

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



**Título: “Mercado de Datos para la Dirección de Salud de la
Administración Provincial de Artemisa.”**

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

Autor (es): Geidy Acosta Méndez

Reynel Iglesias Aldecoa

Tutor: Lic. Yuray Duarte Lugo.

Co-Tutor: Ing. Orelvi Gásquez Martínez.

Artemisa, Junio de 2012

“Año 54 de la Revolución”

*Nuestro miedo más profundo no es que seamos inadecuados.
Nuestro miedo más profundo es que somos poderosos sin límite.
Es nuestra luz, no la oscuridad lo que más nos asusta.
Nos preguntamos: ¿quién soy yo para ser brillante, precioso,
talentoso y fabuloso?
En realidad, ¿quién eres tú para no serlo?*

Nelson Mandela

Declaración de Autoría

Declaración de Autoría.

Declaramos que somos autores de la presente tesis y reconocemos, a la Universidad de las Ciencias Informáticas, sus derechos patrimoniales sobre la misma con carácter exclusivo.

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

	_____	_____
Autor (es)	Geidy Acosta Méndez.	Reynel Iglesias Aldecoa.
	_____	_____

Tutor (es)	Lic. Yuray Duarte Lugo.	Ing. Orelvi Gásquez Martínez
-------------------	-------------------------	------------------------------

De Geidy:

Dedico este trabajo a mi mamá por ser la luz de mis días y mi máxima fuente de inspiración. Je adoro.

De Reynel:

Dedico este trabajo a quien es mi razón de ser, a mi hijo Kevin, a mis padres por su esfuerzo y sacrificio en estos 24 años de edad y 18 años de estudio. A mis hermanos y a toda mi familia por su apoyo y preocupación.

Agradecimientos

De Ambos:

A nuestros tutores Yuray y Orneli por dedicarnos tiempo y esfuerzo, por su asesoramiento en nuestras dudas y por sus críticas que ayudaron al perfeccionamiento del trabajo.

A todo el tribunal por sus recomendaciones y la ayuda que nos dieron.

De Zeidy:

A mi mamá por todo el sacrificio, por su apoyo incondicional, por convertir cada obstáculo en un punto de apoyo.

A mi abuela por ser tan buena y por preocuparse de mis cosas.

A mi tía, a mi tío y a mi prima por estar presentes cuando los necesité.

A mi primo Kendry por ayudarme con mi ppt con solo 6 añitos.

A mi primo Dorge por creer siempre en mí y levantarme tanto los ánimos.

A mis amigas Dainelys, Delia, Yanet por estar en los momentos más importantes de mi vida.

A mis amigas de la universidad Arleny y Yaima por apoyarme en todas mis decisiones y estar siempre en las buenas y en las malas.

A mi compañero de tesis Reynel que es mi ídolo.

A mi amigo Yusnel por ser la personita más buena y de los mejores sentimientos que he conocido.

A toda mi familia mis tías, mis tíos, a mis primos y a mi madrina que siempre han estado pendientes de mí.

Agradezco a todas las personas con las cuales he compartido estos maravillosos años, a la gente de mi grupo, a todos aquellos que alguna vez me preguntaron cómo

Agradecimientos

me iba en la tesis, a los que me han ayudado de una forma u otra a realizar este sueño.

De Reynel:

Agradezco primeramente a Dios por haberme dado el privilegio de conocerle formando en mí una nueva persona; por cuidarme y guiarme en todo momento, por darme sabiduría y entendimiento; por estar siempre conmigo.

A mi madre Isabel Aldecoa Herrera y a mis padres Reinier Iglesias Gomes y Julio Iglesias Gomes por su amor y cariño, por su apoyo en todas mis decisiones, por el esfuerzo y trabajo con el cual me han criado haciendo de mí lo que ahora soy.

A mi esposa por su agradable compañía.

A mis abuelos y al resto de mi familia por su incondicional apoyo y preocupación.

A mis amigos, que he cosechado aquí; de manera especial a Monty, Yasiel, Carliobis, Moro, Alejandro, Juanmy y Fddy.

A mis compañeros de grupo y de apartamento que hemos navegado en el mismo barco y aunque muchos quedaron en el camino la mayoría llegamos a tierra firme.

A mi compañera de tesis Geidy porque sin ella este trabajo no hubiera sido posible.

En fin a todos aquellos que de una manera u otra influyeron positivamente en el desarrollo de la tesis, así como en la estancia en esta universidad.

Resumen

El presente trabajo de diploma va dirigido a la Dirección de Salud de la Administración Provincial de Artemisa, el mismo se enmarca en el tema de los almacenes de datos, los mercados de datos y su utilización para los análisis estadísticos de la información. Se detalla la metodología, topología, herramientas, especificación de requerimientos, junto con las necesidades del cliente, para lograr un buen diseño e implementación de los procesos de integración y análisis de datos de la solución. Como resultado se obtiene un mercado de datos poblado disponible para hacer análisis OLAP, además se tiene la estructura del modelo dimensional que comprende: las dimensiones, las jerarquías, las tablas de hechos, y las medidas necesarias para proceder con los cálculos y análisis estadísticos. Se precisan las reglas del negocio utilizadas y se detalla el proceso de carga de los datos de la fuente al mercado de datos Salud de la Administración Provincial de Artemisa. De igual manera, la solución incluye las políticas de seguridad, respaldo y recuperación de los datos, así como las pruebas para la validación del mercado. Palabras claves: almacén de datos, mercado de datos, dimensión, hecho, medidas.

