

Universidad de las Ciencias Informáticas
Facultad Regional “Mártires de Artemisa”



**Mercado de datos que contribuya a la toma de
decisiones en la dirección de cuadros de la
Administración Provincial de Artemisa.**

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

Autor: Yisel Valdes Rodriguez
Tutor: Ing. Mayrín Ramos Maestre
Co-Tutor: MsC. Raúl Edis Riverón Núñez



"En el mismo hombre suelen ir unidos un corazón pequeño y un talento grande. Pero todo hombre tiene el deber de cultivar su inteligencia, por respeto a sí mismo y al mundo. ..." José Martí

DECLARACIÓN DE AUTORÍA

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

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

Autora: Yisel Valdes Rodriguez

Tutor: Ing. Mayrín Ramos Maestre.

Co-Tutor: MsC. Raúl Edis Riverón Núñez

Dedicatoria

Dedico muy especialmente este trabajo a mi madre Yipsian Rodríguez Soto, quién por su constancia, dedicación, ternura y amor; soy, lo que he llegado a ser. Este también es el fruto de su quehacer. Te amo mamita.

A mi hermana para que siga adelante en sus estudios.

Agradecimientos

A mi madre, que siempre me incitó a luchar por el futuro, por todo ese amor sin límites, por su apoyo, dedicación, sacrificio y por ser ejemplo para mí.

A mi hermana que la adoro.

A mis abuelos Armando y Sara por guiarme y cuidarme siempre.

A mi novio Jandy por brindarme su confianza y amor incondicional.

A mis amigas y amigos por acompañarme siempre en los momentos buenos y malos, a Dailyn, Yaniel, Neybis, Marielis y todos los demás.

A mis compañeros de aula, de proyecto y de residencia.

A mi Tutora Mayrin y mi Co-Tutor Riverón por su apoyo en todo momento y ayudarme a salir adelante.

A mis profesores por guiarme en toda mi vida y a Ricardo por ayudarme siempre.

A todos los que de una forma u otra han tenido que ver con la realización de la tesis.

Resumen

Los almacenes de datos han progresado paulatinamente y son repositorios diseñados para facilitar la confección de los informes y la realización de análisis para la toma de decisiones, lo cual significa una ganancia enorme en el rendimiento de los sistemas cuando se ejecuten las consultas, mientras que los mercados de datos son especializados en un área de negocios concreto y es de gran importancia la calidad de la información guardada en ellos.

Este trabajo se basa en la implementación de un mercado de datos para la dirección de cuadros de la Administración Provincial de Artemisa. La aplicación será capaz de centralizar toda la información referente a las cinco especialidades en tablas de hecho con las medidas correspondientes a cada modelo para ser analizadas por dimensión, facilitando la accesibilidad, el análisis de los datos y apoyar a la toma de decisiones.

Se definió como metodología a usar la metodología basada en Kimball, como sistema gestor de base de datos Postgres SQL v9.0, Visual Paradigm v6.4 como herramienta de modelado y la Suite de Pentaho como herramienta de inteligencia de negocio. Durante el desarrollo de la aplicación se realizó un estudio previo de las herramientas y aplicaciones existentes que ayudaron en la realización de las funcionalidades antes mencionadas.

Palabras claves: Almacén de datos, mercado de datos, dimensión, hecho, medidas.

Índice

Introducción.....	3
Capítulo 1: Fundamentación teórica de los almacenes de datos.	10
Introducción	10
1.2 Tendencias actuales.....	10
1.3 Almacén de datos (Data Warehouse).	11
1.3.1 Características de los almacenes de datos.....	12
1.3.2 Estructura de un almacén de datos.....	13
1.3.3 Componentes de un almacén de datos.....	15
1.3.4 Ventajas del almacén de datos.....	15
1.3.5 Inconvenientes del almacén de datos.....	16
1.4 Mercado de datos.....	16
1.4.1 La fase de construcción:.....	17
1.5 Modelos de almacenamiento de datos para los mercados de datos.....	18
1.6 Integración de datos.....	20
1.7 Inteligencia de negocio.....	21
1.8 Técnicas de almacenamiento de datos.....	22
1.9 Metodología para el diseño de almacenes de datos (<i>Data Warehouse</i>).....	23
1.10 Herramientas de modelado.....	25
1.11 Sistema gestor de base datos y herramienta del sistema gestor de base datos. 26	
1.12 Herramientas de desarrollo.....	26
1.13 Herramientas de inteligencia del negocio.....	27
1.14 Conclusiones parciales.....	29
Capítulo 2: Análisis y diseño del mercado de datos de la dirección de cuadros de la Administración Provincial de Artemisa.....	30
Introducción	30
2.1 Estudio preliminar del negocio.....	30
2.2 Necesidades de información.....	32
2.3 Requisitos funcionales.....	40
2.4 Requisitos no funcionales.....	40
2.5 Roles y permisos	40

2.6	Casos de uso del sistema.....	41
2.7	Modelo de datos dimensional.....	45
2.8	Identificación de dimensiones y hechos.....	47
2.9	Desarrollo de la Matriz BUS:.....	49
2.10	Reglas del negocio.....	50
2.11	Conclusiones parciales.....	51
Capítulo 3: Implementación del mercado de datos de la dirección de cuadros de la Administración Provincial de Artemisa.....		52
Introducción.....		52
3.1	Implementación del modelo de datos físicos.....	52
3.2	Implementación de los subsistemas de integración.....	53
3.3	Implementación de los trabajos (Jobs).....	55
3.4	Implementación del subsistema de visualización de datos.....	56
3.4.1	Cubos OLAP.....	58
3.4.2	Reportes.....	59
3.5	Conclusiones parciales.....	60
Capítulo 4: Validación del mercado de datos de la dirección de cuadros de la Administración Provincial de Artemisa.....		61
Introducción.....		61
4.1	Pruebas.....	61
4.2	Estrategias de pruebas para el plan de pruebas.....	63
4.4	Aporte social y económico.....	66
4.5	Conclusiones parciales.....	66
Conclusiones generales.....		67
Recomendaciones.....		68
Referencias Bibliográficas.....		69
Bibliografía.....		71

Introducción.

En la actualidad el mundo evoluciona vertiginosamente en cuanto a tecnologías, observándose cambios en los sistemas de gestión de bases de datos. La humanidad atesora grandes volúmenes de datos existentes en experiencia, conocimiento y sabiduría, que serían útiles en la toma de decisiones, especialmente en las grandes organizaciones y proyectos científicos. La búsqueda de información relevante es imprescindible para la administración empresarial, el control de la producción, el análisis de los mercados, el diseño en ingeniería y la exploración científica, porque pueden ofrecer las respuestas más apropiadas a las necesidades de información.

El nuevo marco globalizado del mundo actual y sus omnipresentes, imprescindibles y poderosas herramientas TIC, está induciendo una profunda revolución en todos los ámbitos sociales. Esto provoca el constante perfeccionamiento de las empresas en cuanto a tecnologías para su uso y comercialización. La principal arma que se ha desarrollado para contrarrestar esta competencia son los sistemas de análisis de información que permiten estudiar los datos históricos y actuales.

Los sistemas de gestión de bases de datos se han modificado en el uso hasta para los ordenadores personales, siendo muchos de ellos económicos o incluso gratuitos. La tendencia al abaratamiento y disminución en tamaño físico de los sistemas contrasta con la cada vez mayor capacidad, potencia y prestaciones de los sistemas. Los almacenes de datos surgen por la necesidad de resolver problemas de análisis de grandes masas de información.

Generalmente los almacenes de datos se subdividen en unidades lógicas más pequeñas dependiendo del subsistema de la entidad del que procedan o para el que sea necesario. Normalmente, un mercado de datos resuelve estudios a nivel de departamento. Un mercado de datos es una versión del almacén de datos. La diferencia principal es que la creación de un mercado de datos es específica para una necesidad de datos seleccionados, enfatizando el fácil acceso a una información relevante.

El análisis de la estadística es fundamental en el control de los datos para la toma de decisiones, especialmente las relacionadas con el gobierno y los órganos administrativos, en el caso del control de los datos relacionados con los cuadros, constituye una tarea estratégica debido a que hay que garantizar el completamiento de los cuadros y la política de cuadros para un funcionamiento óptimo del sistema.

En Cuba la política de cuadros ha estado orientada a las reglas de cómo se debe trabajar con el sistema de cuadros para no violar los principios éticos de la Revolución, por lo cual se hizo necesaria la toma de medidas que garantizaran un mejor conocimiento sobre las características y condiciones de las personas que se promueven a cargos de dirección en las diferentes nomenclaturas.

La dirección de cuadros tiene como misión: auxiliar al presidente del consejo en la aplicación de la política de cuadros aprobada por el Partido Comunista de Cuba y el Gobierno; implementar y controlar su cumplimiento en las diferentes nomenclaturas de cargos, así como ejercer la autoridad funcional hacia el resto de los órganos y especialistas de cuadros en los niveles de la estructura de base.

La dirección donde se gestiona la información referente a los cuadros, es de gran importancia por la necesidad que existe en la provincia de Artemisa, de tomar decisiones que contribuyan al correcto funcionamiento de la política de cuadros. Por ello el centro de desarrollo de La Facultad Regional Mártires de Artemisa en conjunto con la Administración Provincial de Artemisa, decidieron comenzar el producto mercado de datos para dicha dirección lo cual contribuirá al buen desempeño en la toma de decisiones en las diferentes áreas y direcciones.

La información referente a esta dirección, es recogida a través de diferentes modelos en cada municipio del territorio de la provincia de Artemisa, finalmente se almacena la información en aplicaciones informáticas de oficina (Excel, Word) y documentos impresos.

Situación Problemática:

Durante el estudio preliminar de la situación existente en la dirección de cuadros, de la Administración Provincial de Artemisa, se identificaron algunos elementos que

afectan la gestión de la información de esta dirección, los cuales se exponen a continuación:

- Se detectó que el análisis de la información recibida se realiza de forma manual por los especialistas, tarea que resulta muy engorrosa debido a que los datos se encuentran en múltiples ficheros.
- Existe poca accesibilidad de la información, ya que solamente se puede consultar la misma cuando el especialista está presente, esto conlleva a la pérdida y duplicado de información pues se almacena en formato duro e impreso.
- La elaboración de informes costosos en esfuerzo y tiempo pues se consulta gran cúmulo de información de las diferentes especialidades.

Se dificulta así la disponibilidad de información estadística y el análisis de los datos para los altos mandos de la Administración Provincial de Artemisa por las limitaciones para recuperar indicadores.

Debido a esta situación problemática se establece como **problema de la investigación**: ¿Cómo contribuir a la disponibilidad de la información y facilitar el análisis de los datos para mejorar la toma de decisiones en la dirección de cuadros de la Administración Provincial de Artemisa?

Objeto de estudio: Almacén de datos para la toma de decisiones.

Campo de acción: Mercado de datos para la Administración Provincial de Artemisa.

Para dar solución al problema de la investigación, se identifica como **objetivo general**: Desarrollar un mercado de datos que contribuya a la disponibilidad de la información y facilitar el proceso de análisis de los datos para una mejor toma de decisiones en la dirección de cuadros de la Administración Provincial de Artemisa.

En correspondencia con el mismo, se plantean los siguientes **objetivos específicos**:

- Realizar la fundamentación teórica-metodológica de la investigación.
- Realizar el análisis y diseño del mercado de datos de la dirección de cuadros de la Administración Provincial de Artemisa.
- Implementar el mercado de datos de la dirección de cuadros de la Administración Provincial de Artemisa.
- Validar el mercado de datos de la dirección de cuadros de la Administración Provincial de Artemisa a través de casos de prueba.

Teniendo en cuenta el objetivo general trazado se plantea la siguiente **idea a defender**: Un mercado de datos que contribuirá a la disponibilidad de la información y facilitar el proceso de análisis de los datos para una mejor toma de decisiones en la dirección de cuadros de la Administración Provincial de Artemisa.

Para darle cumplimiento al objetivo planteado, se proponen las siguientes **tareas de la investigación**:

- Definición del marco teórico-metodológico de la investigación y de las herramientas a utilizar en el desarrollo de almacenes de datos.
- Definición de los requisitos, casos de uso, hechos, medidas, dimensiones y diseño del modelo de datos, del subsistema de integración y del subsistema de visualización.
- Desarrollo del modelo de datos, el subsistema de integración y el subsistema de visualización.
- Aplicación de los casos de prueba.

Métodos Teóricos.

- **Histórico - Lógico**: Para la realización de este trabajo se analizó la secuencia lógica desde el surgimiento de los almacenes de datos hasta la actualidad. La trascendencia mundial, en Cuba y más específicamente en la

Universidad de las Ciencias Informáticas (UCI).

- **Analítico – Sintético:** Se utilizó para el análisis de documentos, materiales, y temas relacionados con las mejores prácticas en el desarrollo de mercados de datos. Permitió definir los conceptos fundamentales del tema y se usó para especificar la metodología a utilizar y las dimensiones del mercado.
- **Modelación:** Se desarrollaron los modelos correspondientes al ciclo de vida del desarrollo del mercado de datos propuesto, esto propició facilidades a la hora de cumplir con las tareas de análisis y diseño de los procesos que intervinieron en la aplicación, así como para la implementación del sistema.

Método Empírico.

- **Entrevista:** Este método fue utilizado mediante una conversación previamente definida por el investigador y acordada con el entrevistado, con el objetivo de obtener la mayor cantidad de información posible para diagnosticar y detectar los problemas actuales de la dirección.
- **Encuesta:** La encuesta se realizó mediante un cuestionario que enmarca una serie de interrogantes específicas, sin la intervención directa del investigador, donde las respuestas son seleccionadas de acuerdo con su criterio y se determinaron sus posibles variantes, lo que facilitó su procesamiento estadístico.
- **Análisis documental:** Posibilitó la consulta de materiales y normativas para la realización del presente trabajo.

Se cuenta con una **población** compuesta por los 7 trabajadores de la dirección de cuadros de la Administración Provincial de Artemisa, entre las cuales se interactúa con una **muestra** formada por los 5 especialistas de la dirección de cuadros de la Administración Provincial de Artemisa, representando el 71% de la población actual.

Variable independiente: Mercado de datos.

Variables dependientes: Disponibilidad de la información, facilitación de análisis de datos y toma de decisiones.

Actualidad y Necesidad: Los almacenes de datos han venido naciendo por la carencia de resolver problemas de análisis de grandes volúmenes de información y un mercado de datos especifica su necesidad en la dirección de cuadros de la Administración Provincial de Artemisa por la dificultad de la disponibilidad de información estadística para los altos mandos y por las limitaciones para recuperar indicadores, por lo que se hace necesario la construcción de un mercado de datos para mejorar la toma de decisiones.

Posibles resultados:

- Mercado de datos de la dirección de cuadros de la Administración Provincial de para mejorar la toma de decisiones.
- Informe detallado con toda la base teórico-práctico sobre la cual se sustenta la solución propuesta.

Estructura de la tesis.

Introducción: Se estudió el surgimiento y la evolución de los almacenes y mercados de datos y se realizó todo el análisis metodológico de la investigación científica.

Capítulo 1: Fundamentación teórica de los almacenes de datos.

Se abordan los fundamentos teóricos de los temas relacionados con la tecnología de los almacenes de datos respaldando el presente trabajo, abarcando definiciones, criterios, ventajas y desventajas de este tipo de soluciones, así como las herramientas y metodología a emplear.

Capítulo 2: Análisis y diseño del mercado de datos de la dirección de cuadros de la Administración Provincial de Artemisa.

Se definen los requerimientos del sistema, los requisitos funcionales y no funcionales, se delimitan los roles y permisos del sistema, se diseña el diagrama de casos de uso del sistema, el modelo de datos, los subsistemas de integración y visualización y se especifican las reglas del negocio.

