

reportes.v_edad_cargo

Diseño del sistema de Extracción, Transformación y Carga de los datos

Para comenzar el diseño de las transformaciones de los datos fue necesario realizar antes una configuración de la conexión a la base de datos fuente y el mercado de datos destino. La cual se puede observar en el anexo 8.

El proceso de ETL comienza con la extracción de los datos desde las fuentes de origen que se encuentran en la base de datos SIGEF. Basándose en las necesidades de información de la fiscalía, se explora la fuente de datos disponible, y se extrae la información que se considera relevante. Para la realización de este proceso se crearon un conjunto de consultas SQL que dan cumplimiento a este objetivo. En los anexo 9, 10,11 y 12 se muestra las consultas SQL realizadas.

Una vez que se extraen los datos se realizan las transformaciones, elemento importante dentro de la implementación del proceso ETL. Entre los pasos más comunes que se realizaron se encuentran: entrada de tabla, búsquedas en BD e insertar/actualizar. A continuación se muestra el diseño de la transformación para la dimensión instancia, común para los cuatro mercados de datos (figura 16). Además se muestra la transformación para el hecho denuncias recibidas en la fiscalía del mercado de datos para el departamento Procesos Penales Sumario (figura 17).

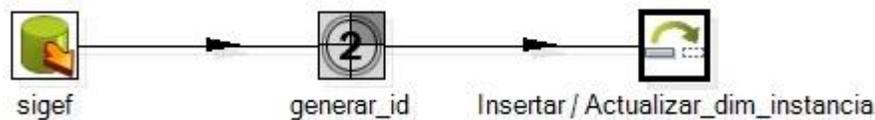


Figura 16: transformación para la dimensión instancia.

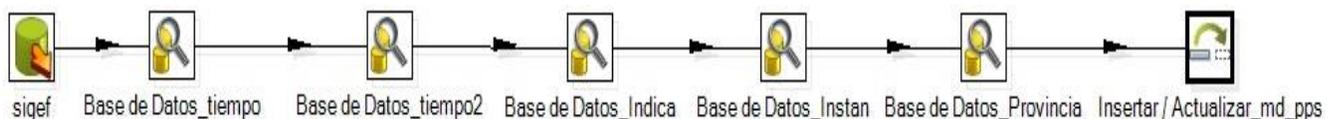


Figura 17: Fragmento de transformación para el hecho denuncias recibidas en la fiscalía.

El diseño de las transformaciones para los demás mercados de datos se muestra en los anexos 13, 14, 15, 16 y 17 mediante Figuras.

Automatización del proceso ETL

El último de los subprocesos de ETL es la carga de la información desde el almacenamiento intermedio hasta el depósito de datos, se diseña un trabajo de automatización, donde se cargarán los datos a las tablas dimensiones y luego a la tabla de hechos. Entre los pasos más comunes que se realizaron se encuentran: inicio de un trabajo y ejecutar una transformación.



CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DE LOS MERCADOS DE DATOS

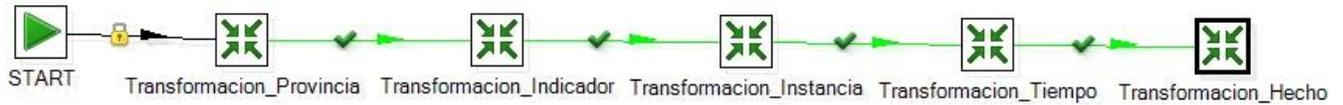


Figura 18: Trabajo para el departamento Procesos Penales Sumario.

Los demás trabajos para los restantes mercados de datos se muestran en los anexos 13, 14, 15, 16 y 17.

Diseño de las dimensiones y cubos de información

Una vez concluido el proceso de extracción, transformación y carga de los mercados de datos, se realiza el diseño multidimensional en la herramienta Schema-Workbench. Para ello se establece la conexión a los mercados de datos que se puede apreciar en el anexo 18.

Primeramente se diseñan los cubos multidimensionales y dentro de estos las medidas y dimensiones correspondientes a cada uno definiendo para esta última las jerarquías y niveles.

En la figura 19 se puede observar la creación del cubo multidimensional para el mercado de datos del departamento Procesos Penales Sumario en el Schema-Workbench, las figuras de los demás cubos multidimensional de los restantes mercados de datos se pueden ver en el anexo 19.