INTRODUCCIÓN	1
CAPÍTULO 1. FUNDAMENTOS TEÓRICOS DE LOS ALMACENES DE DATOS	9
1.1 FUNDAMENTOS TEÓRICOS DE LOS ALMACENES DE DATOS.....	9
1.2 ALMACENES DE DATOS (ADs).....	11
1.3 MERCADO DE DATOS (MDs). CARACTERÍSTICAS Y METODOLOGÍA.....	15
1.4 MODOS DE ALMACENAMIENTO DE DATOS.....	16
1.5 INTEGRACIÓN DE DATOS. CARACTERÍSTICAS.....	18
1.6 INTELIGENCIA DE NEGOCIOS.....	19
1.7 MODELO DE DATOS.....	20
1.8 SELECCIÓN DE TOPOLOGÍA, HERRAMIENTAS DE DESARROLLO.....	24
CONCLUSIONES.....	28
CAPÍTULO 2. ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS SALUD	29
2.1 ANÁLISIS Y DEFINICIÓN DE NEGOCIOS.....	29
2.2 DISEÑO Y PROPUESTA DEL SISTEMA.....	35
2.3 DIMENSIONES Y TABLAS DE HECHOS.....	38
2.3 MODELO DE DATOS.....	40
2.4 ESQUEMA DE SEGURIDAD.....	41
CONCLUSIONES.....	42
CAPÍTULO 3. IMPLEMENTACIÓN DEL MERCADO DE DATOS SALUD	43
3.1 IMPLEMENTACIÓN DE LA BASE DE DATOS.....	43
3.2 USUARIOS Y PRIVILEGIOS.....	44
3.3 IMPLEMENTACIÓN DEL SUBSISTEMA DE INTEGRACIÓN DE DATOS. ARQUITECTURA DEL SUBSISTEMA DE INTEGRACIÓN.....	45
3.4 IMPLEMENTACIÓN DEL SUBSISTEMA DE VISUALIZACIÓN DE DATOS. CUBOS OLAP.....	48
3.5 CONFIGURAR LA SEGURIDAD DE LOS USUARIOS.....	51
CONCLUSIONES.....	52
CAPÍTULO 4. VALIDACIÓN MEDIANTE CASOS DE PRUEBAS	53
4.1 PRUEBAS.....	53
4.2 DISEÑO DE LOS CASOS DE PRUEBA.....	53
CONCLUSIONES.....	57
CONCLUSIONES:	58
RECOMENDACIONES	59
REFERENCIAS	60
BIBLIOGRAFÍAS	62
ANEXOS	64

Introducción

En la antigüedad el hombre primitivo con el fin de buscar su bienestar, comenzó a gestionar la información con la que contaba, por lo que fue necesario que acumulara experiencias y necesidades.

La selección de qué tareas tenía que realizar cada individuo y cómo debían de organizarse para capturar una presa, fue generando poco a poco una manera de búsqueda de respuestas a sus interrogantes, encontró que la experiencia y el análisis de la información de los fenómenos podía ser de utilidad. A partir de ese momento ya no sería igual, el hombre trataría de gestionar la información.

La gestión de la información se convierte en una forma de marcar la diferencia y hacer ventaja competitiva en un mundo globalizado. En este sentido, simples formatos y registros son calificados como herramientas básicas de recopilación de información en especial necesidades de clientes, quejas, reclamos e incluso nuevos servicios solicitados. Esto ayuda también a la incorporación de factores de innovación en las empresas, como nuevas tecnologías de la información y de la comunicación, las cuales hacen más sencillo la incorporación de Bases de Datos.

Las Bases de Datos se han convertido en una herramienta fundamental de control y manejo de las operaciones comerciales. Fue así como en pocos años grandes empresas y negocios a raíz del empleo de esta herramienta comenzaron a generar un considerable número de información almacenadas en diferentes fuentes de datos, alcanzado estas un tamaño considerablemente grande, brindándole ventajas a estas entidades para la toma de decisiones y el desarrollo de las mismas.

Con el transcurrir del tiempo estas Bases de Datos comenzaban a generar gran acumulación de información. Los directivos de tales empresas y negocios se dieron cuenta que estas podían tener un fin útil al estar reflejadas la mayoría de sus operaciones comerciales durante los ciclos de negocios propios del mercado.

Debido a que las empresas demandan mayor rapidez y eficiencia en la entrega de productos y mejora en todos los servicios existentes, se hace imprescindible

encontrar formas más eficaces de distribuir los productos, más facilidades para hacer estudios de mercado basados en la información de las operaciones comerciales de las empresas y de sus clientes logrando mayor rapidez a la hora de tomar decisiones.

Por tanto pensaron en lo ideal que sería unificar las diferentes fuentes de información de las cuales disponían en un único lugar, al que solamente se le incorporaría información relevante sobre la base de una estructura organizada, integrada, lógica, dinámica y de fácil explotación. La respuesta a esto fueron los Almacenes de Datos.