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

Se implementa el modelo de datos físico, el subsistema de integración, los trabajos y el subsistema de visualización, se desarrollan los cubos OLAP, los metadatos y los reportes.

Capítulo 4: Validación del mercado de datos de la dirección de cuadros de la Administración Provincial de Artemisa.

Se valida el mercado de datos a través de los casos de prueba basados en los casos de usos y casos de prueba basado en requisitos. Se establece el plan de pruebas y se verifican los resultados y funcionalidades obtenidas del mercado de datos.

Y finalmente las conclusiones generales, la referencia bibliográfica, la bibliografía consultada, los anexos y el glosario de término.

Capítulo 1: Fundamentación teórica de los almacenes de datos.

Introducción

En el capítulo se abordan los conceptos más importantes relacionados con los almacenes de datos, sus principales características, estructura y las ventajas que proporciona su uso en el mundo empresarial, los mercados de datos y la fase de construcción. Se exponen además aspectos fundamentales, como la metodología de desarrollo a utilizar y las herramientas más empleadas en los procesos de integración de datos e inteligencia de negocio.

1.2 Tendencias actuales

Las compañías a nivel mundial han empezado a migrar hacia su actualización, para lograr una posición ventajosa en este sentido.

A partir de esta problemática un gran número de compañías se dieron a la tarea de implementar la tecnología almacén de datos para convertir sus datos en información útil.

- Mercado de datos para la conserjería de vivienda y mercado de datos para la conserjería de asuntos sociales del Gobierno de La Rioja: El objetivo inicial fue construir un repositorio de información unificado para las diferentes modalidades de ayuda a la adquisición de vivienda, que permitiera analizar visualmente la evolución y repartos de las ayudas, pudiendo tomar decisiones correctivas de forma rápida. La estructura del gobierno es variable, ya que corresponde al presidente designar el número y composición del mismo. La estructura actual (2011) es por departamentos.
- Otro ejemplo de entidades de este tipo que actualmente están utilizando esta tecnología son el instituto nacional de estadística de España y el instituto

nacional de estadísticas en Chile, se tiene la información recopilada en almacenes de datos.

Junto con el desarrollo mundial en Cuba se ha perfeccionado la formación tecnológica sobre el tema de los almacenes de datos y se han visto pasos de avance dentro de esta rama en algunos sectores de la sociedad como se manifiesta en los siguientes ejemplos:

- Se está realizando un almacén de datos para la Oficina Nacional de Estadísticas (ONE).
- En el sector de la salud, el Ministerio de Salud Pública (MINSAP). Trabaja con un almacén de datos y con el aporte de esta herramienta se contribuye con uno de los programas priorizados de la Revolución Cubana.
- El almacén de datos para el Centro de Gestión de la Información y Desarrollo de la Energía (CUBAENERGIA) utilizando un almacén de datos, para alcanzar un entendimiento de los problemas del mundo real de negocios.

Se evidenció la existencia de mercados de datos similares aunque después de analizar las características particulares de cada uno de ellos no pueden ser reutilizados por su arquitectura y por no cumplir con las métricas del negocio de la dirección de cuadros.

1.3 Almacén de datos (Data Warehouse).

El mundo actual evoluciona constantemente y se ha creado un desarrollo en vista a los cambios en el ámbito informático, actualmente un almacén de datos (o *Data Warehouse*) es una gran colección de datos que recoge información de múltiples sistemas fuentes, y cuya actividad se centra en la toma de decisiones -es decir, en el análisis de la información- en vez de en su captura. Una vez reunidos los datos de los sistemas fuentes se guardan durante mucho tiempo, lo que permite el acceso a datos históricos; así los almacenes de datos proporcionan al usuario una interfaz consolidada única para los datos, lo que hace más fácil escribir las

consultas para la toma de decisiones, y lo que mejora la gestión de la información y el aumento de operaciones para el análisis de los datos. Por lo que para cualquier empresa resulta de gran interés guardar su información en un almacén de datos para contribuir a la toma de decisiones maximizando el aumento de ganancias y mejorando el rendimiento del trabajo.

Para proceder a la construcción de la herramienta es necesario realizar un análisis de las metodologías existentes que pueden contribuir al desarrollo de la misma. Si se desea lograr una selección óptima es fundamental el estudio profundo de cada una de ellas, poniendo especial atención a aquellas características que la conviertan en una propuesta factible a utilizar respetando y manteniendo siempre las políticas de desarrollo del centro.

Inmon defiende una metodología descendente (top-down) a la hora de diseñar un almacén de datos, ya que de esta forma se considerarán mejor todos los datos corporativos. En esta metodología los mercados de datos se crearán después de haber terminado el Data Warehouse completo de la organización. [1]

Por otra parte Kimball define un almacén de datos como: *"una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis"*. También fue Kimball quien determinó que un almacén de datos no era más que: *"la unión de todos los mercados de datos de una entidad"*. Defiende por tanto una metodología ascendente (bottom-up) a la hora de diseñar un almacén de datos. [2]

El almacén de datos contiene datos usados en operaciones diarias para la ayuda en la toma de decisiones, para el análisis y para su control. No es más que la unión de varios mercados de datos centrados en un área específica para el análisis de la información almacenada por tiempo histórico.

1.3.1 Características de los almacenes de datos.

Los almacenes de datos solucionan así esta problemática, una de las definiciones más aceptadas es la de Bill Inmon que fue uno de los primeros autores en escribir sobre el tema de los almacenes de datos, define un almacén de datos (Data Warehouse) en términos de las características del repositorio de datos:

- **Orientado a temas.**- Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.
- **Variante en el tiempo.**- Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.
- **No volátil.**- La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas.
- **Integrado.**- La base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.

Dada la bibliografía consultada podemos resumir que un almacén de datos es una colección de datos que está formada por variables (hechos) y dimensiones. Las dimensiones son los elementos para ubicar datos que participan en el análisis y variables los valores que se desean analizar en una estructura de bodega de datos.

1.3.2 Estructura de un almacén de datos.

La estructura básica de la arquitectura del almacén de datos.

- *Datos operacionales.* Origen de datos para el componente de almacenamiento físico del almacén de datos.
- *Extracción de datos.* Selección sistemática de datos operacionales usados para formar parte del almacén de datos.
- *Transformación de datos.* Procesos para resumir y realizar cambios en los datos operacionales.
- *Carga de datos.* Inserción de datos en el almacén.
- *Almacén.* Almacenamiento físico de datos de la arquitectura almacén de datos.
- *Herramienta de acceso.* Herramientas que proveen acceso a los datos. [3]

La estructura lógica de un almacén de datos.

Está compuesta por los siguientes niveles:

- *Metadatos*. Describen la estructura de los datos contenidos en el almacén. Están en una dimensión distinta al resto de niveles.
- *Datos detallados actuales*. Obtenidos directamente del procesado de los datos. Forman el nivel más bajo de detalle. Ocupan mucho espacio. Se almacenan en disco, para facilitar el acceso.
- *Datos detallados históricos*. Igual que los anteriores, pero con datos correspondientes al pasado. Se suelen almacenar en un medio externo, ya que su acceso es poco frecuente.
- *Datos ligeramente resumidos*. Primer nivel de agregación de los datos detallados actuales. Corresponden a consultas habituales. Se almacenan en disco.
- *Datos muy resumidos*. Son el nivel más alto de agregación. Corresponden a consultas que se realizan muy a menudo y que se deben obtener muy rápidamente. Suelen estar separados del almacén de datos, formando *mercados de datos*. [3]

La estructura física puede presentar las siguientes configuraciones.

- *Arquitectura centralizada*. Todo el almacén de datos se encuentra en un único servidor.
- *Arquitectura distribuida*. Los datos del almacén se reparten entre varios servidores. Asignando cada servidor a uno o varios temas lógicos.
- *Arquitectura distribuida por niveles*. Refleja la estructura lógica del almacén, asignando los servidores en función del nivel de agregación de los datos que contienen. Un servidor está dedicado para los datos de detalle, otro para los resumidos y otro para los muy resumidos.

Cuando los datos muy resumidos se duplican en varios servidores para agilizar el acceso se habla de *mercado de datos*. [3]

1.3.3 Componentes de un almacén de datos.

Un almacén está compuesto por varios componentes como son:

Sistemas fuentes operacionales: son utilizados en las empresas para gestionar información que es almacenada en diferentes formatos de acuerdo a las necesidades del negocio. Estos sistemas, conservan pocos datos históricos. Las prioridades principales que poseen son el procesamiento del rendimiento y la disponibilidad.

Área de procesamiento: es un área de almacenamiento donde se realizan un conjunto de procesos conocidos como extracción, transformación y carga (ETL).

Área de presentación: En ella, se almacena toda la información que puede ser de utilidad para el proceso de toma de decisiones en la empresa. Generalmente, es referenciada como un conjunto de mercados de datos integrados, donde cada uno representa a un proceso específico del negocio.

Herramientas de acceso a los datos: en este componente, se utiliza la palabra herramienta para referirse a la variedad de habilidades que pueden ser provistas a los usuarios del negocio, para soportar el proceso de toma de decisiones.

1.3.4 Ventajas del almacén de datos.

Hay muchas ventajas por las que es recomendable usar un almacén de datos.

Algunas de ellas son: [3]

- Los almacenes de datos hacen más fácil el acceso a una gran variedad de datos a los usuarios finales
- Mejoran la productividad de los responsables en la toma de decisiones de la organización debido a que:
 - Se obtiene una base de datos clasificada por temas e histórica.
 - Integración de información procedente de múltiples sistemas externos.
- Los almacenes de datos pueden trabajar en conjunto y, por lo tanto, aumentar el valor operacional de las aplicaciones empresariales, en especial la gestión de relaciones con clientes.

1.3.5 Inconvenientes del almacén de datos.

Utilizar almacenes de datos también plantea inconvenientes, algunos de ellos son:

- A lo largo de su vida los almacenes de datos pueden suponer altos costos. El almacén de datos no suele ser estático. Los costos de mantenimiento son elevados.
- A veces, ante una petición de información estos devuelven una información no óptima, que también supone una pérdida para la organización.
- La subestimación del tiempo requerido para extraer, limpiar y cargar los datos en el almacén.
- Problemas con los sistemas de origen de los datos.
- En almacenes de datos de considerable tamaño puede que la homogeneización de los datos disminuya su valor.
- La construcción de un almacén de datos puede requerir de mucho tiempo.
- La integración de las herramientas de almacén de datos, para conseguir un beneficio en la organización, es muy compleja.

1.4 Mercado de datos.

Un mercado de datos es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. [4]

Un mercado de datos es una versión especial del almacén de datos. Como los almacenes de datos, los mercados de datos contienen una visión de datos operacionales que ayudan a decidir sobre estrategias de negocio basadas en el análisis de tendencias y experiencias pasadas. La diferencia principal es que la creación de un mercado de datos es específica para una necesidad de datos seleccionados, enfatizando el fácil acceso a una información relevante.

Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades. [4]

El mercado de datos se destaca por una definición de requerimientos más fácil y rápida. Debe satisfacer los requerimientos del área o de los departamentos, se identifican los aspectos más importantes del negocio en dimensiones y en medidas y debe contestar las preguntas del negocio en los tres niveles de la toma de decisiones:

- Operativo.
- Táctico.
- Estratégico.

En síntesis, se puede decir que los mercados de datos son pequeños almacenes de datos centrados en un tema o un área de negocios específica dentro de una organización.

1.4.1 La fase de construcción:

La construcción de los mercados de datos atraviesa por varias fases; como son:

- 1.- Construcción del mercado de datos: Esta actividad tiene el objetivo de construir el modelo de datos, los metadatos de la herramienta de explotación y la arquitectura del modelo multidimensional en la herramienta de explotación.
- 2.- Construcción de los procesos de cargas: En esta actividad es cuando se deben desarrollar los procesos de carga de datos, las rutinas de limpieza, los flujos de cargas de datos, las interfaz de acceso, los importadores e integradores de datos, los programas de entrada de datos, etc.
- 3.- Construcción de los reportes analíticos: Consiste en construir los reportes, tableros de control, dashboard, scorecard, etc.
- 4.- Construcción de los procesos de prueba: Se debe construir los programas, reportes, informes que permita probar los procesos de cargas y los reportes entregados. [5]

1.5 Modelos de almacenamiento de datos para los mercados de datos.

El modelo de almacenamiento más usado es el modelo dimensional que tiene como propósito lograr que el manejo de los datos sea el adecuado. A continuación se muestra de forma puntualizada este modelo.

El modelo dimensional: Está basado en dimensiones, las cuales representan categorías de información, atributos que representan un único nivel dentro de una dimensión, pueden existir jerarquías de atributos las cuales expresan relaciones entre diferentes atributos y por último tablas de hechos, las cuales contienen datos de interés que presentan un nivel de granularidad.[5]

Esquemas que se encuentran dentro del modelo dimensional.

Esquema estrella: Su estructura base es similar a una tabla central y un conjunto de tablas que la atienden radialmente. Este modelo, resulta ser asimétrico, pues hay una tabla dominante en el centro con varias conexiones a las otras tablas. Las tablas dimensiones tienen sólo la conexión a la tabla hecho y ninguna más. [5]

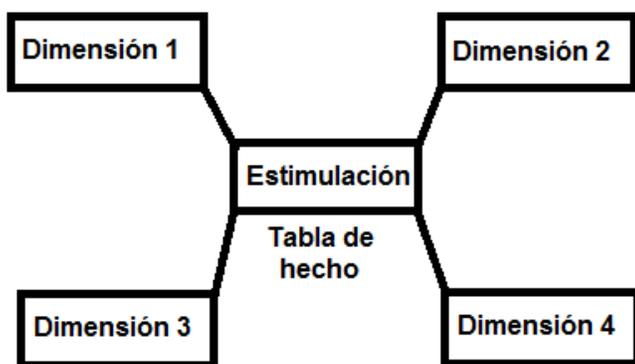


Figura 1: Esquema en estrella

Esquema copo de nieve: Es una estructura más compleja que el esquema de estrella. Se da cuando las dimensiones se implementan con más de una tabla de datos. Aunque puede reducir espacio por la mínima redundancia de datos, tiene la contrapartida de peores rendimientos al tener que crear más tablas de dimensiones y más relaciones entre las tablas lo que tiene un impacto directo sobre el rendimiento. [5]

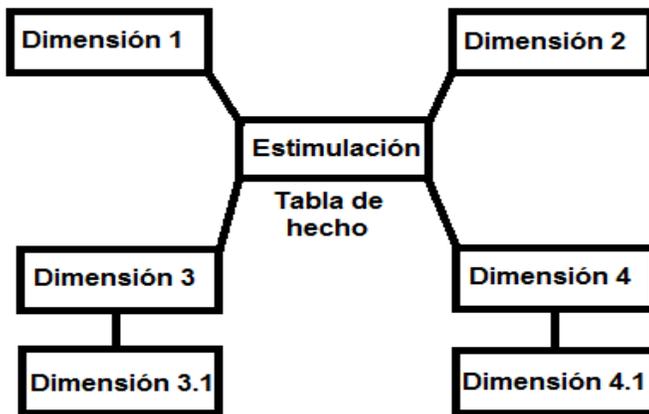


Figura 2: Esquema en copo de nieve.

Tablas de hecho: Habitualmente cada tabla de hecho define un mercado de datos determinado, debido a que en ellas se almacena la información perteneciente al tema en cuestión, ejemplo cuadros, reservas, evaluaciones. La principal condición que deben cumplir las tablas de hechos es que el hecho debe almacenarse de tal forma que su valor sea numérico y a su vez sea aditivo para así poder realizar cálculos sobre él, ya sea por ciento, sumas, entre otros.