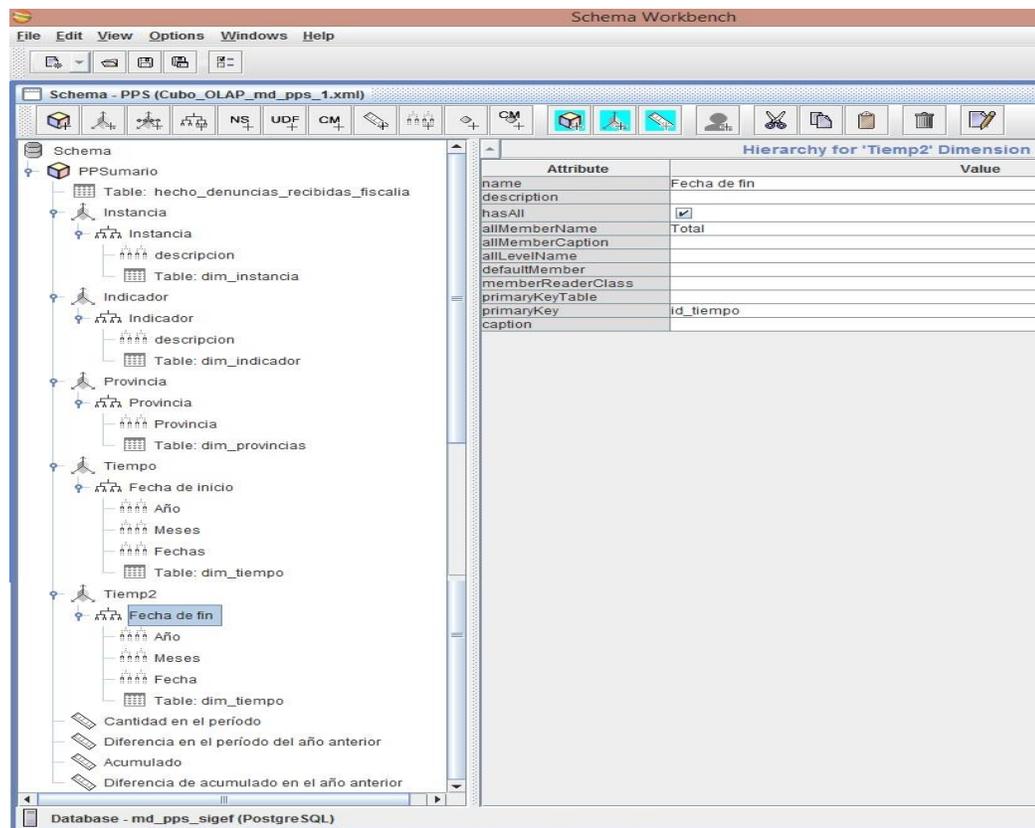


Figura 19: Esquema multidimensional para el mercado de datos del departamento Procesos Penales Sumarios



CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DE LOS MERCADOS DE DATOS

Visualización de los resultados

Los cubos multidimensionales creados se publican en la herramienta Bi Server-ce donde se especifican que dimensiones y medidas se deben mostrar. De esta forma el usuario puede ver la información almacenada en el mercado de datos. A continuación se muestra como se visualizan estos datos en el navegador para el cubo Procesos Penales Sumario

Provincias	Fiscalías	Indicadores	Fecha inicio	Fecha fin	Medidas			
					● Cantidad en el Período	● Diferencia en el período del año anterior	● Acumulado	● Diferencia de acumulado en el año anterior
[-] Total	[-] Total	[+] Total	[+] Total	[+] Total	1.220.218	2.410.307	127.897	3.630.283
Matanzas	[-] Total	[+] Total	[+] Total	[+] Total	1.220.218	2.410.307	127.897	3.630.283
	Fiscalía Municipal Matanzas	[+] Total	[+] Total	[+] Total	619.063	1.207.997	127.897	1.826.818
		1.0 Total de denuncias recibidas en la Fiscalía en el período	[+] Total	[+] Total	68.781	134.216	14.210	202.971
			[-] 2014	[+] Total	68.781	134.216	14.210	202.971
			[-] enero	[+] Total	12.836	22.326	14.210	35.136
			2014-01-01	[+] Total	2.351	1.356	14.210	3.681
				[-] 2014	2.351	1.356	14.210	3.681
				[-] enero	245	123	1.617	376
				[-] febrero	249	137	1.473	362
				2014-02-01	22	7	90	17
				2014-02-02	8	9	61	13
				2014-02-03	12	9	57	16
				2014-02-04	8	2	53	4
				2014-02-05	14	8	27	31

Figura 20: Fragmento de visualización de los datos del cubo Procesos Penales Sumarios para el mercado de datos del departamento Procesos Penales Sumarios.

La visualización de los demás cubos multidimensionales para los restantes mercados de datos se muestra en el anexo 20 mediante figuras.