Sin embargo para hacer un uso eficiente de la información histórica almacenada en un Almacén de Datos para la ayuda a la toma de decisiones, era vital garantizar que estos datos fueran fáciles de obtener, estandarizados y confiables. El problema de la limpieza de datos es poco tratado o evitado por muchas empresas al no considerar adecuadamente el impacto para el negocio de tener almacenada información deficiente.

Los Almacenes de Datos son una tecnología que ha tenido un gran impacto en el medio informático y empresarial desde el momento en que surgió, antes de su nacimiento solamente se trabajaban con datos “fríos”, proporcionados por las Bases de Datos y si se quería hacer un análisis empresarial para el mejoramiento de esta empresa, habían ciertos datos que simplemente eran imposibles de obtener y por lo que el análisis era casi improductivo.

Con la herramienta de los Almacenes de Datos, las empresas se hicieron más grandes, ya que un mejor análisis implica una mejor toma de decisiones y esto a la vez genera un crecimiento de las empresas. Podría decirse sin lugar a dudas, que las empresas grandes que hoy en día se ven, no hubieran podido llegar a donde están sin el Almacén de Datos porque es tanta la información que deben manejar (y no simplemente manejar, sino manejar bien) que el análisis, sin esta herramienta, sería sencillamente imposible.

Los Almacenes de Datos han incursionado en varios ámbitos industriales, también en la especialidad de la medicina donde la calidad de los datos en todos sus componentes debe ser alta.

Para cualquier empresa tanto a nivel nacional como internacional es necesario poder realizar una buena toma de decisión, sin importar su naturaleza es imprescindible conocer, comprender, analizar un problema, para así poder darle solución en algunos casos por ser tan simples y cotidianos, este proceso se realiza de forma implícita y rápida.

Existen otros casos en los cuales las consecuencias de una mala o buena elección puede tener repercusión en la vida y si es en un contexto laboral, en el éxito o fracaso de la organización, es necesario realizar un proceso más estructurado que permita dar seguridad e información y resolver el problema.

Las Tecnologías de la Información y de la Comunicación (TIC) son un soporte para la toma de decisiones debido a que la evolución de estas, permite que la generación de la información sea oportuna. Cuba sostiene que el uso de las TIC no es neutral, responde siempre a los intereses de quienes la poseen y la aplican.

Esta es una de las explicaciones de por qué la extensión de las TIC por el mundo con un enorme potencial de beneficio, contribuye a acentuar la brecha digital socio económica entre ricos y pobres, entre poseedores y desposeídos, entre explotadores y explotados. Cuba ha defendido siempre el concepto de que el uso masivo de las TIC no es un fin sino una herramienta poderosa para lograr el desarrollo.

En Cuba existen sistemas informáticos especializados en el control y gestión de almacenes, pero no se adaptan a las particularidades de todas las empresas o estas se ven imposibilitadas a adquirirlos por no contar con presupuesto para ello. En este contexto el departamento de Almacenes de Datos del centro productivo de la Facultad Regional Mártires de Artemisa perteneciente a la infraestructura productiva de la Universidad de las Ciencias Informáticas (UCI), se dedica al

desarrollo de sistemas informáticos, encaminados a satisfacer las necesidades de gestión de la información existentes en una empresa, una institución o el país en general.

Con el fin de establecer un registro del control y el comportamiento de todos los niveles de información pertenecientes a la rama de la Salud dentro de los distintos municipios de la provincia Artemisa y para garantizar el proceso de toma de decisiones con una mayor validez, el centro productivo de la Facultad Regional Mártires de Artemisa ha elegido la línea de Almacenes de Datos para el desarrollo de dicho sistema informático.

Una vez aprobados todos los documentos y controlado su cumplimiento, se procede al trazo de las estrategias planteadas por dicha dirección, en pos de mejorar el proceso de toma de decisiones.

Donde se hacen evidentes los problemas que se describen a continuación: No existe un estándar específico para el almacenamiento de los datos, existen inconsistencias en los datos que se manejan, ya que no están organizados de la forma más óptima, el análisis de la información resulta engorroso cuando se trata de grandes volúmenes, la seguridad de los datos se encuentra comprometida, no están definidos niveles de accesibilidad en la información que se maneja, los datos no están integrados, ya que existen referencias a la misma información usando diferente codificación.

Esto dificulta a la entidad el análisis y accesibilidad de los volúmenes de información generados durante sus procesos de trabajo con la agilidad y rapidez necesaria, poniendo en riesgo la seguridad de los modelos informativos que se manejan y de esta manera ralentizando el proceso de la toma de decisiones.

A partir de la problemática planteada se genera el siguiente **Problema de investigación**: ¿Cómo contribuir a la toma de decisiones en la Dirección General de Salud de la Administración Provincial (AP) de Artemisa?

El problema de la investigación se enmarca en el **Objeto de Estudio**: Almacenes de Datos.

El Objeto de Estudio está delimitado por el **Campo de Acción**: Mercado de Datos para el apoyo a la toma de decisiones.

En función de dar solución al problema planteado se traza como **Objetivo General**: Desarrollar un Mercado de Datos para mejorar la toma de decisiones en la Dirección General de Salud de la AP de Artemisa.