Tablas de dimensiones: Las tablas de dimensiones concretan como están los datos introducidos lógicamente y proporcionan el medio para analizar el negocio. Representan los ejes del cubo, y los aspectos de interés, mediante los cuales el usuario podrá filtrar y manipular la información almacenada en la tabla de hechos. Dentro de las tablas podemos encontrar:

- **Jerarquía:** Es un conjunto de atributos descriptivos que se rigen por un orden preestablecido. Una jerarquía implica una organización de niveles dentro de una dimensión, con cada nivel representando el total agregado de los datos del nivel inferior.
- **Granularidad:** Es el nivel más bajo de información que será almacenado en la tabla de hechos e indica el grado de detalle asociado a un hecho particular. La granularidad depende directamente del número de dimensiones que se asocian a la tabla de hechos.

Los datos dentro de la tabla dimensión, que proveen información del negocio o que describen alguna de sus características, son llamados datos de referencia. Entonces, se puede afirmar que una tabla de dimensión posee una clave primaria y uno o más datos de referencia.

Justificación de modelo de almacenamiento de datos a utilizar.

Para la realización del mercado de datos de la dirección de cuadros se utilizó el esquema en estrella pues posee mejor rendimiento y velocidad que los demás modelos de almacenamientos. Consiste en estructurar la información en procesos, vistas y métricas a modo de estrella. En la tabla de hechos encontramos los atributos destinados al hecho que constituye el proceso de negocio a medir, es decir, sus métricas. Mientras, en las tablas de dimensión, los atributos se destinan a elementos de nivel (que representan los distintos niveles de las jerarquías de dimensión) y a atributos de dimensión (encargados de la descripción de estos elementos de nivel).

1.6 Integración de datos.

Diseño del sistema de extracción, transformación y carga (ETL).

La integración de datos, consiste en el proceso de unificación de los datos provenientes de múltiples fuentes. Este sistema se hace necesario para eliminar errores que pueden dañar los reportes generados. El sistema de extracción, transformación y carga (ETL) es la base sobre la cual se alimenta el almacén de datos. Si el sistema ETL se diseña adecuadamente, puede extraer los datos de los sistemas de origen de datos, aplicar diferentes reglas para aumentar la calidad y consistencia de los mismos, consolidar la información proveniente de distintos sistemas, y finalmente cargar la información en el almacén de datos en un formato acorde para la utilización por parte de las herramientas de análisis.

1.7 Inteligencia de negocio.

La inteligencia de negocios (Business Intelligence, BI), surge como una alternativa que permite el manejo efectivo de la información para la toma de mejores decisiones. Uno de los componentes tecnológicos de BI, son los almacenes de datos. Al tener los datos históricos de la organización en un solo almacén se simplifica el problema de acceso a la información, por tanto, acelera el proceso de análisis. Para facilitar el acceso de los usuarios a los datos del almacén, se requiere herramientas de explotación. BI, proporciona algunos componentes para realizar esta tarea: los sistemas OLAP (Procesamiento Analítico en Línea) y la minería de datos. Los sistemas OLAP son los sistemas de procesamiento analítico en línea, permiten realizar un análisis multidimensional de los datos, lo cual ofrece un enfoque más realista de los negocios; la minería de datos se enfoca en la explotación de los datos para detectar patrones o relaciones entre los mismos, esto permite a los analistas entender los factores de éxito del negocio.

Sistemas de inteligencia de negocios: La inteligencia de negocios, BI, es un área de especialización dentro de la tecnología de información (TI). Entre sus objetivos están la generación de información estratégica-gerencial e histórica, así como su despliegue y difusión entre los usuarios. Las aplicaciones de BI son herramientas de soporte de decisiones que permiten en tiempo real, acceso interactivo, análisis y manipulación de información crítica para la empresa [6].

Actualmente el acceso a la información clave del negocio en cualquier lugar y en cualquier momento, se está convirtiendo en la mayor prioridad para muchos negocios. Es por ello que los proveedores de inteligencia de negocio, cuyos productos analizan una gran cantidad de datos de ventas, clientes, personal, entre muchos otros, están añadiendo la capacidad de transmisión inalámbrica a su software.

Los sistemas OLAP, son sistemas de soporte a la decisión, facilitan el análisis de los datos corporativos almacenados en los almacenes y mercados de datos mediante herramientas de acceso, consulta y visualización. Los sistemas de minería de datos, mediante diferentes técnicas permite explotar los datos de

grandes bases de datos (como almacenes o mercados de datos) para determinar relaciones y patrones entre esos datos, información que sirve para la toma de decisiones.

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.

Kimball divide a estas aplicaciones en dos categorías basadas en el nivel de sofisticación, y les llama informes estándar y aplicaciones analíticas. [6]

1.8 Técnicas de almacenamiento de datos.

Dentro de los sistemas gestores de base de datos podemos mencionar las bases de datos OLTP y las bases de datos OLAP.

De acuerdo con los sistemas OLTP (Procesamiento de transacciones en línea) son bases de datos orientadas al procesamiento de transacciones. El proceso transaccional es típico de las bases de datos operacionales.

De acuerdo al soporte del SGBD (Modelado Entidad-Relación o Multidimensional) y del esquema de diseño, existen tres categorías de servidores OLAP: [7]

- ROLAP (Relacional OLAP): Almacenamiento en un esquema en estrella (no normalizado) o copo de nieve (normalizado).
- MOLAP (Multidimensional OLAP): Los datos se almacenan en un sistema de matrices (Cubo) en donde cada eje es una dimensión.
- HOLAP (Híbrido OLAP): Proporciona análisis multidimensional accediendo indistintamente a bases de datos multidimensionales o relacionales (es una unión entre ROLAP y MOLAP).

Justificación de la técnica de almacenamiento de datos a utilizar.

Utilizamos el sistema OLAP en la categoría de ROLAP ya que maneja una arquitectura de tres niveles. La base de datos relacional maneja los requerimientos

de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica. El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y obtención del dato. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.

Después de que el modelo de datos para el almacén de datos se ha definido, los datos se cargan desde el sistema operacional. Se ejecutan rutinas de bases de datos para agregar el dato, si así es requerido por el modelo de dato. Se crean entonces los índices para optimizar los tiempos de acceso a las consultas.

Los usuarios finales ejecutan sus análisis multidimensionales, a través del motor ROLAP, que transforma dinámicamente sus consultas a consultas SQL. Se ejecutan estas consultas SQL en las bases de datos relacionales, y sus resultados se relacionan mediante tablas cruzadas y conjuntos multidimensionales para devolver los resultados a los usuarios.

1.9 Metodología para el diseño de almacenes de datos (*Data Warehouse*)

Con la intención de perfeccionar el diseño de almacenes de datos se crearon diferentes metodologías de desarrollo, las cuales definen un conjunto de pasos y técnicas a seguir.

El diseño de un almacén de datos, es una disciplina que ha ido madurando con el paso del tiempo durante el cual, han resaltado dos enfoques principales concernientes a dos de las personalidades más conocidos en el área, Ralph Kimball y William H. Inmon. La principal diferencia entre ambos estilos radica en la forma de enfrentar el problema.

La metodología de Kimball proporciona una base empírica y metodológica adecuada para las implementaciones de almacenes de datos pequeños y medianos, dada su gran versatilidad y su enfoque ascendente, que permite construir los almacenes en forma escalonada. [1]

Debido a sus características, esta metodología es reconocida como bottom-up por su arquitectura ascendente. Además presenta una serie de herramientas, tales

como gráficos y documentos, que proporcionan una gran ayuda para iniciarse en el ámbito de la construcción de un almacén de datos.

La metodología se basa en lo que Kimball denomina ciclo de vida dimensional del negocio para los almacenes de datos o Data Warehouse.

Estas rutas se combinan cuando se instala finalmente el sistema. En la parte de abajo de la figura se muestra la actividad general de administración del proyecto.

Sin embargo, el enfoque propuesto por Inmon presenta una arquitectura descendente por lo que se conoce como top-down. Este método, propone la creación de un repositorio de datos como fuente de información consolidada, persistente, histórica y de calidad. De esta forma, se crea primero un almacén de datos del cual se van a sostener los diferentes mercados de datos que van a satisfacer las carencias específicas de cada proceso del negocio.

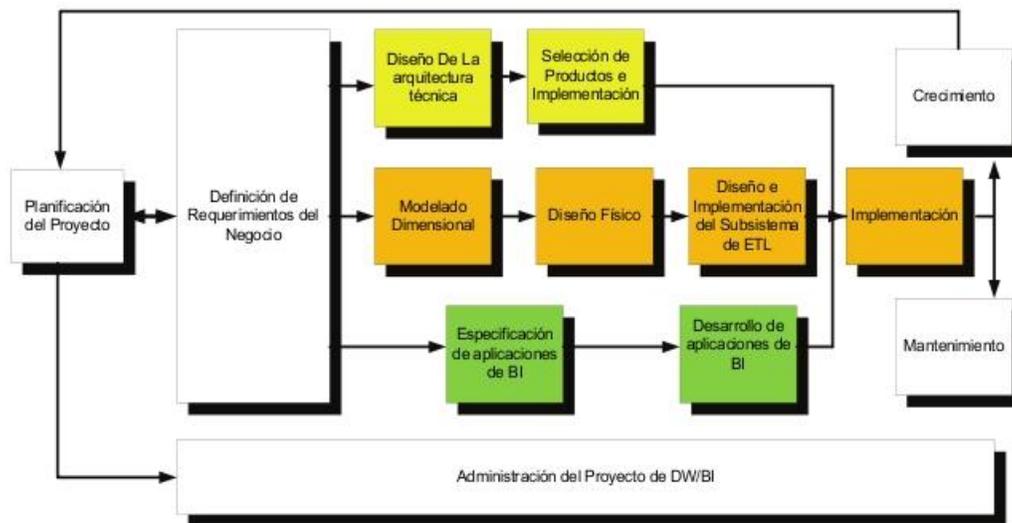


Figura 3: Ciclo de vida, Metodología Kimball [1]

Justificación de la metodología a utilizar.

La Metodología propuesta para desarrollar un mercado de datos es la Metodología de Proceso de Desarrollo en la Línea Soluciones de Almacenes de Datos e Inteligencia de Negocio. Fue elaborada a partir del estudio y análisis de la metodología de Kimball y teniendo en cuenta la experiencia de los recursos humanos en otros proyectos desarrollados. Es una metodología mixta que reúne

elementos de varias metodologías de desarrollo de proyectos de integración de datos, ya que cubre todas las fases por las que pasa su construcción, desde el levantamiento de información inicial, la arquitectura, el diseño, la implementación de la herramienta de inteligencia de negocio, prueba, despliegue, soporte, hasta la gestión del proyecto.

1.10 Herramientas de modelado.

Case

Las herramientas CASE (Computer Aided Software Engineering), son un conjunto de métodos, utilidades y técnicas, que facilitan la automatización del ciclo de vida del desarrollo de un software.

Visual Paradigm.

Visual Paradigm ofrece un entorno de creación de diagramas para UML, diseño centrado en caso de uso y enfocado al negocio que generan un software de mayor calidad, uso de un lenguaje común a todo el equipo de desarrollo que facilita la comunicación, capacidades de ingeniería directa (versión profesional) e inversa, modelo y código que permanece sincronizado en todo el ciclo de desarrollo, disponibilidad de múltiples versiones para cada necesidad, disponibilidad de integrarse en los principales IDE, disponibilidad en múltiples plataformas. [8]

Se cuenta con una versión libre y gratuita denominada edición de la comunidad.

Justificación de herramientas de modelado a utilizar.

Utilizamos la herramienta Visual Paradigm. Esta permite la creación de diagramas para el modelamiento del negocio en la ingeniería de software, permitiendo resaltar los detalles más importantes. Posibilita seguir un diseño común entre todos los programadores de un proyecto específico lo cual brinda mayor legibilidad y organización del trabajo en equipo. Es fácil de instalar y actualizar y soporta aplicaciones web.

1.11 Sistema gestor de base datos y herramienta del sistema gestor de base datos.

PostgreSQL v9.0

Es un potente motor de bases de datos, que tiene prestaciones y funcionalidades equivalentes a muchos gestores de bases de datos comerciales. Está considerado como el gestor de bases de datos de código abierto más avanzado del mundo, ya que proporciona un gran número de características que normalmente sólo se encontraban en gestores comerciales como Oracle. [9]

PgAdmin III

PgAdmin III es una aplicación gráfica para gestionar el gestor de bases de datos PostgreSQL, siendo la más completa y popular con licencia Open Source. Está escrita en C++ usando la librería gráfica multiplataforma wxWidgets, lo que permite que se pueda usar en Linux, FreeBSD, Solaris, Mac OS X y Windows. Es capaz de gestionar versiones a partir de la PostgreSQL 7.3 ejecutándose en cualquier plataforma, así como versiones comerciales de PostgreSQL como Pervasive Postgres, EnterpriseDB, Mammoth Replicator y SRA PowerGres. [11]

Esta aplicación está diseñada para satisfacer las necesidades de todos los usuarios. El interfaz gráfico soporta todas las características de PostgreSQL y facilita enormemente la administración, ya que es de código abierto. La conexión al servidor puede hacerse mediante conexión TCP/IP o Unix Domain Sockets (en plataformas *nix), y puede encriptarse mediante SSL para mayor seguridad. La aplicación lleva consigo un generador de consultas y está disponible en más de 30 idiomas.

1.12 Herramientas de desarrollo.

Pentaho Data Integration

Pentaho Data Integration (PDI) es un motor de transformación, y desde el principio se observa que ha sido diseñado para cubrir las necesidades en la integración de datos. La plataforma Open Source Pentaho Business Intelligence cubre amplias necesidades de análisis de los datos y de los informes empresariales. Abre, limpia

e integra la 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 TIC hoy en día.

Las soluciones de Pentaho están escritas en Java y tienen un ambiente de implementación también basado en este lenguaje. Eso hace que sea una solución muy flexible para cubrir una amplia gama de necesidades empresariales. Es una herramienta concreta para el desarrollo ágil de soluciones de inteligencia del negocio.

En el mundo del desarrollo de software las metodologías ágiles cada día van adquiriendo mayor aceptación y adopción dada la versatilidad y flexibilidad que ofrecen, características que en la Inteligencia de Negocios (BI) también se aprecian bastante.

Las características básicas de esta herramienta son:

- Entorno gráfico de desarrollo.
- Uso de tecnologías estándar: Java, XML, JavaScript.
- 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 cuatro herramientas:

Spoon: Para diseñar transformaciones ETL usando el entorno gráfico.

PAN: Para ejecutar transformaciones diseñadas con spoon.

CHEF: Para crear trabajos.

Kitchen: Para ejecutar trabajos. [12]

1.13 Herramientas de inteligencia del negocio.

Pentaho Metadata Editor v4.1:

Es una herramienta que permite crear dominios de metadatos y modelos. El objetivo es mapear la estructura física de la base de datos a un modelo lógico de negocio.

Pentaho Schema Workbench v3.2:

Es una herramienta gráfica que permite la construcción de esquemas de Mondrian. Permite la conexión a la BD de forma fácil y segura y accede a la publicación de esquemas en el servidor BI para ser utilizados en los análisis por los usuarios. Permite crear esquemas donde se agregan los cubos OLAP que se componen de archivos XML donde se definen las dimensiones con su jerarquía.