2.4 Conclusiones del capítulo

- Se detallaron aspectos relacionados con el diseño como los estándares y patrones, concluyendo utilizar el patrón constelación de hechos y el patrón esquema en estrella, para el modelado de las bases de datos multidimensionales, debido a que son los adecuados según las necesidades de información.
- Se definió la arquitectura que tendrán los mercados de datos. Identificándose 5 dimensiones y 23 hechos, además de las uniones establecidas entre las mismas que dieron como resultado el diseño de los mercados de datos, permitiendo la implementación del proceso de extracción, transformación y carga de los datos.
- Se realizó el diseño multidimensional de los datos haciendo uso de la herramienta Workbench para facilitar el análisis de la información almacenada en los mercados de datos a los usuarios del negocio.



Capítulo 3: Pruebas de Validación

Introducción

Las pruebas son un elemento importante para los sistemas de almacenamiento de datos, debido a que ayudan a detectar problemas que puedan surgir cuando el sistema se encuentre en explotación. Para garantizar la calidad de los mercados de datos se realizarán pruebas de rendimiento tales como: pruebas de carga, volumen y estrés, las cuales validarán el correcto funcionamiento de los mercados de datos.

3.1 Pruebas de rendimiento

Prueba de volumen

Para la prueba de volumen se utilizó la herramienta Datanamic Data Generator para PostgreSQL en su versión 5.0. Esta herramienta permitió llenar los mercados de datos con una cantidad determinada de información en función de los pronósticos de crecimiento en cada mercado, para comprobar que pueden manejar esta carga y funcionar adecuadamente con el máximo de datos esperados en las fiscalías. A continuación en la siguiente figura se muestra el resultado de esta prueba para el mercado de datos del departamento de Procesos Penales Sumarios. Los resultados para los demás mercados de datos se encuentran en el anexo 21.

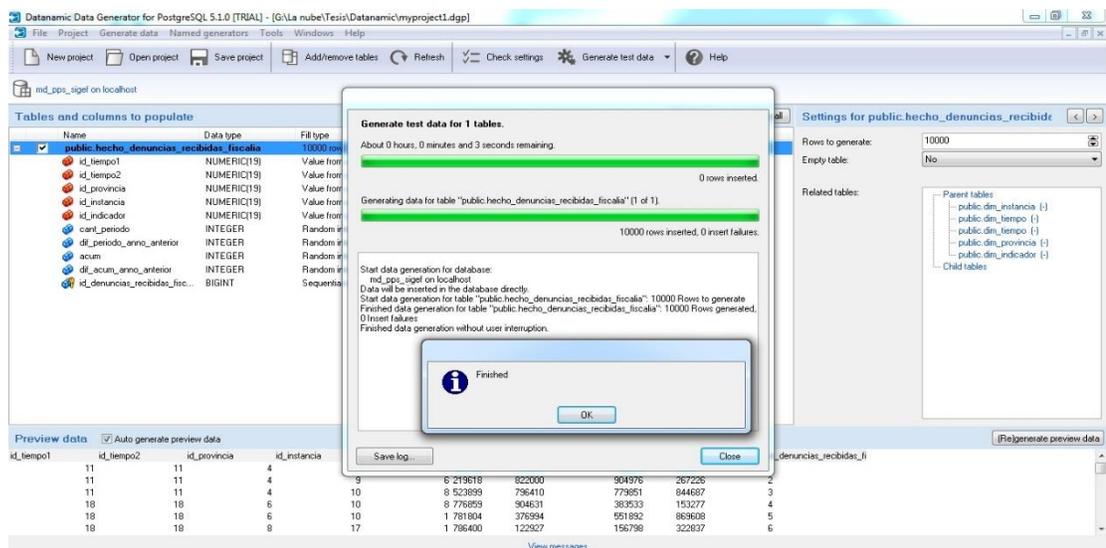


Figura 21: Prueba de volumen para el mercado de datos para el departamento Procesos Penales Sumarios.

Una vez concluida la prueba de volumen los resultados obtenidos para los cuatro mercados de datos fueron satisfactorios. No se presentaron problemas ni errores al introducir los datos, por lo que se



PRUEBAS DE VALIDACIÓN

puede asegurar que los mercados de datos son capaces de soportar la carga de trabajo generada en las fiscalías.

Pruebas de carga

Las pruebas de carga consisten en simular un plan de pruebas con una cantidad de usuarios conectados concurrentemente haciéndole peticiones al mercado de datos. Para ello se realiza una consulta a cada mercado de datos (que significan las peticiones de los usuarios) y se establece la cantidad de usuarios conectados concurrentemente con la herramienta JMeter en su versión 2.9, con el objetivo de comprobar si el tiempo de respuesta de los mercados de datos a dichas consulta es aceptable o no.

La cantidad de usuarios que tendrán los mercados de datos en los distintos niveles de la FGR son los siguientes:

Tabla 9: Representación de las instancias y su respectiva cantidad de usuarios.