Objetivos Específicos

- Investigar acerca del estado del arte, de las metodologías, herramientas y tecnologías a utilizar en el desarrollo del Mercado de Datos Salud de la AP.
- Realizar el análisis y diseño del Mercado de Datos Salud de la AP.
- Implementar el Mercado de Datos Salud de la AP.
- Validar el Mercado de Datos Salud de la AP.

Para desarrollar el presente trabajo se ha planteado la siguiente **idea a defender**:

Con el desarrollo de un Mercado de Datos para la Dirección Salud de la Administración Provincial de Artemisa se contribuye a mejorar la toma de decisiones.

Las Tareas de la Investigación trazadas para la realización del presente trabajo son:

- Definición del marco teórico de la investigación y elaboración del estado del arte.
- Diagnóstico de la situación existente en la Dirección de Salud con relación a estado del proceso de toma de decisiones y levantamiento de requisitos.
- Implementación del modelo de datos y subsistemas de integración y visualización a partir de la definición de la arquitectura del mercado de datos y diseño de los subsistemas de integración y visualización.
- Validación por medio de la aplicación de los casos de prueba.

Con el fin de lograr mayor éxito en la elaboración del Almacén de Datos Salud, se emplearon los siguientes Métodos Científicos de la Investigación:

Métodos Teóricos:

- **Análisis-síntesis:** Se utilizaron para el estudio y la determinación de las posiciones teóricas existentes sobre el Mercado de Datos, así como para la descripción y caracterización de los procesos observados; para realizar el análisis y la explicación de las distintas concepciones sobre el desarrollo del Mercado de Datos Salud.
- **Histórico y lógico:** Se utilizó para conocer, con mayor profundidad, los antecedentes del objeto que se investiga y las tendencias actuales, nacionales e internacionales.
- **Inducción-deducción:** Se utilizó para establecer conclusiones y generalizaciones sobre los procesos inherentes al objeto de estudio.
- **Enfoque de sistema:** Para la elaboración e instrumentación práctica de la estrategia pedagógica a fin de establecer las relaciones de dependencia, jerarquización y la estructuración de los componentes y los contenidos en el proceso de la dirección de salud de la Administración Provincial de Artemisa.

Métodos Empíricos:

- **La observación:** Se utilizó para valorar la actuación de los usuarios respecto al objeto de estudio con relación desempeño de su trabajo en las condiciones previas a la realización del presente trabajo y luego de aplicada la propuesta.
- **La encuesta:** Se utilizó para conocer los criterios de los usuarios antes y después de aplicada la propuesta y verificar su grado de identificación con la herramienta, así como las dificultades que estos presentan en la aplicación del mismo.
- **Análisis documental:** Posibilitó la consulta de materiales y normativas para la realización del presente trabajo.

Método Estadístico Matemático

➤ **Método de análisis porcentual:** Que no es más que comprobar los resultados reales de lo que se analiza, con la muestra.

Se cuenta con una población de 30 trabajadores de la Dirección Salud de la AP de Artemisa, interactuarán con el sistema una muestra de 12 personas, representando la misma un 40% de la población actual.

Como **variable independiente** de la investigación se identificó el Mercado de Datos Salud y como **variable dependiente** la toma de decisiones en la Dirección General de Salud de la AP de Artemisa.

Actualidad y necesidad:

Este trabajo pone a disposición de los usuarios una aplicación que les ayudará en el proceso de la toma de decisiones, con el objetivo de poder en un futuro manejar de forma más factible las informaciones útiles para el desarrollo de la dirección de Salud de la AP. Se definen para el desarrollo del Mercado de Datos Salud novedosas herramientas como Visual Paradigm 6.4, PgAdmin III 1.10, Pentaho Data Integration 4.0.1, Pentaho Schema Workbench 3.2.0, Pentaho BI server 3.6.0, Mondrian 3.0.4, Servidor Web Apache Tomcat, permitiendo las mismas el desarrollo completo del mercado de datos antes mencionado.

Aportes prácticos esperados del trabajo.

➤ Mercado de datos para la Dirección General de Salud.

Para una mejor comprensión del presente documento, la estructura del contenido queda conformada de la siguiente manera:

Capítulo 1. Fundamentos teóricos: En este capítulo se abordará toda la fundamentación teórica sobre los principales conceptos, metodologías y herramientas para el desarrollo de los Mercados de Datos que permitirá a la Dirección Salud apoyar el proceso de toma de decisiones.

Capítulo 2. Análisis y diseño del Mercado de Datos Salud: En este capítulo se realizará el refinamiento al análisis y diseño realizado en el trabajo precedente.

Capítulo 3. Implementación del Mercado de Datos Salud: En este capítulo se realizarán todas las transformaciones necesarias como Extraer, Transformar y Cargar (ETL), proceso que deja toda la información lista para la fase siguiente, Inteligencia del Negocio (BI); donde se realizará la implementación del modelo de datos, los cubos OLAP, los reportes candidatos y la política de seguridad de los usuarios.

Capítulo 4. Validación mediante casos de pruebas: En este capítulo se aplicarán los casos de prueba para validar el Mercado de Datos.