Funcionalidades:

- Editor de esquemas integrados con el origen de datos subyacente para su validación.
- Prueba de MDX consultas en el esquema y la base de datos.
- Navegar por la estructura subyacente de bases de datos. [13]

Apache Tomcat 5.5

Fue creado bajo el proyecto Jakarta de la fundación Apache, siendo mantenido por una comunidad de desarrollo, logrando destacar como un producto robusto y altamente eficiente. Esta herramienta es gratis, fácil de instalar y se puede ejecutar en máquinas con pocos recursos, así como también es compatible con las API más recientes de Java. Debido a que su código binario posee un tamaño total de un poco más de un megabyte, no ocupa mucho espacio de modo que no resulta extraño que se ejecute tan rápidamente. Además, su remarcada estabilidad y la capacidad de ejecución multiplataforma, lo han colocado como uno de los servidores más utilizados para el despliegue de aplicaciones web basadas en la tecnología Java. [14]

Es un servidor HTTP o se utiliza conectado a otro servidor HTTP y un contenedor de *servlets*. En la instalación para poder acceder a las aplicaciones de gestión y administración es necesario crear un usuario de Tomcat con los roles adecuados. En el caso de NetBeans, incluye un Tomcat interno. No hace falta ni siquiera instalarlo.

1.14 Conclusiones parciales

En este capítulo se estudiaron temas como antecedentes, definiciones, características y tecnologías, además se realizó un análisis de la evolución y definición de los almacenes y mercados de datos, como son: ventajas, inconvenientes y características, dejando de manera clara que un mercado de datos es la solución para este negocio. Se definió que para darle solución al problema en cuestión, se adoptó el estudio de la metodología de solución basada en Kimball, se analizaron algunos de los modos de almacenamiento (ROLAP, MOLAP, HOLAP), justificando la selección de ROLAP, y dentro de las herramientas a utilizar se encuentran Visual Paradigm v6.4, Power Architect v3.0, PostgreSQL v9.0, suite de la comunidad de Pentaho v3.6.

Capítulo 2: Análisis y diseño del mercado de datos de la dirección de cuadros de la Administración Provincial de Artemisa.

Introducción

En el capítulo se realiza un análisis del negocio, y se propone un diseño de la solución con las características necesarias para satisfacer las necesidades presentadas por el cliente. Se describen los requisitos informativos, funcionales y no funcionales, así como los casos de uso del sistema y la trazabilidad que existe entre ellos.

2.1 Estudio preliminar del negocio.

La dirección de cuadros de la Administración Provincial de Artemisa tiene la misión de auxiliar al presidente del consejo en la aplicación de la política de cuadros. Está compuesta por 5 especialidades.

- Estimulación, estadística, evaluación, reserva, superación.

Para las áreas de análisis, la solución más factible es implementar un área de análisis por cada una de las especialidades y por el balance general, quedando de la siguiente forma:

Áreas de análisis:

- Especialidad de evaluación.
- Especialidad de estimulación.
- Especialidad de superación.
- Especialidad de reserva.
- Especialidad de estadística.
- Balance general de la dirección de cuadros.

Temas de análisis:

- Análisis de los indicadores de evaluación de la dirección de cuadros.

- Análisis de los indicadores de estimulación de la dirección de cuadros.
- Análisis de los indicadores de chequeo médico de la dirección de cuadros.
- Análisis de los indicadores de superación de la dirección de cuadros.
- Análisis de los indicadores de reserva de la dirección de cuadros.
- Análisis de los indicadores de movimientos de cuadros en la dirección de cuadros.
- Análisis de los indicadores de estadísticas de la dirección de cuadros.
- Análisis del balance general de la dirección de cuadros.

En la especialidad de evaluación, se propone la organización, planificación, realización y control del proceso de evaluación de los cuadros y una vez aprobado, controla su cumplimiento; elabora y controla los balances, así como el proceso de ejecución de las medidas disciplinarias aplicadas a los cuadros de la Administración Municipal y la Administración Provincial.

En la especialidad de estimulación, se controla el proceso de atención y estimulación de los cuadros, registro de vacaciones y el chequeo médico de los cuadros.

En la especialidad de estadística, se implementa y controla los análisis integrales y valoraciones sobre el estado del completamiento cuantitativo y cualitativo de los cuadros.

En la especialidad de superación, propone de conjunto con los restantes órganos y especialistas, el Plan de Preparación y Superación de los Cuadros y sus Reservas, en correspondencia con lo establecido en la estrategia nacional y orientar y controlar su realización en las entidades subordinadas.

En la especialidad de reserva, se controla el proceso de proponer y tramitar los movimientos y otras decisiones sobre los cuadros y sus reservas, así como controlar los datos de las reservas de los cuadros.

2.2 Necesidades de información.

En la presente investigación se propone dar solución a problemas existentes en la dirección de cuadros de la Administración Provincial de Artemisa, relacionados con la administración de la información referente a las cinco especialidades. Por tal motivo, es de vital importancia identificar las necesidades de información que tienen los especialistas de dichas áreas, pues constituyen la base para un correcto diseño del mercado de datos.

RI1. Obtener total de cuadros, nivel de completamiento cubiertos, % de cubiertos, nivel de completamiento vacantes, % de vacantes, sexo femenino, % femeninos, sexo masculino, % masculino, total de mujeres que laboran en la entidad, % mujeres del total que laboran, edad entre 25-40 años, % entre 25-40 años, edad entre 41-50 años, % entre 41-50 años, edad entre 51-60 años,% entre 51-60 años, edad con más de 61 años,% más de 61 años, raza blanca,% de piel blanca, raza negra,% de piel negra, raza mestiza, % de piel mestiza, raza negros y mestizos,% negros y mestizos, nivel escolar 12grado,% con 12grado, nivel escolar técnico medio,% con técnico medio, nivel escolar nivel superior,% nivel superior, militancia UJC,% de UJC, militancia PCC,% que pertenecen al PCC, militancia doble militancia,% que pertenecen doble militancia, mujeres con cargos decisorios,% de mujeres con cargos decisorios por municipio y mes.

RI2. Obtener total de cuadros, nivel de completamiento cubiertos, % de cubiertos, nivel de completamiento vacantes, % de vacantes, sexo femenino, % femeninos, sexo masculino, % masculino, total de mujeres que laboran en la entidad, % mujeres del total que laboran, edad entre 25-40 años, % entre 25-40 años, edad entre 41-50 años, % entre 41-50 años, edad entre 51-60 años,% entre 51-60 años, edad con más de 61 años,% más de 61 años, raza blanca,% de piel blanca, raza negra,% de piel negra, raza mestiza, % de piel mestiza, raza negros y mestizos,% negros y mestizos, nivel escolar 12grado,% con 12grado, nivel escolar técnico medio,% con técnico medio, nivel escolar nivel superior,% nivel superior, militancia UJC,% de UJC, militancia PCC,% que pertenecen al PCC, militancia doble

militancia,% que pertenecen doble militancia, mujeres con cargos decisorios,% de mujeres con cargos decisorios por dirección y mes.

RI3. Obtener total de cuadros, total nivel de completamiento cubiertos, total nivel de completamiento vacantes, total sexo femenino, total sexo masculino, total de mujeres que laboran en la entidad, total edad entre 25-40 años, total edad entre 41-50 años, total edad entre 51-60 años, total edad con más de 61 años, total raza blanca, total raza negra, total raza mestiza, total raza negros y mestizos, total nivel escolar 12grado, total nivel escolar técnico medio, total nivel escolar nivel superior, total militancia UJC, total militancia PCC, total militancia doble militancia, total mujeres con cargos decisorios por mes.

RI4. Obtener del chequeo físico el total de cuadros a chequear, de chequeo físicos los chequeados, % de chequeados, de chequeo físicos los pendientes, % de pendientes, chequeo estomatológico el total de cuadros a chequear, chequeo estomatológico los chequeados, % de chequeados estomatológico, de chequeo estomatológico los pendientes, % de pendientes estomatológicos por dirección y mes.

RI5. Obtener total de cuadros, estimulados, % de estimulados, no estimulados, % de no estimulados, total de fallos, fallos cubiertos con otros cuadros por municipio y mes.

RI6. Obtener total de cuadros, estimulados, % de estimulados, no estimulados, % de no estimulados, total de fallos, fallos cubiertos con otros cuadros por dirección y mes.

RI7. Obtener total de cuadros, total de cuadros a evaluar, % de total de cuadros a evaluar, evaluados, % de evaluados, no evaluados en el tiempo establecido, % de no evaluados en el tiempo establecido, cantidad de evaluados que coinciden con la evaluación del cumplimiento de objetivos, fecha de aprobación en la comisión, resultados avanza, % de resultados avanza, resultados estables, % de resultados estables, resultados estancados, % resultados estancados, resultados en retroceso, % en retroceso, a superarse, % a superarse, a liberar, % a liberar, total de recomendaciones, de los cuadros de reserva el total, de los cuadros de

reserva se ratifican, % que se ratifican del total, de los cuadros de reserva los listos, % de listos, estabilidad a mantenerse, % de estabilidad a mantenerse, estabilidad a mover, % de estabilidad a mover, sobre el resultado en desacuerdo, % en desacuerdo, sobre el resultado de reclamaciones el total, % del total de reclamaciones, sobre el resultado de reclamaciones con razón, % con razón, sobre el resultado de reclamaciones sin razón, % sin razón por municipio y mes.

RI8. Obtener total de cuadros, total de cuadros a evaluar, % de total de cuadros a evaluar, evaluados, % de evaluados, no evaluados en el tiempo establecido, % de no evaluados en el tiempo establecido, cantidad de evaluados que coinciden con la evaluación del cumplimiento de objetivos, fecha de aprobación en la comisión, resultados avanza, % de resultados avanza, resultados estables, % de resultados estables, resultados estancados, % resultados estancados, resultados en retroceso, % en retroceso, a superarse, % a superarse, a liberar, % a liberar, total de recomendaciones, de los cuadros de reserva el total, de los cuadros de reserva se ratifican, % que se ratifican del total, de los cuadros de reserva los listos, % de listos, estabilidad a mantenerse, % de estabilidad a mantenerse, estabilidad a mover, % de estabilidad a mover, sobre el resultado en desacuerdo, % en desacuerdo, sobre el resultado de reclamaciones el total, % del total de reclamaciones, sobre el resultado de reclamaciones con razón, % con razón, sobre el resultado de reclamaciones sin razón, % sin razón por dirección y mes.

RI9. Obtener total de cuadros, total de cuadros a evaluar, total evaluados, total no evaluados en el tiempo establecido, total cantidad de evaluados que coinciden con la evaluación del cumplimiento de objetivos, fecha de aprobación en la comisión, total resultados avanza, total resultados estables, total resultados estancados, total resultados en retroceso, total a superarse, total a liberar, total de recomendaciones, total de los cuadros de reserva, total de los cuadros de reserva se ratifican, total de los cuadros de reserva los listos, total estabilidad a mantenerse, total estabilidad a mover, total sobre el resultado en desacuerdo, total sobre el resultado de reclamaciones el total, total sobre el resultado de

reclamaciones con razón, total sobre el resultado de reclamaciones sin razón por mes.

RI10. Obtener total de cuadros, categorías de cuadro directivo superior, % directivo superior, categorías de cuadro directivo, % directivos, categorías de cuadro ejecutivo, % ejecutivos, de medidas aplicadas amonestaciones privadas, % de amonestaciones privadas, de medidas aplicadas amonestaciones públicas, % de amonestaciones públicas, de medidas aplicadas democión temporal, % de democión, de medidas aplicadas democión definitiva, % democión definitiva, de medidas aplicadas separación definitiva, % de separación definitiva, causas resultados deficientes, % de resultados deficientes, causas incumplimiento de funciones, % de incumplimientos, causas descontrol administrativo, % de descontrol, apelación, % apelación, ratificación de la medida muy débil, % de débil, ratificación de la medida muy severa,% de severa, nueva medida por municipio y mes.

RI11. Obtener total de cuadros, categorías de cuadro directivo superior, % directivo superior, categorías de cuadro directivo, % directivos, categorías de cuadro ejecutivo, % ejecutivos, de medidas aplicadas amonestaciones privadas, % de amonestaciones privadas, de medidas aplicadas amonestaciones públicas, % de amonestaciones públicas, de medidas aplicadas democión temporal, % de democión, de medidas aplicadas democión definitiva, % de democión definitiva, de medidas aplicadas separación definitiva, % de separación definitiva, causas resultados deficientes, % de resultados deficientes, causas incumplimiento de funciones, % de incumplimientos, causas descontrol administrativo, % de descontrol, apelación, % apelación, ratificación de la medida muy débil, % de débil, ratificación de la medida muy severa,% de severa, nueva medida por dirección y mes.

RI12. Obtener total de cuadros, total categorías de cuadro directivo superior, total categorías de cuadro directivo, total categorías de cuadro ejecutivo, total de medidas aplicadas amonestaciones privadas, total de medidas aplicadas amonestaciones públicas, total de medidas aplicadas democión temporal, total de

medidas aplicadas democión definitiva, total de medidas aplicadas separación definitiva, total causas resultados deficientes, total causas incumplimiento de funciones, total causas descontrol administrativo, total apelación, total ratificación de la medida muy débil, total ratificación de la medida muy severa por mes.

RI13. Obtener cuadros que asisten a las preparaciones la fecha, total general de cuadros a asistir, total general de cuadros que asistieron, % que asistieron, cantidad que no asistieron, % que no asistieron, defensa formación, % formación, defensa actualización, % actualización en la defensa, defensa coden, % coden en la defensa, defensa total, % total defensa, tipos de curso técnica profesional postgrado, % postgrado en técnica profesional, tipos de curso técnica profesional diplomado, % diplomado en técnica profesional, tipos de curso técnica profesional maestría, % maestría en técnica profesional, tipos de curso técnica profesional doctorado, % doctorado en técnica profesional, total de técnica profesional, % total de técnica profesional, tipos de curso otros economía, % economía a otros cursos, tipos de curso otros curso de dirección, % dirección a otros cursos, total que asistieron a otros cursos, % de otros cursos por municipio y mes.

RI14. Obtener cuadros que asisten a las preparaciones la fecha, total general de cuadros a asistir, total general de cuadros que asistieron, % que asistieron, cantidad que no asistieron, % que no asistieron, defensa formación, % formación, defensa actualización, % actualización en la defensa, defensa coden, % coden en la defensa, defensa total, % total defensa, tipos de curso técnica profesional postgrado, % postgrado en técnica profesional, tipos de curso técnica profesional diplomado, % diplomado en técnica profesional, tipos de curso técnica profesional maestría, % maestría en técnica profesional, tipos de curso técnica profesional doctorado, % doctorado en técnica profesional, total de técnica profesional, % total de técnica profesional, tipos de curso otros economía, % economía a otros cursos, tipos de curso otros curso de dirección, % dirección a otros cursos, total que asistieron a otros cursos, % de otros cursos por dirección y mes.

RI15. Obtener cuadros que asisten a las preparaciones la fecha, total general de cuadros a asistir, total general de cuadros que asistieron, total que no asistieron,

total defensa formación, total defensa actualización, total defensa coden, total defensa, total tipos de curso técnica profesional postgrado, total tipos de curso técnica profesional diplomado, total tipos de curso técnica profesional maestría, total tipos de curso técnica profesional doctorado, total de técnica profesional, total tipos de curso otros economía, total tipos de curso otros curso de dirección, total que asistieron a otros cursos por mes.

RI16. Obtener total de reservas, sexo femenino, % femenino, sexo masculino, % hombres, raza blanca, % blancos, raza negra,% negros, raza mestiza, % mestizos, nivel escolaridad técnico medio(M/S),% técnicos medios, nivel escolaridad nivel superior,% nivel superior, militancia UJC,% UJC, militancia PCC,% PCC, reservas que son cuadro,% reservas cuadros, año de ingreso a la reserva, año de baja de la reserva, año de promoción de la reserva por municipio y mes.