Instancia	Cantidad de usuarios
Fiscalía General de la República	25
Fiscalías provinciales	12
Fiscalías municipales	7

Se eligen como casos más críticos para realizar las pruebas la cantidad de usuario que existen en las fiscalías provinciales y en la Fiscalía General de la República, por ser los tipos de entidades donde puede acumularse mayor carga de trabajo.

El rango de tiempo aceptable y no aceptable definido para las consultas es el siguiente:

Tabla 10: Representación del tiempo de respuesta de las consultas.

Tipo de respuesta	Tiempo
Aceptable	Menor o igual a los 10 segundos
No aceptable	Mayor a los 10 segundos

Las consultas utilizadas para los mercados de datos se muestran a continuación:

Índice de Peligrosidad Predelictiva:



PRUEBAS DE VALIDACIÓN

```
SELECT cant_periodo, dif_periodo_anno_anterior, acum, dif_acum_anno_anterior FROM
hecho_eipp_apelaciones_celebradas where id_tiempo1= 1 and id_tiempo2=365 and id_indicador>=10
and id_indicador<=18 and id_instancia=43
```

Proceso Penales Sumario:

```
SELECT cant_periodo, dif_periodo_anno_anterior, acum, dif_acum_anno_anterior FROM"
hecho_denuncias_recibidas_fiscalia " where id_tiempo1= 1 and id_tiempo2=365 and id_indicador>=1
and id_indicador<=8 and id_instancia=43
```

Gestión de Cuadros y Personal de Apoyo:

```
SELECT total_primera, total_segunda, total_tercera FROM hecho_cantidad_fiscales_categoria;
```

Procesos Penales Ordinarios:

```
SELECT cant_periodo, dif_periodo_anno_anterior, acum, dif_acum_anno_anterior FROM
hecho_efp_devuelto_organo_instruccion where id_tiempo1= 1 and id_tiempo2=365 and
id_indicador>=16 and id_indicador<=30 and id_instancia=43
```

El reporte resumen del JMeter define los siguientes campos (Baedke, 2015):

Tabla 11: Representación de los campos de la herramienta JMeter.

Campos	Significado
Etiqueta	Nombre de la etiqueta de muestra
# Muestras	Cantidad de peticiones con la misma etiqueta
Media	El tiempo promedio de respuesta del conjunto de peticiones en milisegundos.
Min	El menor tiempo de respuesta de una petición en milisegundos
Max	El mayor tiempo de respuesta de una petición en milisegundos.

Seguidamente se muestran los resultados de la prueba de carga realizada para el mercado de datos perteneciente al departamento de Gestión de Cuadros y Personal de Apoyo, los resultados de esta prueba para los demás mercados de datos se encuentran en los anexos 22 y 23.



PRUEBAS DE VALIDACIÓN

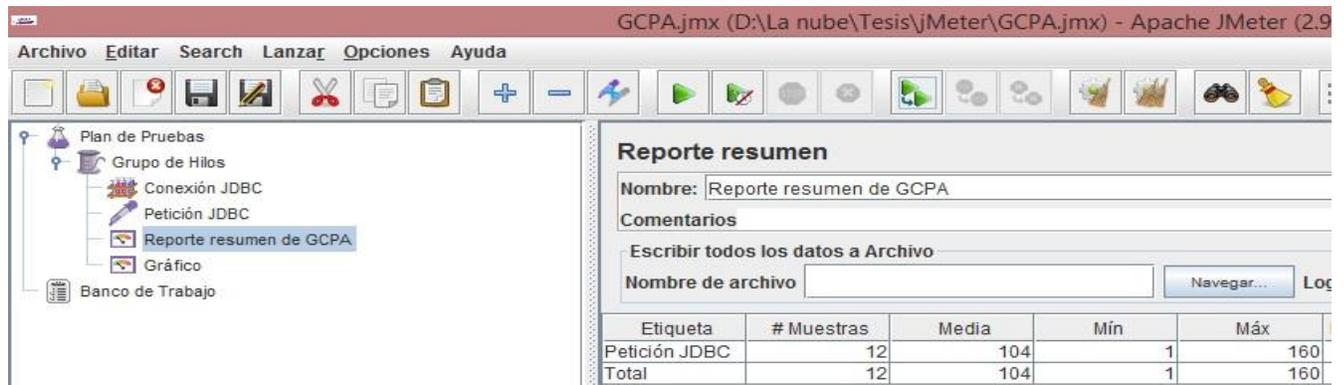


Figura 22: Resultados del plan de pruebas simulando para la Fiscalía Provincial.