CAPÍTULO 1. FUNDAMENTOS TEÓRICOS DE LOS ALMACENES DE DATOS.

En el presente capítulo se describen los aspectos teóricos relacionados con la evolución de las tecnologías de almacén de datos, haciendo un estudio de las mismas a nivel nacional y mundial. Se presenta la metodología y herramientas a usar para el desarrollo de la solución propuesta. Además se exponen conceptos importantes que permiten una comprensión significativa de la investigación y justifican la solución escogida.

1.1 Fundamentos teóricos de los almacenes de datos.

En el mundo muchas empresas han tenido resultados factibles con el empleo de los Almacenes de Datos entre las cuales podemos encontrar empresas minoristas como TV Azteca, Wal-Mart, Visa, Telefónica de Argentina, Ipostel, GNP, Baxter, también existen compañías como American Stores, Canadian Tyre, BonPreu, Supermercados Casino en Francia, Otto Versand en Alemania, Helene Curtis en España, las cuales han obtenido avances económicos con la utilización de estos depósitos de datos.

En el área de salud existen importantes instituciones que han empleado los Almacenes de Datos, permitiendo estos agilizar el proceso de toma de decisiones y desarrollo de las mismas, entre ellas se encuentra el Almacén de Datos que fue implementado en Francia con el fin de controlar toda información relevante acerca de casi 50 millones de ciudadanos.

En 2001 el Departamento Nacional de Identificación de Malasia despliega a nivel masivo su sistema multipropósito de identificación por tarjetas inteligentes, las mismas obtienen información directa de Almacenes de Datos acerca de todo lo relacionado con la Salud.

En España en el año 2008 se realizó el diseño, construcción e implantación de un Almacén de Datos corporativo para las listas de espera quirúrgicas, consultas

Capítulo 1. Fundamentación teórica

externas y pruebas complementarias, en el ámbito de la atención especializada del Servicio de Salud.

Dos importantes empresas argentinas del sector salud eligieron a Neoris, uno de los principales proveedores mundiales de servicios, con el fin de efectuar un Depósito de Datos, para el tratamiento de fraudes y recargas en el área de la salud Argentina.

Estos Almacenes de Datos antes mencionados, poseen desventajas a la hora de determinar las funcionalidades que se pueden aprovechar y cuáles se deben implementar debido a la cantidad de información que se maneja, resultando costoso implementar operaciones no necesarias o dejar de implementar alguna que sí vaya a necesitarse. También emplean para su implementación herramientas de software privativo, los cuales poseen curso de aprendizaje muy costoso, una absoluta dependencia de los proveedores entre otras desventajas

En Cuba se ha comenzado a irrumpir en las tecnologías de los Almacenes de Datos, existen diferentes entidades que han utilizado estas obteniendo grandes logros entre ellas podemos encontrar: La Corporación Cimex, implementó un almacén que centra su atención en la actividad del comercio con el objetivo de apoyar a la toma de decisiones en dicha entidad. La Empresa de Proyectos de Arquitectura e Ingeniería (EMPAI) de Matanzas creada para dar respuesta a importantes inversiones del sector de la Construcción.

La UCI ha incursionado en el desarrollo de los Almacenes de Datos para importantes instituciones del país, como por ejemplo, la Oficina Nacional de Estadística, la Administración Provincial de Artemisa, el Ministerio del Comercio Exterior y la Inversión Extranjera (MINCEX).

El presente trabajo va dirigido a la Dirección de Salud de la Administración Provincial de Artemisa, realizando para la misma el análisis, diseño e implementación de un Mercado de Datos, con el objetivo de facilitar el proceso de la toma decisiones, para la ejecución del sistema se emplea herramientas

pertencientes a software libre brindando ventajas en el ámbito de la seguridad, privacidad, calidad, actualización y bajo costo, resultando sustentable para la economía del país con respecto al software privativo.

1.2 Almacenes de Datos (ADs).

Un ADs es un repositorio de datos de fácil acceso, alimentado de numerosas fuentes, transformadas en grupos de información sobre temas específicos de negocios, para permitir nuevas consultas y análisis. Es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes, para luego procesarla permitiendo su análisis desde diferentes perspectivas y con grandes velocidades de respuesta. La creación de un ADs representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implantar una solución completa y fiable de inteligencia de negocios (ISAITH WILLIAN, 2007).

Metas de los almacenes de datos.

Con la necesidad de que la información cada día sea de mayor utilidad para los directivos de las organizaciones, y con la complejidad de las soluciones empresariales, se obliga a la realización de sistemas más abiertos y dinámicos que permitan su adaptación ante los cambios que se imponen en la actualidad. Ante esta disyuntiva, con la maduración de las técnicas de desarrollo de ADs y Mercados de Datos (MDs) se han definido un conjunto de metas que deben cumplir estos sistemas (KIMBALL RALPH, 2002).

- Deben hacer fácilmente accesible la información.
- La información de la organización debe ser presentada de forma consistente.
- Deben ser adaptables y resistentes al cambio.
- Deben ser un baluarte seguro que apoye los recursos de información.
- Debe servir como base para mejorar la toma de decisiones.

Características de los almacenes de datos.