RI17. Obtener total de reservas, sexo femenino, % femenino, sexo masculino, % hombres, raza blanca, % blancos, raza negra,% negros, raza mestiza, % mestizos, nivel escolaridad técnico medio(M/S),% técnicos medios, nivel escolaridad nivel superior,% nivel superior, militancia UJC,% UJC, militancia PCC,% PCC, reservas que son cuadro,% reservas cuadros, año de ingreso a la reserva, año de baja de la reserva, año de promoción de la reserva por dirección y mes.

RI18. Obtener total de reservas, total sexo femenino, total sexo masculino, total raza blanca, total raza negra, total raza mestiza, total nivel escolaridad técnico medio (M/S), total nivel escolaridad nivel superior, total militancia UJC, total militancia PCC, total reservas que son cuadro, total año de ingreso a la reserva, año de baja de la reserva, año de promoción de la reserva por mes.

RI19. Obtener total de propuesta del mes, tipos de movimientos propuestos promoción,% promoción, tipos de movimientos propuestos igual nivel,% igual nivel, tipos de movimientos propuestos democión, % democión, tipos de movimientos propuestos pasar cursos,% pasar cursos, tipos de movimientos propuestos otros, % otros, efecto del movimiento en la entidad traslado interno,% traslado interno, efecto del movimiento en la entidad altas,% altas, efecto del movimiento en la

entidad bajas,% bajas, total de movimiento que proviene de la reservas del cargo,% total de movimientos, total de aprobados, total de denegado por dirección y mes.

RI20. Obtener total de propuesta del mes, tipos de movimientos propuestos promoción,% promoción, tipos de movimientos propuestos igual nivel,% igual nivel, tipos de movimientos propuestos democión, % democión, tipos de movimientos propuestos pasar cursos,% pasar cursos, tipos de movimientos propuestos otros, % otros, efecto del movimiento en la entidad traslado interno,% traslado interno, efecto del movimiento en la entidad altas,% altas, efecto del movimiento en la entidad bajas,% bajas, total de movimiento que proviene de la reservas del cargo,% total de movimientos, total de aprobados, total de denegado por año.

RI21. Obtener total de cuadros, nivel de completamiento cubiertos, % de cubiertos, nivel de completamiento vacantes, % de vacantes, sexo femenino, % femeninos, sexo masculino, % masculino, edad entre 25-40 años, % entre 25-40 años, edad entre 41-50 años, % entre 41-50 años, edad entre 51-60 años,% entre 51-60 años, edad con más de 61 años,% más de 61 años, raza blanca,% de piel blanca, raza negra,% de piel negra, raza mestiza, % de piel mestiza, raza negros y mestizos,% negros y mestizos, nivel escolar 12grado,% con 12grado, nivel escolar técnico medio,% con técnico medio, nivel escolar nivel superior,% nivel superior, militancia UJC,% de UJC, militancia PCC,% que pertenecen al PCC, % de mujeres con cargos decisorios, total de reservas, reservas que son cuadro, % reservas cuadros, total general de cuadros que asistieron, % que asistieron, defensa total, % total defensa, total de técnica profesional, % total de técnica profesional, tipos de curso otros curso de dirección, % dirección a otros cursos, total que asistieron a otros cursos, % de otros cursos, de chequeo físicos los chequeados, % de chequeados, estimulados, % de estimulados, a liberar, % a liberal, total de movimiento que proviene de la reservas del cargo,% total de movimientos, evaluados, % de evaluados, resultados avanza, % de resultados avanza, resultados estables, % de resultados estables, resultados estancados, % resultados estancados, resultados en retroceso, % en retroceso, total de cuadros sancionados, de medidas aplicadas amonestaciones privadas, % de amonestaciones privadas, de medidas

aplicadas amonestaciones públicas, % de amonestaciones públicas, de medidas aplicadas democión temporal, % de democión, de medidas aplicadas democión definitiva, % de democión definitiva, de medidas aplicadas separación definitiva, % de separación definitiva, causas resultados deficientes, % de resultados deficientes, causas incumplimiento de funciones, % de incumplimientos, causas descontrol administrativo, % de descontrol, apelación, % apelación, ratificación de la medida muy débil, % de débil, ratificación de la medida muy severa,% de severa por dirección y mes.

RI22. Obtener total de cuadros, nivel de completamiento cubiertos, % de cubiertos, nivel de completamiento vacantes, % de vacantes, sexo femenino, % femeninos, sexo masculino, % masculino, edad entre 25-40 años, % entre 25-40 años, edad entre 41-50 años, % entre 41-50 años, edad entre 51-60 años,% entre 51-60 años, edad con más de 61 años,% más de 61 años, raza blanca,% de piel blanca, raza negra,% de piel negra, raza mestiza, % de piel mestiza, raza negros y mestizos,% negros y mestizos, nivel escolar 12grado,% con 12grado, nivel escolar técnico medio,% con técnico medio, nivel escolar nivel superior,% nivel superior, militancia UJC,% de UJC, militancia PCC,% que pertenecen al PCC, % de mujeres con cargos decisorios, total general de cuadros que asistieron, % que asistieron, defensa total, % total defensa, total de técnica profesional, % total de técnica profesional, tipos de curso otros curso de dirección, % dirección a otros cursos, total que asistieron a otros cursos, % de otros cursos, estimulados, % de estimulados, evaluados, % de evaluados, resultados avanza, % de resultados avanza, resultados estables, % de resultados estables, resultados estancados, % resultados estancados, resultados en retroceso, % en retroceso, total de cuadros sancionados, de medidas aplicadas amonestaciones privadas, % de amonestaciones privadas, de medidas aplicadas amonestaciones públicas, % de amonestaciones públicas, de medidas aplicadas democión temporal, % de democión, de medidas aplicadas democión definitiva, % de democión definitiva, de medidas aplicadas separación definitiva, % de separación definitiva, causas resultados deficientes, % de resultados deficientes, causas incumplimiento de

funciones, % de incumplimientos, causas descontrol administrativo, % de descontrol, apelación, ratificación de la medida muy débil, ratificación de la medida muy severa por municipio y mes.

2.3 Requisitos funcionales.

Para lograr satisfacer las necesidades del cliente precisadas con anterioridad, se hace necesario identificar las funcionalidades que debe tener el sistema. Son conocidas como requisitos funcionales de la aplicación y se definen durante la etapa de análisis.

Seguidamente se enumeran los requisitos funcionales de la solución:

RF1. Autenticar (usuario)

RF2. Realizar vistas de análisis de los indicadores de la dirección de cuadros.

RF3. Realizar los reportes de los indicadores de la dirección de cuadros.

RF4. Realizar la extracción de los datos.

RF5. Asegurar la carga de los datos desde distintas fuentes.

RF6. Asegurar la limpieza de los datos.

2.4 Requisitos no funcionales.

Los requerimientos no funcionales, simbolizan aquellas características del sistema que le reportan al cliente alguna ventaja, como el rendimiento, la robustez y la fiabilidad, dándole más confianza y seguridad en la aplicación. Luego de un análisis profundo, se definieron 32 requisitos no funcionales para el mercado de datos, los cuales están descritos en el artefacto especificación de requisitos. (Ver expediente del proyecto).

2.5 Roles y permisos

Todo sistema de información, debe detallar con un mecanismo de protección contra acciones que puedan afectar la integridad de los datos. Por tal motivo, la seguridad estará avalada a través de la autenticación de los usuarios, a los cuales se les asigna un rol con un grupo de permisos para interactuar con la aplicación.

Seguidamente se muestran los roles y permisos definidos para la solución:

Administrador: Tiene acceso total para gestionar los permisos, roles y usuarios, pero solo tiene acceso de lectura a las Áreas de Análisis (AA) del mercado de datos.

Analista: Tiene acceso de solo lectura a las AA. También, puede visualizar todos los reportes de estas áreas.

Administrador ETL: Tiene acceso total a todas las AA del mercado de datos.

Además de los roles y permisos definidos, se cuenta con la seguridad que proporciona la Plataforma Pentaho BI, la cual posee un mecanismo de seguridad compuesto por cuatro áreas fundamentales que son: [15]

- Seguridad de acceso a datos de objetos: incluye usuarios, contraseñas, autorizaciones permitidas, recursos web y protección a datos.
- Autenticación: tiene que ver con el procesamiento de información interactiva de inicio de sesión (por ejemplo nombre de usuario y contraseña) comparándola con la información recuperada del almacén de datos de seguridad.
- Autorización de recursos web (URL): brinda protección a las URL para responder a cada usuario si pueden o no acceder a una determinada página. Esto es decidido por el administrador de recursos web, el cual le brinda a cada usuario autenticado un permiso de seguridad, delimitando las páginas a las que tiene acceso y a las que no.
- Autorización a objetos del dominio: en el sistema los únicos objetos del dominio protegidos por la plataforma, son los objetos del repositorio otorgados al usuario autenticado. Es responsabilidad de los objetos del dominio autorizar las operaciones solicitadas por este.

2.6 Casos de uso del sistema.

Durante la fase de análisis y diseño de un mercado de datos, se definen los casos de uso del sistema, que tienen como objetivo describir la relación entre la aplicación

y los usuarios. En ellos, se agrupan todos los requisitos funcionales y de información que hayan sido identificados con anterioridad, proporcionando una mejor comprensión del sistema. En la siguiente figura se muestra el diagrama de casos de uso del sistema (DCUS) para la solución.

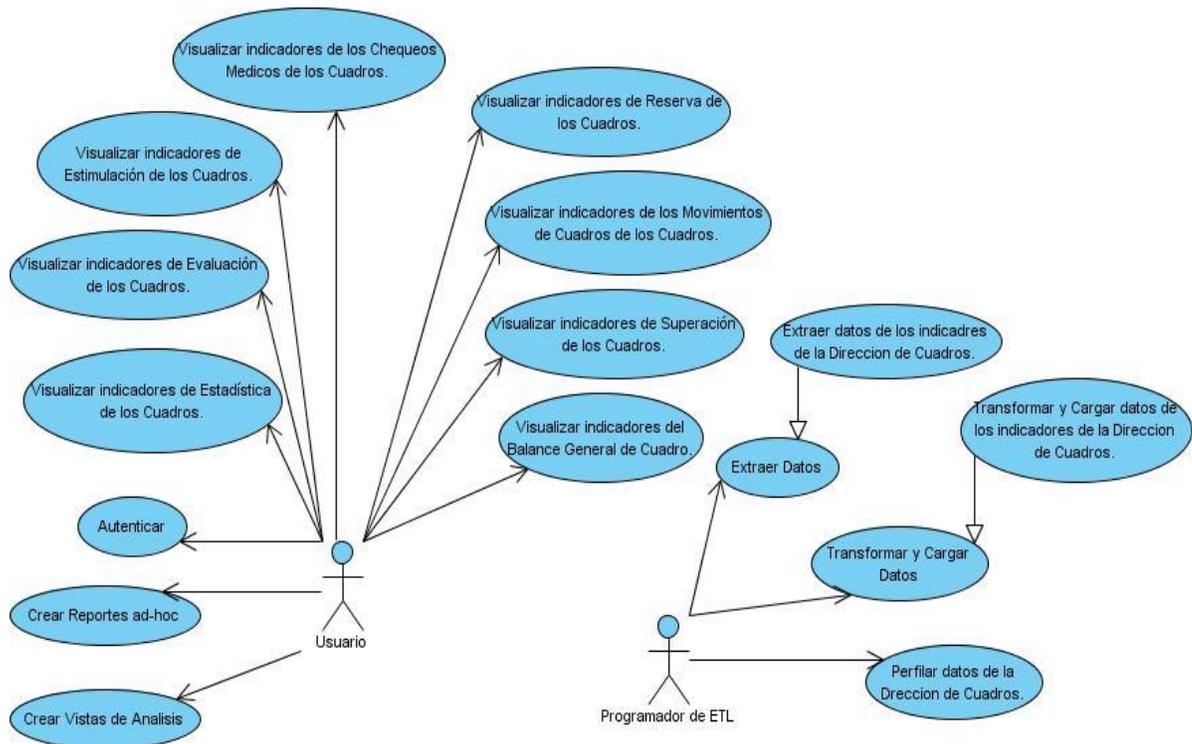


Figura 4: Casos de usos del sistema.

Un caso de uso del sistema, agrupa los requisitos de información que se definen teniendo en cuenta las privaciones del cliente. Por su parte, los casos de uso funcionales se encargan de agrupar las funcionalidades que debe tener el sistema.

Casos de uso del sistema:

Visualizar indicadores de estadísticas de los cuadros: comprende toda la información estadística de los cuadros en la dirección de cuadros.

Visualizar indicadores de estimulación de los cuadros: comprende toda la información de estimulación de los cuadros en la dirección de cuadros.

Visualizar indicadores de los chequeos médicos de los cuadros: comprende

toda la información de los chequeos físicos y estomatológicos de los cuadros en la dirección de cuadros.

Visualizar indicadores de evaluación de los cuadros: comprende toda la información de la evaluación de los cuadros en la dirección de cuadros.

Visualizar indicadores de superación de los cuadros: comprende toda la información de superación en los cursos de los cuadros en la dirección de cuadros.

Visualizar indicadores de reserva de los cuadros: comprende toda la información de reserva en la dirección de cuadros.

Visualizar indicadores de los movimientos de cuadros de los cuadros: comprende toda la información de los movimientos de los cuadros en la dirección de cuadros.

Visualizar indicadores del balance general de cuadros: comprende todo el balance de los cuadros en la dirección de cuadros.

Descripción del caso del uso del sistema: **Visualizar indicadores de estadísticas de los cuadros.**

Caso de Uso:	Visualizar indicadores de estadísticas de los cuadros.	
Tipo:	Información.	
Actores:	Rol o Grupo (Usuario).	
Resumen:		
Precondiciones:	Compleitud del almacén. Carga de los datos.	
Referencias		
Prioridad		
Flujo Normal de Eventos		
Sección “Indicadores de Estadísticas”		
Acción del Actor	Respuesta del Negocio	
Entra al sistema.	Muestra opciones de reportes.	

Selecciona reporte de estadísticas.	Muestra el reporte de Estadísticas deseado.
Prototipo de Interfaz	
Poscondiciones	Disponibilidad de opciones (cruces de variables) de reportes relacionados con la estadísticas

Tabla 1: Caso del uso del sistema visualizar indicadores de estadísticas de los cuadros.

Casos de uso funcionales:

Autenticar: Consiste en el proceso de autenticación por el que deben pasar los usuarios, para controlar el acceso a los datos.

Crear reportes ad-hoc. Consiste en el proceso de creación de los metadatos para los reportes futuros en la dirección de cuadros.

Crear vista de análisis. Realiza la navegación dimensional sobre los cubos desde la plataforma BI y se visualizan los resultados de las consultas, las mismas se muestran al usuario en formato web.

Extraer datos de los indicadores de la dirección de cuadros: Se refiere al proceso durante el cual se extraen los datos de las fuentes de información.

Transformar y cargar datos de los indicadores de la dirección de cuadros: Se refiere a los procesos de transformación y carga por el que pasan los datos, luego de ser extraídos de las fuentes.

Perfilar datos de los indicadores de la dirección de cuadros: Se refiere a perfeccionar el proceso de transformación.

Descripción del caso del uso: **Autenticar.**

Caso de Uso:	Autenticar.
Tipo:	Usuario
Actores:	
Resumen:	El sistema debe estar funcionando correctamente

Precondiciones:	
Referencias	Alta
Prioridad	Autenticar.
Flujo Normal de Eventos	
Sección “Autenticar”	
Acción del Actor.	Respuesta del almacén.
<ul style="list-style-type: none"> El usuario introduce su nombre de usuario y la contraseña. 	El sistema verifica si el usuario y la contraseña sean válidos. 1.2). Si los datos están correctos y el usuario está registrado, el sistema le asigna sus privilegios según su rol y le da acceso al sistema.
Prototipo de Interfaz	
Flujos Alternos	
Acción del Actor	Respuesta del Negocio
Prototipo de Interfaz	
Poscondiciones	La autenticación se realice de forma correcta.