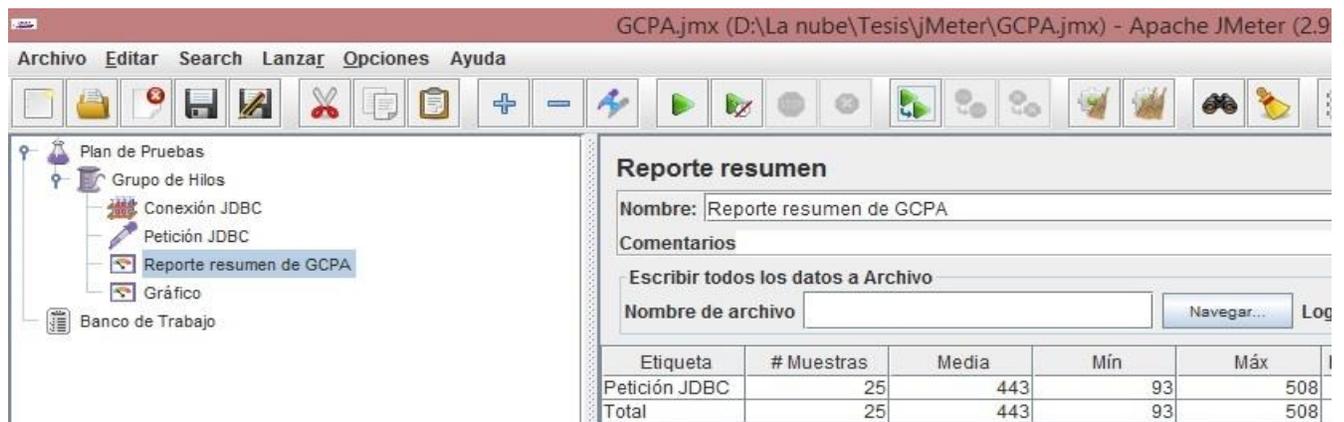


Figura 23: Resultados del plan de pruebas simulando para la Fiscalía General.

En el mercado de datos perteneciente al departamento de Gestión de Cuadros y Personal de Apoyo se realizó una prueba con 12 usuarios para una Fiscalía Provincial y con 25 usuarios para la Fiscalía General, realizando peticiones concurrentes al mercado de datos. Los tiempos promedios de respuesta a las peticiones de los usuarios fueron aproximadamente de 104 y 443 milisegundos respectivamente para cada prueba, siendo aceptables. Además el mayor tiempo de respuesta entre las dos pruebas, a una petición fue de 508 milisegundos, siendo también aceptable, según se definió en la tabla 10. La prueba resultó satisfactoria y para los demás mercados de datos los resultados también fueron exitosos debido a que los tiempos promedios de respuesta están entre 1 y 5 segundos.

Por otra parte la siguiente tabla muestra una comparación entre los tiempos de respuesta de los mercados de datos y de la base de datos del SIGEF. Se puede observar que el tiempo de respuesta de los mercados de datos a una consulta está aproximadamente entre 0,1 y 8 segundos, mientras que el tiempo de respuesta de la base de datos del SIGEF a esa misma consulta está entre 70 y 80 segundos, por lo que se puede asegurar que el desarrollo de los mercados de datos solucionó el problema de la investigación anteriormente planteado.



PRUEBAS DE VALIDACIÓN

Tabla 12: Comparación de los tiempos de respuesta.

Mercados de datos	Base de datos del SIGEF
Aproximadamente de 1-8 segundos	Aproximadamente de 70-80 segundos

Prueba de estrés

En la prueba de estrés se realizó el mismo procedimiento que en la prueba de carga pero con una diferencia, la cantidad de usuarios conectados concurrentemente se aumenta, con el objetivo de probar los mercados de datos en condiciones fuera de lo normal y conocer su comportamiento.

Tabla 13: Representación de la cantidad de usuarios para la prueba de estrés.

Instancia	Cantidad de usuarios	
Fiscalía General de la República	25 (real)	50(prueba)

A continuación se muestra los resultados de la prueba de estrés realizada para el mercado de datos perteneciente al departamento de Gestión de Cuadros y Personal de Apoyo, los resultados de esta prueba para los demás mercados de datos se encuentran en el anexo 24.