➤ Organizado en torno 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 (PADRÓN LIUDMINA, 2006), es decir, los datos están almacenados por diferentes materias o temas, las cuales se organizan teniendo como objetivo principal la perspectiva del usuario final.

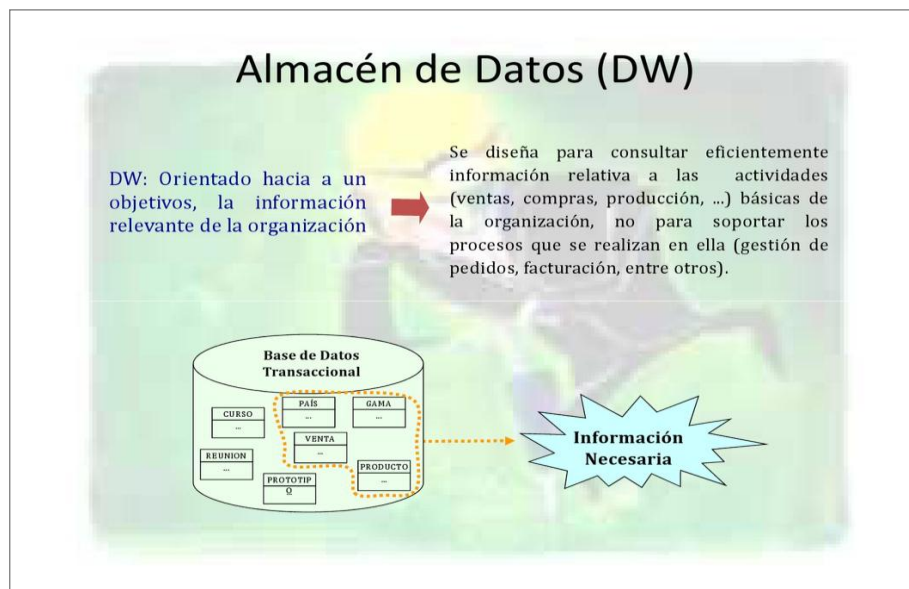


Figura 1. Organizado en torno a temas.

➤ Integrado

Se puede decir con certeza que la integración de datos es el aspecto más importante, ya que consiste en convenciones de nombres, codificaciones consistentes, medida uniforme de variables, etc. Un ejemplo claro de datos integrados se puede apreciar en la siguiente figura.

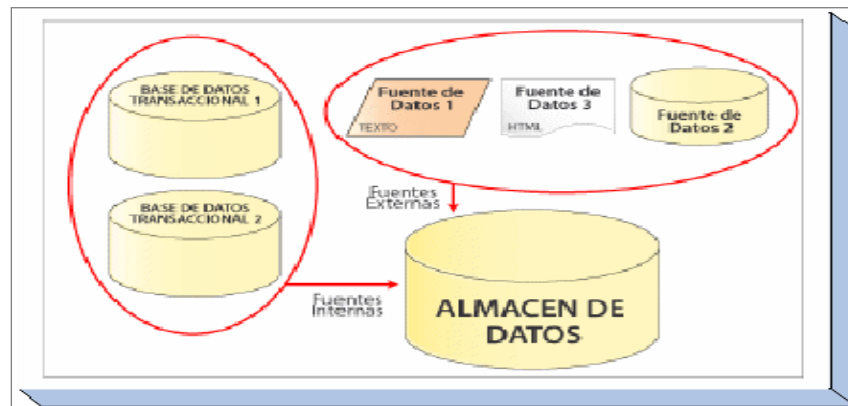


Figura 2. Integración de los Datos.

➤ **Variante en el tiempo**

Esta dependencia aparece de tres formas: La información representa los datos sobre un horizonte largo de tiempo, cada estructura clave contiene (implícita o explícitamente) un elemento de tiempo (día, semana, mes, etc.), la información, una vez registrada correctamente, no puede ser actualizada (HERNANDO ROBERTO, 2007).

Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones, ya que los datos son relativos a un período de tiempo y estos deben ser integrados periódicamente.

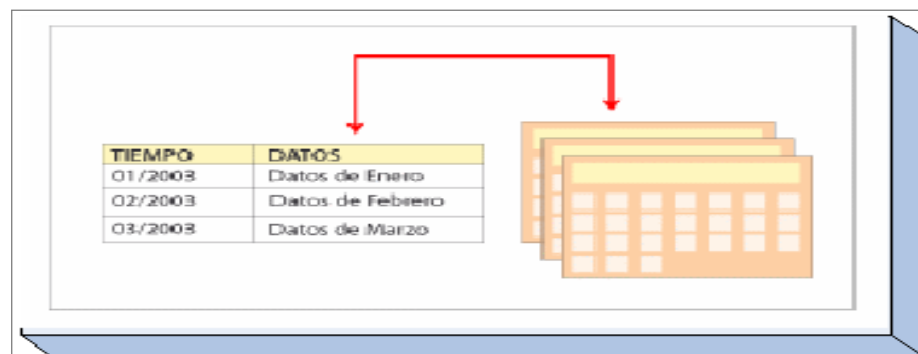


Figura 3. Variante en el tiempo.