Tabla 2: Caso del uso autenticar.

2.7 Modelo de datos dimensional.

El modelado dimensional, es el nuevo nombre por el que se conoce a una vieja técnica para lograr que las bases de datos fueran más simples y entendibles, pero ha llegado a ser ampliamente aceptada como la técnica dominante para la presentación de los almacenes de datos. Además, ha surgido como la única arquitectura coherente para la construcción de estos sistemas, demostrando ser la estrategia más efectiva en cuanto al costo. El modelado dimensional, también resulta beneficioso en cuanto a diseño, pues posibilita una mejor comprensión de la aplicación por parte de los usuarios, un mayor rendimiento en las consultas y flexibilidad ante los cambios. [16]

En el modelo de datos, se puntualizan las dimensiones y hechos que serán las futuras tablas de la base de datos de la solución. Debido a esto, se hace preciso conocer cada uno de sus elementos, los cuales se explican a continuación con la intención de facilitar su entendimiento:

Hechos

En un modelo dimensional, una tabla de hechos representa una transacción o un evento del negocio y en ella se almacenan un conjunto de medidas o atributos, que permiten cuantificar o medir el rendimiento en los diferentes procesos del mismo. Generalmente, estas tablas poseen su propia llave primaria que se forma por la unión de las llaves pertenecientes a las dimensiones que se relacionan con ella, por lo que también se conoce como llave compuesta. [16]

Dimensiones

Por su parte, las tablas de dimensiones tienden a ser poco profundas en cuanto a la cantidad de filas, pero suelen ser bastante amplias en cuanto al número de columnas o atributos. Estos últimos, son la fuente principal de las restricciones de las consultas y los agrupamientos, por lo que en una consulta o una solicitud de reporte, son los diferentes aspectos por los que se pueden analizar las medidas de los hechos. Además, cada dimensión debe tener una llave primaria que sirva como base para la integridad referencial con cualquier tabla de hechos con la que se relacione. [16]

A continuación, se muestran algunos de los modelos de datos diseñados para la solución, los cuales están descritos en el artefacto Modelo Dimensional (ver expediente de proyecto), para representar mejor las relaciones entre los hechos y las dimensiones:

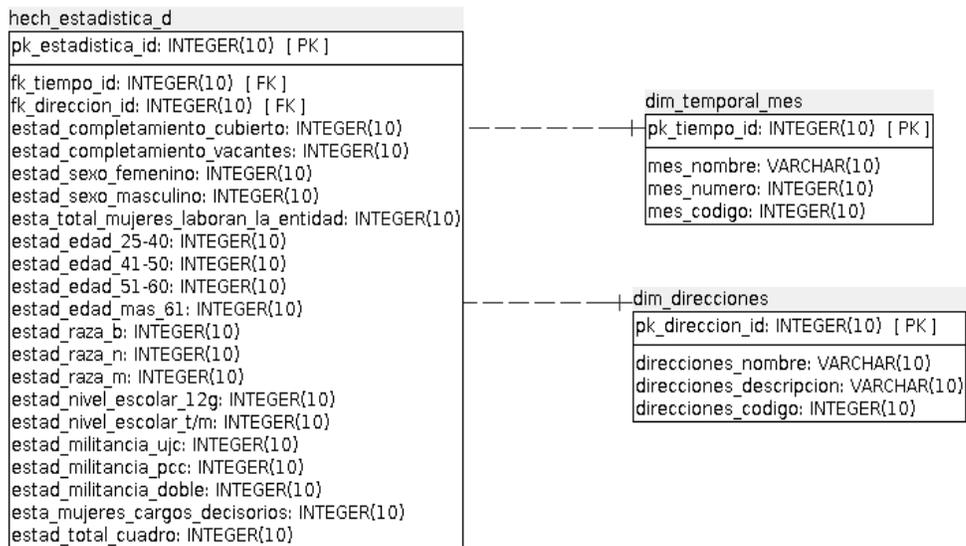


Figura 5: Hecho estadística direcciones

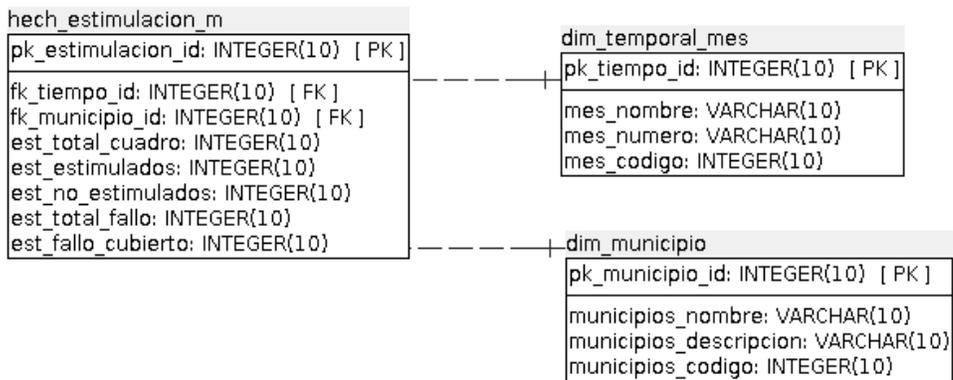


Figura 6: Hecho estimulación municipios

2.8 Identificación de dimensiones y hechos.

Luego de un estudio realizado sobre toda la información que se recoge en la dirección de cuadros, fueron identificados los siguientes hechos y dimensiones:

Lista de hechos:

- hech_chequeo_medico: agrupa los indicadores sobre el chequeo médico físico y estomatológico recogidos en el modelo.
- hech_estadistica_d: agrupa los indicadores sobre los cuadros recogidos en el modelo de direcciones.
- hech_estadistica_m: agrupa los indicadores sobre los cuadros recogidos en el modelo de municipios.

- hech_estimulacion_d: agrupa los indicadores sobre los cuadros estimulados recogidos en el modelo de direcciones.
- hech_estimulacion_m: agrupa los indicadores sobre los cuadros estimulados recogidos en el modelo de municipios.
- hech_evaluacion_d: agrupa los indicadores sobre los cuadros evaluados recogidos en el modelo de direcciones.
- hech_evaluacion_m: agrupa los indicadores sobre los cuadros evaluados recogidos en el modelo de municipios.
- hech_medidas_evaluacion_d: agrupa los indicadores sobre las medidas disciplinarias de los cuadros recogidos en el modelo de direcciones.
- hech_medidas_evaluacion_m: agrupa los indicadores sobre las medidas disciplinarias de los cuadros recogidos en el modelo de municipios.
- hech_movimientos_cuadros: agrupa los indicadores sobre los movimientos de cuadros recogidos en el modelo.
- hech_reserva_d: agrupa los indicadores sobre las reservas de los cuadros recogidos en el modelo de direcciones.
- hech_reserva_m: agrupa los indicadores sobre las reservas de los cuadros recogidos en el modelo de municipios.
- hech_superacion_d: agrupa los indicadores sobre la superación de los cuadros recogidos en el modelo de direcciones.
- hech_superacion_m: agrupa los indicadores sobre la superación de los cuadros recogidos en el modelo de municipios.
- hech_balance_d: agrupa los indicadores sobre las cinco especialidades de los cuadros recogidos en el modelo de direcciones.
- hech_balance_m: agrupa los indicadores sobre las cinco especialidades de los cuadros recogidos en el modelo de municipios.

Lista de dimensiones:

- dim_temporal_mes: agrupa información relacionada con la periodicidad con que se recoge la información de los modelos.
- dim_direcciones: agrupa información relacionada con todas las direcciones que se encuentran en la administración.
- dim_municipios: agrupa información relacionada con todos los municipios de la provincia Artemisa.

2.9 Desarrollo de la Matriz BUS:

La Matriz BUS o matriz dimensional, representa la relación que existe entre las tablas de hechos y las tablas de dimensiones con las que se relaciona. Posteriormente, se muestra la matriz correspondiente al mercado de datos para la dirección de cuadros, donde se reflejan las relaciones.

Matriz BUS			
Hechos	Dimensiones		
	dim_temporal_mes	dim_municipios	dim_direcciones
hech_chequeo_medico	X		X
hech_estadistica_m	X	x	
hech_estadistica_d	X		X
hech_estimulacion_d	X		X
hech_estimulacion_m	X	x	
hech_evaluacion_d	X		X
hech_evaluacion_m	X	x	
hech_medidas_evaluacion_d	X		X
hech_medidas_evaluacion_m	X	x	
hech_movimientos_cuadros	X		X
hech_reserva_d	X		X
hech_reserva_m	X	x	
hech_superacion_d	X		X
hech_superacion_m	X	x	

hech_balance_d	X		X
hech_balance_m	X		X

Tabla 3: Matriz BUS

2.10 Reglas del negocio.

Las reglas del negocio, son políticas que deben cumplirse o condiciones que deben satisfacerse, las cuales regulan algún proceso del negocio. Estas normas o restricciones, son definidas por los especialistas de la dirección de cuadros de acuerdo a sus necesidades de información. A continuación, se muestran las reglas del negocio identificadas para el mercado de datos:

- Los nombres ficheros.xls guardan la información mensual y anualmente.
- Debe existir nomenclador de municipios y entidades.
- Los campos predefinidos se comprueban con los nomencladores.
- Para validar que los datos de los modelos dados por la dirección de cuadros sean los correctos se aplican nuevamente las fórmulas utilizadas por parte del centro de información.
- Se quedan en cero los datos nulos. El responsable de introducir la información debe llenar estos datos con cero.
- Cada modelo tiene especificado cuando lleva campos vacíos y cuando no; en caso de existir campos nulos o vacíos donde no se debe se devuelven los modelos a la dirección de cuadros.
- En caso de encontrar errores en algunos campos debe dirigirse a los jefes de la dirección correspondiente para ser orientados sobre cómo solucionar el error. También se lanzará una alerta con los campos que presenten problemas.
- En la dirección de cuadros cuando un ministerio necesita información, estos se la solicitan al presidente y a su vez el presidente se la solicita al centro de información. Luego el centro de Información se la hace llegar al presidente el cual la brinda al ministerio que realizó la solicitud.

- En caso de error en la información el DOPI (Dirección, Organización, Planificación, Información.) lo soluciona realizando una visita al municipio que brindó la información.
- Cuando la dirección de cuadros necesita información se la solicita al Presidente, el cual aprueba o no la solicitud. En caso de ser aprobada la dirección de cuadros se acerca al presidente de la dirección municipal y solicita la información deseada.
- En caso de encontrar errores en algunos campos debe dirigirse al director o subdirector de la dirección correspondiente para ser orientados sobre cómo solucionar el error.

2.11 Conclusiones parciales

En el capítulo se realizó un estudio de la información que es analizada en la dirección de cuadros de la Administración Provincial de Artemisa, luego se identificaron los requisitos de información, funcionales y no funcionales del sistema. Además, se definió el diseño de caso de uso del sistema y el modelo de datos dimensional donde se reflejan las dimensiones y hechos identificados, así como también se definieron las reglas del negocio conjuntamente con el cliente. Posteriormente, se explicó cómo se operará la seguridad de los datos en la aplicación y las estrategias a seguir para la recuperación de los datos, en caso de ocurrir algún suceso que atente contra la integridad de los mismos.

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

Introducción

El capítulo está dirigido a la implementación de los aspectos relacionados con los procesos de integración de datos, esperando brindar una mayor comprensión de las estrategias y procedimientos utilizados, además se abordan elementos relacionados con la implementación de la capa de inteligencia de negocio, incluyendo la creación de las estructuras necesarias para la navegación y el análisis de los datos.

3.1 Implementación del modelo de datos físicos.

Durante la etapa de análisis y diseño del mercado de datos, fueron definidos los hechos y las dimensiones que se convertirían en las tablas de la base de datos. Para el mercado de datos de la dirección de cuadros se definieron dos esquemas principales, el primero para almacenar las dimensiones comunes a todas las direcciones de la administración y un segundo esquema para las dimensiones y hechos específicos de la dirección.

A continuación se pueden apreciar dichos esquemas con sus tablas correspondientes:

Esquemas	Tablas que lo componen
dimensiones	dim_temporal_mes dim_municipios
mart_cuadro	dim_direcciones hech_chequeo_medico hech_estadistica_d hech_estadistica_m hech_estimulacion_d hech_estimulacion_m

	hech_evaluacion_d hech_evaluacion_m hech_medidas_evaluacion_d hech_medidas_evaluacion_m hech_superacion_d hech_superacion_m hech_movimientos_cuadros hech_reserva_d hech_reserva_m hech_balance_d hech_balance_m
--	--

Tabla 4: Esquemas y tablas de la base de datos.

3.2 Implementación de los subsistemas de integración.

El proceso ETL permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos, limpiarlos, y cargarlos en el mercado de datos para analizar o en otro sistema operacional para apoyar un proceso de negocio.

Extraer: La primera parte del proceso ETL consiste en extraer los datos desde los sistemas de origen, tales como Excel y tablas de la base de datos.

Transformar: La fase de transformación aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados.

Carga: La fase de carga es el momento en el cual los datos de la fase anterior (transformación) son cargados en el sistema de destino del mercado de datos para la dirección de cuadros dependiendo de los requerimientos de la organización.

A continuación, se describe de forma general la estrategia de integración definida para realizar los procesos de ETL, correspondientes a los hechos y dimensiones del mercado de datos.

Primeramente, se extraen los datos de las fuentes conformadas por ficheros Excel para las dimensiones y para los hechos. Una vez efectuada la extracción de los datos, se realizan las validaciones necesarias teniendo en cuenta las reglas del

negocio identificadas y en caso de encontrarse valores incorrectos, el flujo se desvía hacia un fichero Excel donde es almacenado con la correspondiente descripción del error. Si se comprueba que los datos poseen la calidad requerida, se les realiza un conjunto de transformaciones y búsquedas y se procede a su inserción en las tablas del mercado de datos. Seguidamente se muestra ejemplos de transformaciones realizadas a tablas de hechos del mercado de datos.

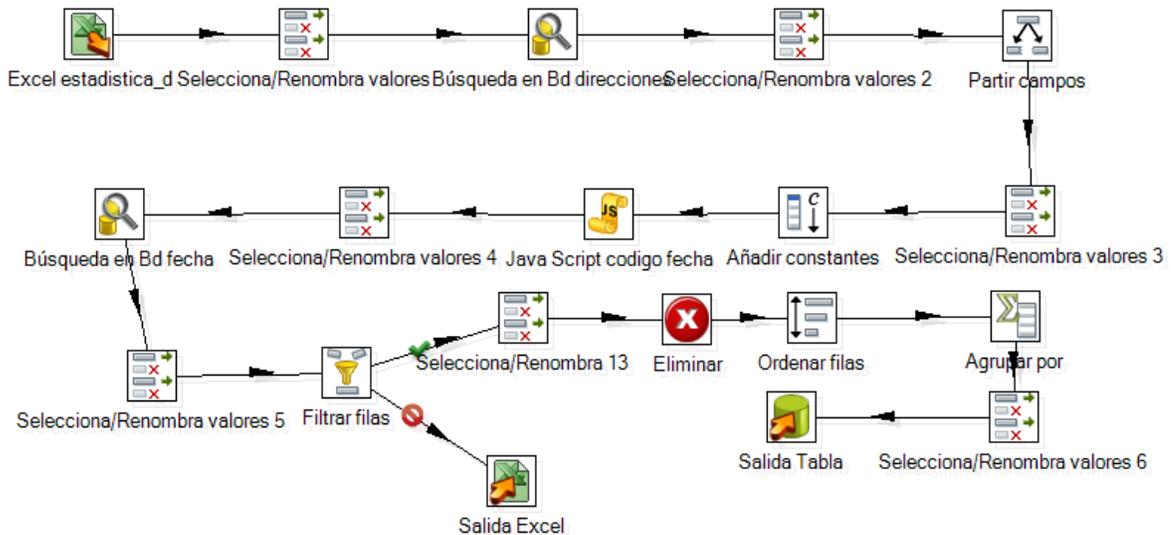


Figura 7: ETL Hecho estadística de las direcciones.