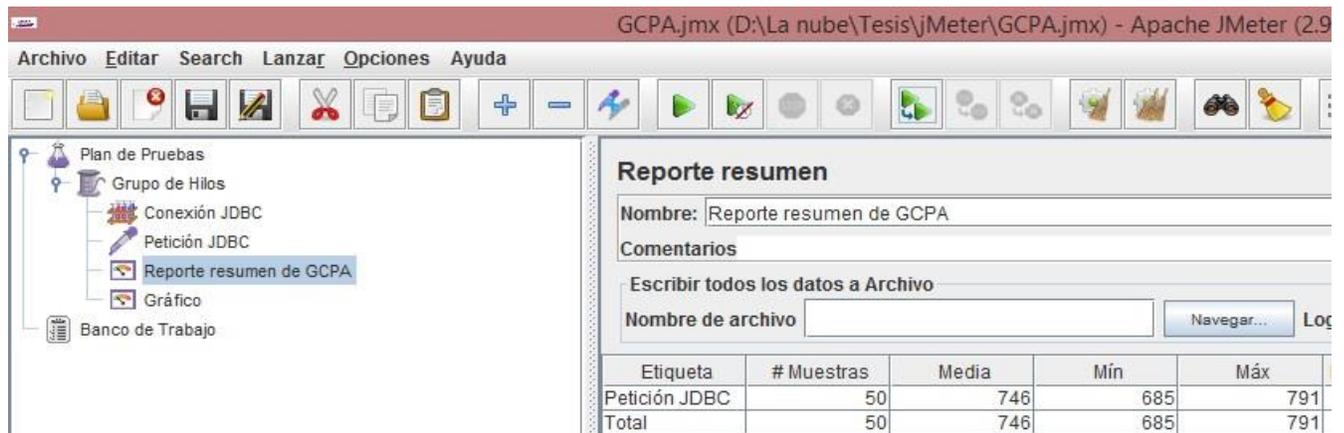


Figura 24: Resultados del plan de prueba simulado.

En el mercado de datos perteneciente al departamento de Gestión de Cuadros y Personal de Apoyo se realizó una prueba con 50 usuarios conectados concurrentemente. El tiempo promedio de respuesta a las peticiones realizadas por los usuarios fue aproximadamente de 746 milisegundos y el mayor tiempo de respuesta a una petición fue de 791 milisegundos, siendo aceptables según se definió en la



PRUEBAS DE VALIDACIÓN

tabla 10. La prueba terminó exitosamente y los resultados obtenidos para los demás mercados de datos también fueron satisfactorios debido a que los tiempos promedios de respuesta están entre los 4 y 6 segundos.

3.2 Conclusiones del capítulo

- Se demostró que los mercados de datos están listos para soportar la carga de trabajo en condiciones reales de la FGR, brindando un tiempo de respuesta aceptable a las peticiones que se le realizan y mejores que los tiempos de respuesta que brinda la base de datos SIGEF.
- Los mercados de datos están aptos para manejar el volumen de datos con el que se trabaja en los distintos niveles de la fiscalía.
- Se demostró que los mercados de datos son capaces de mantener el máximo de conexiones estimadas de manera concurrente sin presentar problemas.



CONCLUSIONES GENERALES

Conclusiones Generales

- Con el estudio de los principales conceptos relacionados con los mercados de datos, se pudo conocer y seleccionar las diferentes metodologías y herramientas existentes para su desarrollo.
- Las necesidades de información identificadas para los departamentos de Procesos Penales y Gestión de Cuadros y Personal de Apoyo de la Fiscalía General de la República de Cuba, permitió modelar los mercados de datos haciendo uso de la metodología propuesta por Kimball, que resultó ser la más adecuada para adaptarse a la solución propuesta.
- El proceso de extracción, transformación y carga resultó ser un elemento fundamental que permitió seleccionar específicamente la información relevante de los departamentos de Procesos Penales y Gestión de Cuadros y Personal de Apoyo de la Fiscalía General de la República de Cuba para ser almacenada en los mercados de datos.
- Se validaron los mercados de datos para los departamentos de Procesos Penales y Gestión de Cuadros y Personal de Apoyo de la Fiscalía General de la República de Cuba haciendo uso de las pruebas de volumen, carga y estrés, garantizando la correcta implementación de los mismos. Además se compararon los tiempos de respuesta entre los mercados de datos y la base de datos del SIGEF, donde se pudo comprobar que los mercados de datos brinda un mejor tiempo de respuesta a las necesidades de información del usuario asegurándose de esta forma que el desarrollo de los mercados de datos para los departamentos de Procesos Penales y Gestión de Cuadros y Personal de Apoyo solucionó el problema de la investigación planteado anteriormente.



RECOMENDACIONES

Recomendaciones

Con el propósito de mejorar la solución antes desarrollada se plantean la siguiente recomendación:

- Una vez concluida la implementación de todos los mercados de datos para el Sistema de Gestión Fiscal integrarlos en un almacén de datos único para todo el sistema.
- Se propone para la implementación de futuros mercados de datos con características similares en otras áreas, utilizar la metodología y las herramientas seleccionadas en la investigación.