➤ No volátil

Únicamente existen dos tipos de operaciones en el ADs: la carga de los datos procedentes de los entornos operacionales (carga inicial y carga periódica) y la consulta de los mismos. La actualización de datos no forma parte de la operativa normal de un ADs (PADRÓN LIUDMILA, 2006).

El Almacén de Datos solamente permite cargar nuevos datos y acceder a los ya almacenados, pero no permite ni borrar ni modificar los datos, los datos que son almacenados no sufren ninguna actualización solamente son incrementados.

La información es útil solamente cuando es estable. Los datos operacionales cambian sobre una base momento a momento. La perspectiva más grande, esencial para el análisis y la toma de decisiones, requiere una Base de Datos estable (HERRERA CRISTIAN, 2007).

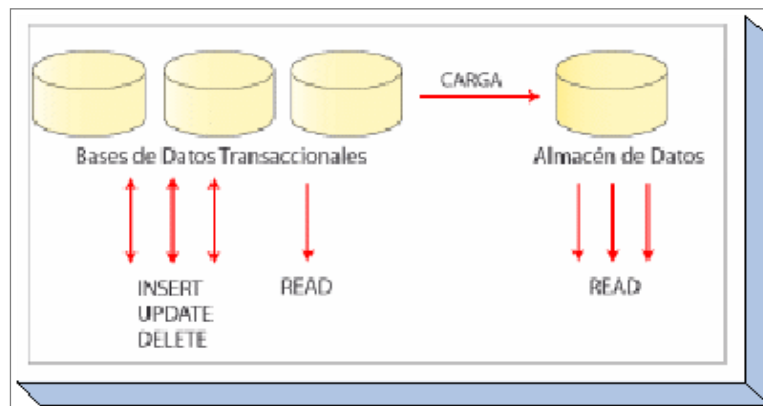


Figura 4. Información no volátil.

Principales ventajas y desventajas.

➤ Ventajas

1. Brindan rentabilidad en las inversiones realizadas para su creación.
2. Aumenta la competitividad en el mercado.
3. Aumenta la productividad de los técnicos de dirección.
4. Los ADs hacen más fácil el acceso a una gran variedad de datos a los usuarios finales

➤ **Desventajas**

1. Sub-valoración del esfuerzo necesario para su diseño y creación.
2. Sub-valoración de los recursos necesarios para la captura, carga y almacenamiento de datos.
3. Incremento continuo de los requisitos de los usuarios.
4. A lo largo de su vida los ADs pueden suponer altos costos. El AD no suele ser estático. Los costos de mantenimiento son elevados. (RODRIGUEZ PEDRO, 2009)

1.3 Mercado de Datos (MDs). Características y metodología.

Los MDs son un subconjunto de datos de un AD donde se almacenan la mayoría de las actividades de análisis que en el entorno de Inteligencia de Negocio se llevará a cabo.

La visión de Inmon se basa en un enfoque descendente, propone construir primero el AD, y a partir de este los MDs. Plantea la creación de un repositorio de datos corporativo como fuente de información consolidada, persistente, histórica y de calidad.

A diferencia de la anterior, la propuesta de Kimball se basa en dividir el mundo de Inteligencia de Negocio entre los hechos y las dimensiones, ésta es eficaz y conduce a una solución completa en un corto período de tiempo. Además, tiene abundante documentación y se puede encontrar una respuesta a casi todas las preguntas que se puedan tener. Entre sus características principales, está el hecho de poseer una arquitectura ascendente, plantea que se debe crear por cada departamento un conjunto de MDs independientes orientados a los temas que estén relacionados con él.

Metodología a utilizar en el Mercado de Datos Salud del AP.

La metodología que emplearon los autores para la implementación del Mercado de Datos Salud de la AP es la de Proceso de Desarrollo en la Línea Soluciones de Almacenes de Datos Inteligencia de Negocio la que posee las siguientes ventajas:

- La técnica posee una gran cantidad de documentación y generalmente se puede encontrar una respuesta a casi todas las problemáticas que puedan presentar.
- Esta metodología de dividir el mundo de BI entre el hecho y las dimensiones es muy eficaz y conduce a una solución completa en un tiempo razonable.
- Es iterativo, donde se construye una pieza a la vez (MDs) garantizando mayor velocidad de respuesta a los clientes.
- La forma de almacenar la información es de fácil entendimiento por parte del usuario lo que permite mayor comprensión para el análisis de los datos que se encuentran integrados.
- Es una metodología resistente y adaptable ante los cambios.

1.4 Modos de almacenamiento de datos.

La tecnología de Procesamiento Analítico en Línea OLAP (Online Analytical Processing) permite un uso más eficaz de los ADs para el análisis de datos en línea, proporciona respuestas rápidas a consultas analíticas complejas e iterativas utilizada generalmente para sistemas de ayuda para la toma de decisiones, presenta los datos a los usuarios a través de un modelo de datos intuitivo y natural. Permitiendo a los usuarios finales ver y entender la información de su Base de Datos. OLAP acelera la entrega de información a los usuarios finales que ven estas estructuras de datos como cubos, denominadas multidimensionales debido a que la información es vista en varias dimensiones. Las características principales del OLAP son:

- **Rápido:** Proporciona la información al usuario a una velocidad constante. La mayoría de las peticiones se deben de responder al usuario en cinco segundos o menos.
- **Análisis:** Realiza análisis estadísticos y numéricos básicos de los datos, predefinidos por el desarrollador de la aplicación o definido “ad hoc” por el usuario.