En la figura 7 se carga el Excel de los modelos correspondientes, seguidamente se realizan las búsquedas en la base de datos para comparar que el nombre de la dirección exista en la dimensión direcciones, a continuación se divide el campo fecha, se añade una constante código y mediante un código Java Script se crea el código fecha, el año concatenado con el mes, después de aplicar lo antes mencionado se busca dicho código en la dimensión temporal mes en la base de datos y si coinciden se filtran las filas y si la transformación es correcta se elimina por la fecha para que no existan valores duplicados, se ordenan las filas y se agrupa mediante el tipo de consulta, se inserta en la tabla de la base de datos señalada y en caso contrario de existir algún error los datos se guardan en el Excel de error.

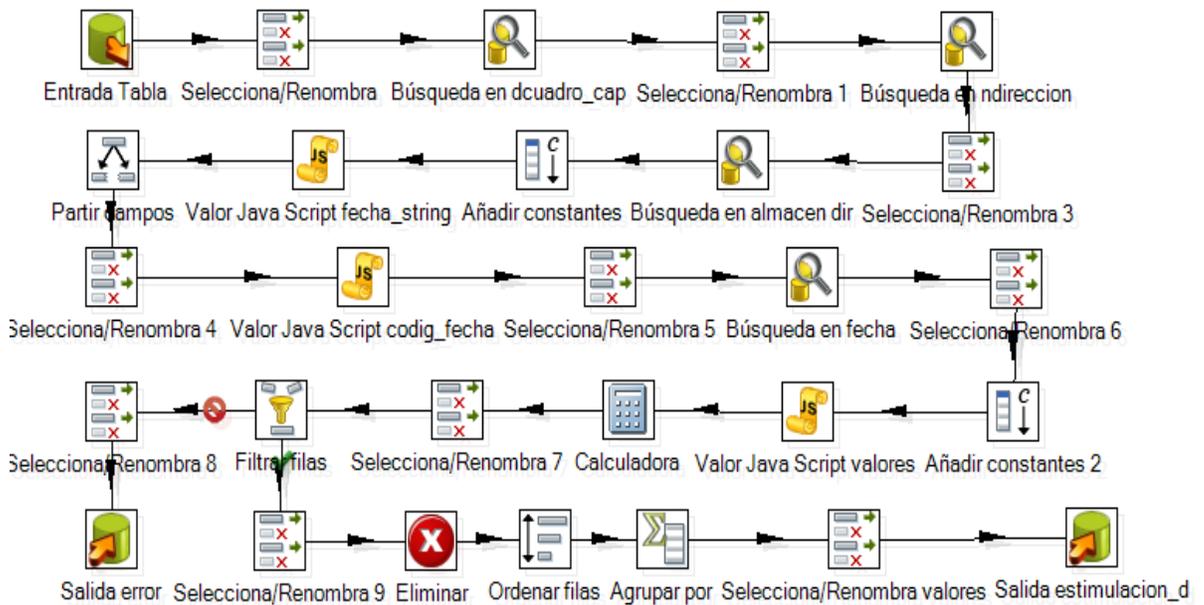


Figura 8: ETL Hecho estimulación de las direcciones.

En la figura 8 se cargan los datos desde la tabla de la base de datos del sistema de gestión de información, se realizan las búsquedas en la base de datos para comparar que el nombre de las direcciones exista en la dimensión direcciones, seguidamente se realiza otra búsqueda para capturar la fecha, se convierte el valor de tipo date a string, se divide el campo fecha, se añade una constante código, mediante un código Java se crea el código fecha, el año concatenado con el mes, después de aplicar lo antes mencionado; se busca dicho código en la dimensión temporal mes en la base de datos, a continuación mediante un código Java se calculan los datos, se filtran las filas y si la transformación es correcta, se elimina por la fecha para que no existan valores duplicados, se ordenan las filas y se agrupa mediante el tipo de consulta, se inserta en la tabla de la base de datos señalada y en caso contrario de existir algún error los datos se guardan en el esquema errores_mart_cuadro.

3.3 Implementación de los trabajos (Jobs).

Un job o trabajo, es una tarea que es asignada para que se ejecute a determinada hora o fecha, es el encargado de ejecutar todas las transformaciones realizadas,

hay diversas formas de hacer un job, esto de acuerdo a las necesidades o ejecutar procedimientos almacenados. Una vez que la conexión al mercado de datos se encuentra en perfecto estado, se procede a la carga del mismo, y el job o trabajo es la forma en que se realiza la carga de los datos hacia el mercado.



Figura 9: Job para la transformación estimulación direcciones.

Una vez realizada todas las transformaciones necesarias se realiza una transformación general para cargar los datos hacia todas las tablas de hechos.

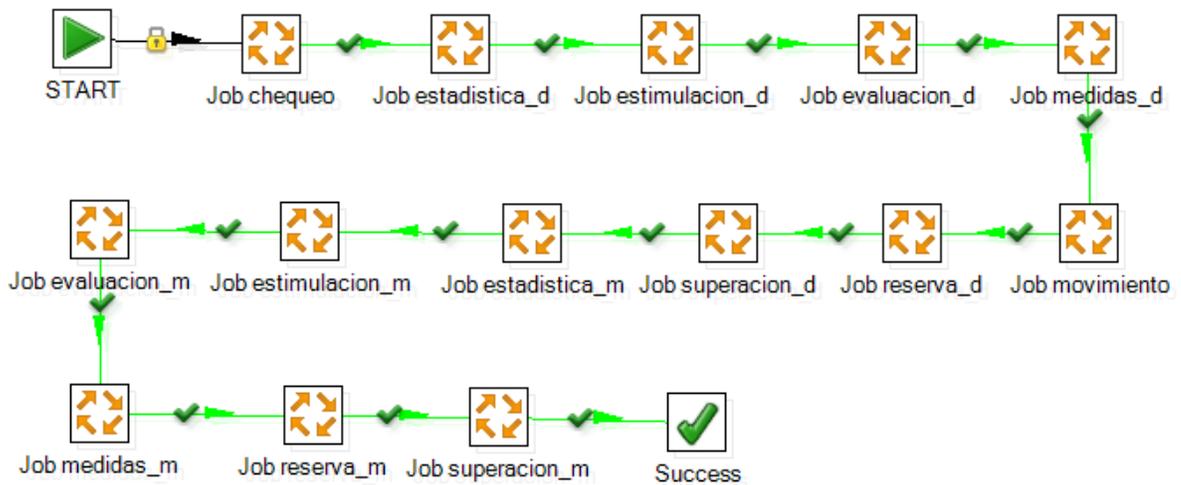


Figura 10: Job para las transformaciones de la base de datos.

3.4 Implementación del subsistema de visualización de datos.

En correspondencia con los requisitos de información, se definieron 22 reportes agrupados en 8 libros de trabajo (LT) ubicados dentro de 8 áreas de análisis (AA). Dichas áreas se corresponde con las áreas de las cuales se van a analizar los reportes, mientras que los LT representan las diferentes categorías a las que pueden pertenecer los reportes, ver Figura 11.



Figura 11: Mapa de navegación.

Cuando se accede al reporte deseado se pueden observar las vistas de análisis generadas, la siguiente figura muestra la vista de análisis referente al cubo correspondiente del hecho (hech_estadistica_d).

		Medidas				
Tiempo	Dirección	• Total Cuadros	• Completamiento Cubierto	• % Cubiertos	• Completamiento Vacantes	• % Vacantes
[-] Fecha	[+] Direcciones	245	187	76,327	60	24,49
[-] 2011	[+] Direcciones	231	185	80,087	48	20,779
febrero	[-] Direcciones	222	182	81,982	40	18,018
	DG Inf y Ser	5	3	60	2	40
	Vivienda	6	4	66,667	2	33,333
	Rec Hidrául	5	2	40	3	60
	Comercio	5	5	100	0	0
	Transporte	3	3	100	0	0

Figura 12: Vista de análisis estadística direcciones.

Se crearon 16 vistas materializadas, que no son más que tablas de hechos con funciones creadas con el fin de mostrar los datos almacenados solamente del último año para el análisis en la dirección de cuadros.

3.4.1 Cubos OLAP

El total de hechos definidos para un mercado de datos, indica la cantidad de cubos OLAP que deben ser creados para la visualización de la información. En correspondencia con los hechos definidos para la dirección de cuadros, se implementaron 16 cubos OLAP y en correspondencia con las vistas materializadas creadas se realizaron 16 cubos OLAP.

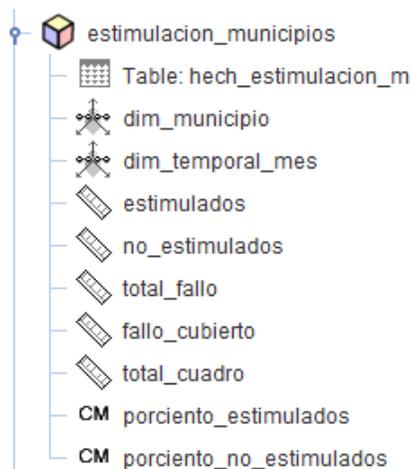


Figura 13: Cubo OLAP correspondiente al hecho (hech_estimulacion_m).



Figura 14: Esquema de cubos OLAP correspondiente a los hechos.

3.4.2 Reportes

Con el fin de generar reportes que precisan ser entregados para su análisis se realizaron una serie de ellos en la herramienta Pentaho Report Designer la cual pertenece a la suite de Pentaho. Un ejemplo se evidencia en el reporte de estadística total, a continuación se muestra cómo quedaría después de su diseño.

Administración Provincial de Artemisa

Dirección de Cuadros

Informe Estadístico Total

Mes: agosto Año: 2011

Total Cuadros	156
Completamiento Cubiertos	2
Completamiento Vacantes	154
Sexo femenino	0
Sexo Masculino	0
Mujeres Laboran Entidad	0

Figura 15: Reporte estadística total.

Es posible que el usuario construya sus propios reportes a partir de la información de los metadatos definidos en el sistema con el uso de la herramienta Pentaho Metadata Editor. Los metadatos son los nombrados reportes ad-hoc, y se clasifican como reportes inmediatos que el usuario puede crear al instante. Para ello se construyeron 16 metadatos descritos a continuación.

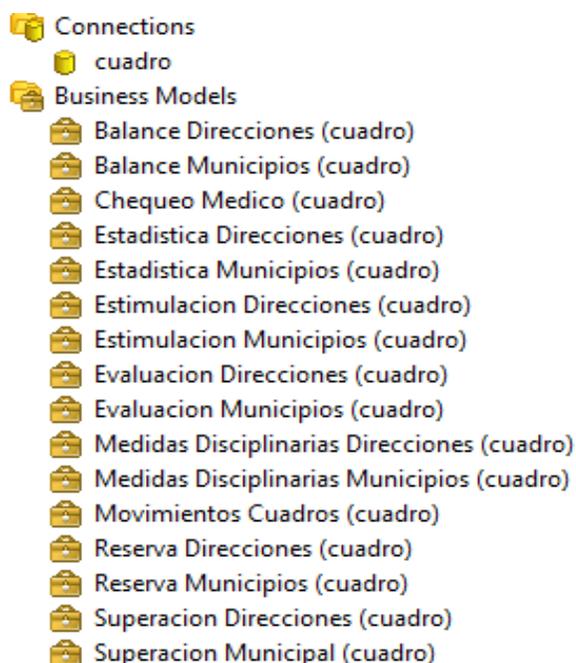


Figura 16: Metadatos del mercado de datos.

3.5 Conclusiones parciales

En el capítulo se abordaron los elementos más significativos relacionados con la implementación del subsistema de integración, donde se evidencian las 32 transformaciones realizadas para la carga de los datos necesarios, así como el orden en que deben ser ejecutadas. Además, se expusieron aspectos de interés relacionados con la organización y visualización de la información, como son la estructura definida para la navegación y la creación de los 32 cubos OLAP con el fin de diseñar las vistas de análisis y fueron precisados los 6 reportes generados para examinar la estadística en la dirección de cuadros de la Administración Provincial de Artemisa, así como también los 16 metadatos diseñados para los reportes ad-hoc.

Capítulo 4: Validación del mercado de datos de la dirección de cuadros de la Administración Provincial de Artemisa.

Introducción

En el capítulo se hace referencia a las pruebas y se muestran los casos de prueba por casos de usos y por requisitos diseñados para probar el funcionamiento del mercado de datos, comprobándose que el resultado logrado sea el anhelado de acuerdo a las necesidades de los clientes de la dirección de cuadros de la Administración Provincial de Artemisa.

4.1 Pruebas

En el desarrollo de cualquier producto de software se realizan diferentes actividades desde que surge la idea inicial hasta la obtención del producto final.

Para llevar a cabo las validaciones del mercado de datos se desarrollan una serie de pruebas la cual son un conjunto de acciones en las cuales un sistema o componente es ejecutado bajo unas condiciones específicas, donde los resultados son observados y registrados para dar una evaluación de algún aspecto del sistema. En ellas se describen cómo comprobar cada versión del sistema durante la integración del mismo, verificando que todos los requerimientos hayan sido implementados.

Algunas de las pruebas que pueden ser utilizadas para la validación de un software:
[17]

Pruebas unitarias: esta prueba centra el proceso de verificación en la menor unidad del diseño del software: el componente de software o módulo.

Pruebas de integración: es una técnica sistemática cuyo objetivo consiste en probar el sistema como un todo, para detectar errores asociados con la interacción entre los diferentes módulos que lo componen.

Pruebas de validación: proporciona una seguridad final que el software satisface todos los requisitos funcionales, de comportamiento y de rendimiento.

Pruebas de regresión: consiste en volver a ejecutar un subconjunto de pruebas que se han llevado a cabo anteriormente, para asegurarse que los cambios que se hayan realizado no introduzcan un comportamiento no deseado o errores adicionales.

Pruebas de aceptación: se realizan para probar que el sistema cumpla con los requerimientos y expectativas del cliente.

Es importante aclarar que las pruebas no pueden asegurar la ausencia de defectos; solo pueden demostrar que existen desperfectos en el software y que cada prototipo que se quiera entregar debe ser probado y evaluado a diferentes niveles.

4.1.1 Diseño de los casos de pruebas.

Para la conformidad del mercado de datos de la dirección de cuadros, se diseñaron ocho casos de prueba, con el propósito de verificar los requisitos, agrupados en ocho casos de uso del sistema que fueron definidos previamente durante la etapa de análisis.

En la siguiente tabla, se muestra un fragmento del caso de prueba correspondiente al caso de uso: Visualizar indicadores de estimulación de los cuadros, donde se aprecia el reporte correspondiente al LT: indicadores de estimulación, además de los perfiles de análisis correspondientes, los indicadores que se miden en el reporte, la respuesta que debe dar el sistema y el flujo central de ejecución del caso de prueba, los cuales están descritos en los artefactos Casos de Pruebas. (Ver expediente del proyecto).