Bibliografía

- Aguilera, Sonia. 2010.** Marketing directo. *Marketing directo*. [Online] 09 2010. [Cited:]
<http://www.marketingdirecto.com/actualidad/marketing/aplicacion-del-analisis-olap-en-la-empresa/>.
- Alegsa, Leandro. 2009.** Alegsa. *Alegsa*. [Online] 2009. [Cited: 05 01, 2015.]
<http://www.alegsa.com.ar/Dic/sistema%20transaccional.php>.
- Avella, Angela. 2011.** calameo. [Online] 2011. [Cited: 4 2, 2015.]
<http://es.calameo.com/books/002299301667571c7ab05>.
- Avila, Katty.** CAVSI. [Online] [Cited: 2 6, 2015.] <http://www.cavsi.com/preguntasrespuestas/que-es-procesamiento-analitico-en-liea-olap/>.
- Baedke, Manfred. 2015.** The Apache software. *The Apache software*. [Online] 2015. <http://www.apache.org/>.
- Bernabeu, Ricardo Dario. 2010.** *Hefesto: Metodología para la construcción de un Datawarehouse*. Córdoba, Argentina : s.n., 2010.
- . **2007.** *HEFESTO: Metodología propia para la Construcción de un Data Warehouse*. Córdoba, Argentina : s.n., 2007.
- Blaha, Michel. 2010.** Patterns of Data Modeling. Emerging Directions in Database Systems and Applications. Taylor &. [book auth.] Michel Iaha. *Patterns of Data Modeling. Emerging Directions in Database Systems and Applications*. Taylor &. Atlanta, Giorgio, USA : Sham Navathe, 2010.
- Blanco, Hector Fuentes. 2012.** *Linea_Base_de_Arquitectura_SIGEF II*. 2012.
- Chen, Doris. 2006.** Microsoft Developer Network. *Microsoft Developer Network*. [Online] 2006. [Cited: 05 01, 2015.]
<https://msdn.microsoft.com/es-es/es-s/library/ms175367%28SQL.90%29.aspx>.
- Dario, Ing. Bernabeu R. 2009.** Dataprix. [Online] 5 2009. [Cited: 2 7, 2015.] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse/34-datawarehouse-manager#x1-420003.4.3>.
- Diaz, Alonso Javier Perez. 2012.** AJPDSof. *AJPDSof*. [Online] 2012. [Cited: 2 9, 2015.]
<http://www.ajpdsoft.com/modules.php?name=Encyclopedia&op=content&tid=769>.
- Eduardo. 2010.** Blogroll. [Online] 2010. [Cited: 2 6, 2015.] <https://eduardo39.wordpress.com/unidad-i/>.
- Frahling, Gereon. 2015.** Linguee. *Linguee*. [Online] 2015. <http://www.linguee.es/espanol-ingles/search?source=auto&query=oltp>.
- George, Sansu. 2012.** searchbusinessintelligence. *searchbusinessintelligence*. [Online] 2012. [Cited: 05 1, 2015.]
<http://searchbusinessintelligence.techtarget.in/tip/Inmon-vs-Kimball-Which-approach-is-suitable-for-your-data-warehouse>.
- Hidalgo-Gato, Gisel González. 2010.** redalyc. *redalyc*. [Online] 4 2, 2010.
<http://www.redalyc.org/articulo.oa?id=193915954004>.



BIBLIOGRAFÍA

- Ibarra, María de los Angeles. 2006.** Procesamiento Analítico en Línea. Corrientes, Argentina : s.n., 2006.
- Inmon, Bill.** Dataprix. [Online] [Cited: 2 2, 2015.] <http://www.dataprix.com/qu-es-un-data-warehouse>.
- . **2014.** Intellego. [Online] 2014. [Cited: 2 2, 2015.] <http://www.intellego.com.mx/es/noticias/bill-inmon-el-padre-del-data-warehousing-dwh-en-mexico>.
- Keller Rodríguez Cabanillas, Angela Lucia Mendoza Peña. 2011.** Análisis diseño e implementación de una solución de inteligencia de negocios para el área de compras y ventas de una empresa comercializadora de electrodomésticos. Lima : s.n., 2011.
- Kimball, Ralph. 2004.** *The Data Warehouse Lifecycle Toolkit*. New York : s.n., 2004.
- Lanzillotta, Analía. 2007.** [Online] 2007. [Cited: 2 4, 2015.] <http://www.mastermagazine.info/termino/6841.php>.
- Lekstraat, Leiden. 2011.** Datanamic Tools for database developers. *Datanamic Tools for database developers*. [Online] 2011. [Cited: 2 9, 2015.] <http://es.downloadv.com/get-Datanamic-Data-Generator-for-PostgreSQL-265096.htm>.
- Luis Flores Mez. 2012.** Prezi. [Online] 2012. [Cited: 2 3, 2015.] https://prezi.com/umfw_tr51ecb/datamart/.
- Martinez, Rafael. 2013.** PostgreSQL-es. *PostgreSQL-es*. [Online] 2013. [Cited: 2 9, 2015.] http://www.postgresql.org.es/sobre_postgresql.
- Martínez, Xiomara. 2009.** Gravatar. *Gravatar*. [Online] 2009. [Cited: 2 9, 2015.] <http://gravitar.biz/pentaho/>.
- Méndez, Geidy Acosta. 2013.** Informática en salud 2013. *Informática en salud 2013*. [Online] 2013. <http://www.informatica2013.sld.cu/index.php/informaticasalud/2013/paper/view/195>.
- Mingo, Pedro Sebastian. 2013.** Servicios especializados en testing. [Online] 2 4, 2013. [Cited: 4 12, 2015.] <http://pedrosebastianmingo.com/para-que-sirven-las-pruebas-de-rendimiento-i-introduccion/>.
- Nima, Jonathan David Ramos. 2015.** EUMEDNET. *EUMEDNET*. [Online] 2015. [Cited: 05 01, 2015.] <http://www.eumed.net/libros-gratis/2009c/574/TECNOLOGIA%20OLTP.htm>.
- Pablo Tactuk . 2013.** Oficina Nacional de Estadística. *Oficina Nacional de Estadística*. [Online] 2013. [Cited: 2 9, 2015.] <http://www.one.gob.do>.
- pentaho-bi-suite.** [Online] [Cited: 2 9, 2015.] <http://www.dataprix.com/empresa/productos/pentaho-bi-suite>.
- Pompa, Yisel González. 2014.** *MERCADOS DE DATOS PARA EL ANÁLISIS ESTADÍSTICO DE LA INFORMACIÓN*. 2014.
- Ramos, Salvador. 2011.** Microsoft Business Intelligence:vea el cubo medio lleno. *Microsoft Business Intelligence:vea el cubo medio lleno*. España : SolidQ Press, 2011.
- Reyes, Juan Gerardo Mata. 2013.** [Online] 2013. [Cited: 2 8, 2015.] <http://es.slideshare.net/juangerardomatareyes/data-warehouse-16842696>.



BIBLIOGRAFÍA

Rivadera, Gustavo R. 2010. *La metodología de Kimball para el diseño de almacenes de datos.* 2010.

Rud, Olivia Parr. 2000. *Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management.* *Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management.* s.l. : John Wiley & Sons, 2000, 2000.

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

Sanchez, Antonio. 2013. *Servidores de Aplicaciones. Servidores de Aplicaciones.* [Online] 2013. [Cited: 2 9, 2015.] <http://servidoresdeaplicaciones.com/2013/08/03/aprendiendo-a-usar-apache-jmeter/>.

Sierra, daniel. 2007. *slideshare. slideshare.* [Online] 11 15, 2007. [Cited: 02 08, 2015.] <http://es.slideshare.net/vanquishdarkenigma/visual-paradigm-for-uml>.

Stackowiak, Robert. 2007. *Oracle Data Warehousing and Business Intelligence Solutions.* [Online] 2007. http://books.google.com.co/books?id=Gxy6_drRWRgC&dq=%22Oracle+Data+Warehousing+and+Business+Intelligence+Solutions%22&printsec=frontcover&source=bn&hl=es&ei=W0uJSqmGsqIgtgewwtjnDA&sa=X&oi=book_result&ct=result&resnum=4#v=onepage&q=&f=false.

TechNet. 2013. [Online] 11 2013. [Cited: 2 5, 2015.] <https://technet.microsoft.com/es-es/library/hh916543.aspx>.

Tedeschi, Nicolás. 2014. *Microsoft Developer Network.* [Online] 2014. [Cited: 3 13, 2015.] <https://msdn.microsoft.com/es-es/library/bb972240.aspx>.

Trujillo, Juan Carlos. 2009. *Universitat Jaume I. Universitat Jaume I.* [Online] 2009. www3.uji.es.

Velasco, Roberto Hernando. *Almacenes de datos.* [Online] [Cited: 2 2, 2015.] <http://www2.rhernando.net/modules/tutorials/doc/bd/dw.html>.

Ventura, Alberto. 2014. *guia-ubuntu. guia-ubuntu.* [Online] 2014. [Cited: 2 9, 2015.] http://www.guia-ubuntu.com/index.php?title=PgAdmin_III.

Wikispaces. *Wiki.* [Online] [Cited: 2 6, 2015.] <http://oltp.wikispaces.com/Desventajas+de+OLTP>.

Zapata, Carlos Mario. 2011. *Revista Avances en Sistemas e Informática. Revista Avances en Sistemas e Informática.* [Online] 8 2, 2011. [Cited: 5 12, 2015.] <http://www.redalyc.org/articulo.oa?id=133119867014>.