Escenario	EC1.1 Indicadores de estimulación.
Descripción	Muestra los reportes de los indicadores de estimulación.
Perfiles de análisis	Municipio, Temporal_mes, Direcciones.
Indicadores a medir	est_total_cuadro, est_estimulados, est_no_estimulados, est_total_fallo, est_fallo_cubierto,

	est_total_cuadro_d, est_estimulados_d, est_no_estimulados_d, est_total_fallo_d, est_fallo_cubierto_d.
Respuesta del sistema	El sistema muestra los reportes de los indicadores de estimulación.
Flujo central	<p>Se abre la aplicación.</p> <p>Se autentifica.</p> <p>Se entra al sistema.</p> <p>En el lateral izquierdo donde aparece el navegador se selecciona el área de análisis A.A: Análisis de los indicadores de estimulación de la dirección de cuadros.</p> <p>LT: Indicadores seleccionados sobre estimulación.</p> <p>En la parte inferior izquierda se selecciona el reporte deseado.</p> <p>En el área de trabajo se visualiza la tabla correspondiente al reporte y las vistas correspondientes a esta área.</p>
Resultado de las pruebas	Satisfactorio.

Tabla 5. Diseño del caso de prueba: Visualizar indicadores de estimulación de los cuadros.

4.2 Estrategias de pruebas para el plan de pruebas.

Con el fin de comprobar la calidad del proceso de implementación, en el cual, se recogen precondiciones y medidas que facilitan la identificación de resultados verídicos y eficientes se muestra a continuación algunos ejemplos de las pruebas establecidas para la capa de visualización, los cuales están descritos en el artefacto Plan de Pruebas (Ver expediente del proyecto):

Prueba de acceso a la información por roles o niveles de usuario:

Descripción	Precondiciones	Medidas
Verifica el acceso y privilegios que posee cada usuario o rol de la dirección de cuadros, definido con respecto al sistema.	Acceso restringido a la información, solo se puede ver la información necesaria del Mercado de datos para la dirección de cuadros para el trabajo de cada usuario o rol. Cada usuario o rol debe tener acceso pleno a la información que maneja, no pueden existir información inaccesible para ellos si en el negocio fue definido su acceso.	Cumple con los privilegios de acceso.

Tabla 6. Prueba de acceso a la información por roles o niveles de usuario.

Prueba de consistencia de los datos:

Descripción	Precondiciones	Medidas
Se insertan un número finito de datos a través de proceso de ETL en las tablas de cada hecho. Buscando que se ejecute el ciclo completo hasta la capa de visualización, donde se verifica que los datos insertados sean los mismos que los visualizados en la aplicación.	Datos confiables, sin existencia de duplicidad, u otros errores de cálculos en las estructuras.	Eficiente: el conjunto de datos de entrada debe ser el mismo que el de salida. Parcial: el conjunto de datos de salida es el 80 % del conjunto de datos de entrada. Deficiente: el conjunto de datos de salida es menor del 80 % del conjunto de datos de entrada.

Tabla 7. Prueba de Consistencia de los Datos.

El resultado obtenido de la prueba de consistencia de los datos fue: Eficiente

Prueba de las estructuras multidimensionales:

Descripción	Precondiciones	Medidas
Se verifica que las estructuras construidas (cubos) respondan a todo el alcance de información que fue definido en el negocio por los clientes y analistas de la dirección de cuadros.	Estructuras sólidas que respondan a las necesidades de información que se definieron en el negocio.	<p>Eficiente: Las estructuras cumplen con el alcance de información previsto.</p> <p>Parcial: Las estructuras cumplen en un 80% con el alcance de información previsto.</p> <p>Deficiente: Las estructuras cumplen menos del 80% con el alcance de información previsto.</p>

Tabla 8. Prueba de las estructuras multidimensionales.

El resultado obtenido de la prueba de las estructuras multidimensionales fue: Eficiente.

4.3 Resultados y Funcionalidades Obtenidas.

Como resultado de este trabajo de diploma se tiene disponible en su versión 1.0 el mercado de datos para la dirección de cuadros de la Administración Provincial de Artemisa. Se obtuvo un sistema que cumple con todas las especificaciones, para garantizar la accesibilidad, calidad, disponibilidad y el análisis de la información en apoyo a la toma de decisiones en la dirección.

Entre las funcionalidades más significativas del mercado de datos se pueden señalar:

- Permite crear vistas de análisis.
- Permite realizar los reportes del negocio.
- Permite crear reportes ad-hoc.

- Proporciona al usuario una interfaz consolidada, única para los datos, que hace más fácil el trabajo con las consultas para la toma de decisiones.

4.4 Aporte social y económico

En la dirección de cuadros, según la problemática planteada, hay consecuencias desfavorables en cuanto a accesibilidad a datos, dificultad de la disponibilidad de información estadística para los altos mandos de la Administración Provincial de Artemisa y limitaciones para recuperar indicadores.

El mercado de datos para la dirección de cuadros ofrece a los especialistas de la dirección guardar los datos durante períodos de tiempo extensos, brinda una vía eficiente y rápida de realizar su trabajo evitando la elaboración de informes costosos en esfuerzo y tiempo pues se consulta gran cúmulo de información de las diferentes especialidades.

El mercado de datos propuesto favorece el desempeño profesional para la administración de información lo que permite facilitar la toma de decisiones de la dirección de cuadros, esta propuesta ahorra recursos dando la posibilidad de que la información se administre utilizando las nuevas tecnologías y permitiendo el acceso a esta en el momento preciso sin necesidad de que esté presente el especialista, evita la pérdida de tiempo que ocasionaba el análisis que hasta este momento realizaban los especialistas manualmente.

4.5 Conclusiones parciales

En este capítulo se estableció la estrategia seguida para la aprobación del mercado de datos. Y se diseñaron y aplicaron ocho casos de pruebas para verificar el resultado del mercado de datos para la dirección de cuadros de la Administración Provincial de Artemisa. También se estableció el plan de pruebas para probar la habilidad del proceso de implementación y para comprobar el cumplimiento de los objetivos definidos, conjuntamente con los resultados obtenidos luego de ser aplicadas.

Conclusiones generales

Se obtuvo un mercado de datos para la dirección de cuadros de la Administración Provincial de Artemisa, que integra toda la información perteneciente a las cinco especialidades de la dirección, facilitando el análisis de los datos para la toma de decisiones.

En esta investigación se abordaron los siguientes temas:

- Aspectos teóricos relacionados con almacenes y mercados de datos, realizando un estudio de los sistemas a nivel mundial y nacional, se evidenció la existencia de mercados de datos similares pero que no pueden ser utilizados por su arquitectura.
- Se definió la metodología para definir las fases de construcción del mercado y las herramientas a utilizar en su desarrollo.
- Con el análisis y diseño se obtuvo 16 tablas de hecho y 16 vistas materializadas para el modelado de los datos.
- Se realizó la implementación de todas las transformaciones y trabajos necesarios para la carga de los datos.
- Se dio respuesta a 32 vistas de análisis y 6 reportes, permitiendo la disponibilidad de los datos para el análisis y desarrollo del mercado de datos de la dirección de cuadros de la Administración Provincial de Artemisa.
- Se efectuaron 8 casos de prueba por casos de uso y 22 casos de prueba por requisito, comprobando el funcionamiento óptimo del sistema.

Recomendaciones

- Incentivar el estudio sobre los almacenes y mercados de datos.
- Desplegar el mercado de datos para la dirección de cuadros de la Administración Provincial de Artemisa.
- Mantener actualizado el mercado de datos ante los posibles cambios que ocurran para con los datos almacenados.
- Abrir líneas de investigación cuyos objetivos sean la aplicación de algoritmos de minería de datos a la base existente de manera que contribuyan a la transformación del conocimiento tácito a conocimiento explícito.

Referencias Bibliográficas

- [1]. Alberto Límia Navarro, Anisley Delfino, Asnioby Hernández López, Doris Medina Mustelier, Iván M. Cárdenas Tandrón, Julio Ernesto Ortiz, Marisleidys Socas, Osniel Calvo, Yanet Peña Vazquez, Yanisbel Gonzales Hernández, otros compañeros del Centro UCI-Villa Clara. Metodología de Desarrollo de Soluciones BI y Warehousing (BI & W). Ciudad de La Habana.
- [2]. Mundy & Thornthwaite 2006, Kimball et al 1998, Kimball & Caserta 2004, Kimball & Ross 2002, Kimball & Merz 2000, Kimball & Ross 2010.
- [3]. Roberto Hernando Velasco [En línea] [Citado el 20 de octubre del 2011.] <http://www.rhernando.net>
- [4]. Mercado de datos [En línea] [Citado el 20 de octubre del 2011.] <http://www.mitecnologico.com/Main/DefinicionesConceptosMercadosDatos>
- [5]. Mercado de datos [En línea] [Citado el 18 de noviembre del 2011.] http://datamarting.org/info/index.php?option=com_content&task=view&id=28&Itemid=1
- [6]. Mundy & Thornthwaite, The Microsoft Data Warehouse Toolkit—With SQL Server 2005 and the Microsoft Business Intelligence Toolset, Indianapolis, Wiley, 2006.
- [7]. Inteligencia del negocio [En línea] [Citado el 21 de noviembre del 2011.] www.sinnexus.com/business_intelligence/olap_avanzado.aspx
www.ucasal.net/templates/unid-academicas/ingenieria/apps/5-p56-rivadere-formateado.pdf
- [8]. Visual Paradigm. [En línea] 2010. [Citado el 25 de noviembre del 2011.] <http://www.visual-paradigm.com>
- [9] Oracle. World's Largest Data Warehouse, 2006.
- [10] González, Carlos D. Curso Base de Datos PostgreSQL, SQL avanzado y PHP. [En línea] [Citado el: 28 de Noviembre del 2011.] http://www.usabilidadweb.com.ar/postgre.php_

- [11]. PgAdmin [En línea] [Citado el 29 noviembre del 2011.] http://www.guia-ubuntu.org/index.php?title=PgAdmin_III
- [12]. Pentaho [En línea] [Citado el 29 noviembre del 2011.] www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/
- [13]. Schema-Workbench [En línea] [Citado el 24 de octubre del 2011.] churriwifi.wordpress.com/2010/07/04/17-3-preparando-el-análisis-dimensional-definición-de-cubos-utilizando-schema-workbench/
- [14]. Chopra, Vivek, Li, Sing y Genender, Jeff. Professional Apache Tomcat 6. Indianapolis: Wiley Publishing, 2007. ISBN: 9780471753612.
- [15]. Pentaho. [En línea] [Último acceso: 19 de febrero de 2011] <http://wiki.pentaho.com/display/ServerDoc1x/03.+Platform+Security+Implementation>
- [16]. Kimball, Ralph y Ross, Margy. The Data Warehouse Toolkit. s.l.: Wiley Computer Publishing. ISBN: 0-471-20024-7.
- [17] Scalone, Fernanda. “Estudio comparativo de los modelos y estándares de calidad del software”. Universidad tecnológica nacional, Facultad regional de Buenos Aires. 2006.

Bibliografía

1. Alberto Límia Navarro, Anisley Delfino, Asnioby Hernández López, Doris Medina Mustelier, Iván M. Cárdenas Tandrón, Julio Ernesto Ortiz, Marisleidys Socas, Osniel Calvo, Yanet Peña Vazquez, Yanisbel Gonzales Hernández, Otros compañeros del Centro UCI-Villa Clara. Metodología de Desarrollo de Soluciones BI y Warehousing (BI & W). Ciudad de La Habana.
2. Bouman, Roland y van Dongen, Jos. Pentaho Solutions, Business Intelligence and Data Warehousing with Pentaho and MySQL. s.l.: Wiley Publishing. ISBN: 978-0-470-48432-6.
3. Chopra, Vivek, Li, Sing y Genender, Jeff. Professional Apache Tomcat 6. Indianapolis: Wiley Publishing, 2007. ISBN: 9780471753612.
4. González, Carlos D. Curso Base de Datos PostgreSQL, SQL avanzado y PHP. [En línea] [Citado el: 28 de Noviembre del 2011.] <http://www.usabilidadweb.com.ar/postgre.php>
5. Hernández, A. Coello (2002). El paradigma cuantitativo de la investigación científica. Ciudad de La Habana. Editorial universitaria.
6. Integración de datos [En línea] [Citado el 21 de noviembre del 2011.] <http://www.ucasal.net/templates/unid-academicas/ingenieria/apps/5-p56-rivadera-formateado.pdf>
7. Inteligencia del negocio [En línea] [Citado el 21 de noviembre del 2011.] www.sinnexus.com/business_intelligence/olap_avanzado.aspx
www.ucasal.net/templates/unid-academicas/ingenieria/apps/5-p56-rivadera-formateado.pdf
8. Kimball, Ralph y Caserta, Joe. The Data Warehouse ETL Toolkit Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. s.l.: Wiley Publishing. ISBN: 0-764-57923-1.
9. Kimball, Ralph y Ross, Margy. The Data Warehouse Toolkit. s.l.: Wiley Computer Publishing. ISBN: 0-471-20024-7.
10. Kimball, Ralph, Reeves, Laura, Ross, Margy y Thorntwhaite, Warren. The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing,

and Deploying Data Warehouses. s.l.: Wiley Publishing. ISBN: 978-0-471-25547-5.

11. Mercado de datos [En línea] [Citado el 18 de noviembre del 2011.]
http://datamarting.org/info/index.php?option=com_content&task=view&id=28&Itemid=1
12. Mercado de datos [En línea] [Citado el 20 de octubre del 2011.]
<http://www.mitecnologico.com/Main/DefinicionesConceptosMercadosDatos>
13. Mundy & Thornthwaite 2006, Kimball et al 1998, Kimball & Caserta 2004, Kimball & Ross 2002, Kimball & Merz 2000, Kimball & Ross 2010.
14. Mundy & Thornthwaite, The Microsoft Data Warehouse Toolkit—With SQL Server 2005 and the Microsoft Business Intelligence Toolset, Indianapolis, Wiley, 2006.
15. Oficina Nacional de Estadísticas, ONE. [En línea] [Citado el: 30 de noviembre del 2011.] <http://www.one.cu>
16. ONE [En línea] [Citado el 25 de octubre del 2011.]
http://www.ecured.cu/index.php/Oficina_Nacional_de_Estad%C3%ADsticas
17. Pentaho [En línea] [Citado el 29 noviembre del 2011.]
www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/
18. Pentaho, open source business intelligence. [En línea] [Citado el: 4 de diciembre de 2011.]
http://mondrian.pentaho.com/documentation/workbench.php_
19. Pentaho. [En línea] [Último acceso: 19 de febrero de 2011]
<http://wiki.pentaho.com/display/ServerDoc1x/03.+Platform+Security+Implementation>
20. PgAdmin [En línea] [Citado el 29 noviembre del 2011.] http://www.guia-ubuntu.org/index.php?title=PgAdmin_III
21. Pressman, R.S. Ingeniería del software. Un enfoque práctico. s.l.: McGraw-Hill. ISBN: 84-481-3214-9.
22. Roberto Hernando Velasco [En línea] [Citado el 20 de octubre del 2011.]
<http://www.rhernando.net>

23. Scalone, Fernanda. “Estudio comparativo de los modelos y estándares de calidad del software”. Universidad tecnológica nacional, Facultad regional de Buenos Aires. 2006.
24. Schema-Workbench [En línea] [Citado el 20 de octubre del 2011.] <http://www.upscene.com/displaynews.php?item=20110330>
25. Schema-Workbench [En línea] [Citado el 24 de octubre del 2011.] churriwifi.wordpress.com/2010/07/04/17-3-preparando-el-análisis-dimensional-definición-de-cubos-utilizando-schema-workbench/
26. Visual Paradigm. [En línea] 2010. [Citado el 25 de noviembre del 2011.] <http://www.visual-paradigm.com>