Capítulo 1. Fundamentación teórica

- **Compartida:** Implementa los requerimientos de seguridad necesarios para compartir datos potencialmente confidenciales a través de una gran población de usuarios.
- **Multidimensional:** Llena la característica esencial del OLAP, que es ver la información en determinadas vistas o dimensiones.
- **Información:** Acceden a todos los datos y a la información necesaria y relevante para la aplicación, donde sea que ésta resida y no esté limitada por el volumen.

Existen tres modelos para el proceso analítico en línea (OLAP) de la información: ROLAP, MOLAP y HOLAP. El proceso de análisis se realiza de igual forma lo que varía en uno y otro caso es la metodología de almacenamiento. La forma de almacenamiento es crítica para garantizar la velocidad de recuperación de la información, las zonas de ubicación de las agregaciones y el procesamiento de los datos en general (FALCÓN YOLANDA Y LEYVA REYNALDO, 2010).

Características			
	MOLAP	ROLAP	HOLAP
Almacenamiento de las Agregaciones	Modelo Multidimensional	Base de datos relacional	Modelo Multidimensional
Almacenamiento de los datos	Modelo Multidimensional	Base de datos relacional	Base de datos relacional
Facilidad de Creación	Sencillo	Muy Sencillo	Sencillo
Velocidad de respuesta	Buena	Regular o Baja	Buena para consultas que posean agregaciones, Regular para datos de bajo nivel
Escalabilidad	Problemas de escalabilidad	Son más escalables	
Recomendados para	Cubos con uso frecuente	Datos que no son frecuentemente usados	Si el cubo requiere una rápida respuesta

Figura 5. Comparación de las principales características de los tres modelos para el proceso analítico en línea.

En el presente trabajo se empleará Procesamiento Analítico Relacional en Línea (ROLAP).

1.5 Integración de datos. Características.

ETL - este término viene de inglés de las siglas Extract-Transform-Load que significan Extraer, Transformar y Cargar organizando el flujo de los datos entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un AD, limpiarlos y cargarlos en otra Base de Datos (HERNÁNDEZ ASNIOBY, 2009).

Etapas del proceso de integración de datos.

Debido a que los datos deberán ser extraídos, transformados, limpiados y cargados desde el conjunto de archivos DBF hacia el MD, es imprescindible conocer como se realizarán cada una de estas actividades.

Extracción: Obtención de la información de las distintas fuentes tanto internas como externas.

Transformación: Luego de realizarse el proceso de extracción los datos provenientes de las diferentes fuentes pueden ser incoherentes, tener errores o estar incompletos. Con esto se busca obtener datos lo más precisos, completos, consistentes, interpretables y accesibles. Después del proceso de limpieza se lleva a cabo la integración de los datos con el propósito de eliminar problemas de redundancia e identificar las fuentes de datos más fiables. Una vez realizado el proceso de extracción y limpieza se procede a transformar los datos para de esta forma estandarizar los códigos, corregir los datos, eliminar registros duplicados, usar conversiones y combinaciones para generar nuevos campos.

Carga: Organización y actualización de los datos y los metadatos en la Base de Datos.

Si no se realiza un correcto proceso de ETL se pudieran obtener datos incorrectos lo que afectaría el proceso de toma de decisiones, es por eso que este proceso constituye aproximadamente un 70% del trabajo de la construcción de un AD (ITATÍ PAOLA, 2010).

1.6 Inteligencia de negocios.

Es un factor estratégico dentro de una empresa u organización, genera una potencial ventaja competitiva, proporcionando información privilegiada que responden a los problemas de negocio: entrada a nuevos mercados, promociones u ofertas de productos, eliminación de islas de información, control financiero, optimización de costes, planificación de la producción, análisis de perfiles de clientes, y rentabilidad de un producto concreto.

Componentes de una solución de inteligencia de negocios.

Todas las soluciones de BI tienen funciones similares, pero deben reunir al menos los siguientes componentes:

Multidimensionalidad: La información multidimensional se puede encontrar en hojas de cálculo, Base de Datos, etc. Una herramienta de BI debe de ser capaz de reunir información dispersa en toda la empresa e incluso en diferentes fuentes para así proporcionar a los departamentos la accesibilidad, poder y flexibilidad que necesitan para analizar la información.

Minería de datos: Las aplicaciones de minería de datos pueden identificar tendencias y comportamientos, no solamente para extraer información, sino también para descubrir las relaciones en Base de Datos que pueden identificar comportamientos que no son muy evidentes.

Agentes: Los agentes son programas que "piensan". Ellos pueden realizar tareas a un nivel muy básico sin necesidad de intervención humana. Por ejemplo, un agente puede realizar tareas un poco complejas, como elaborar documentos, establecer diagramas de flujo.

AD es la respuesta de la tecnología de información a la descentralización en la toma de decisiones. Coloca información de todas las áreas funcionales de la organización en manos de quien toma las decisiones. También proporciona herramientas para búsqueda y análisis.

1.7 Modelo de datos.

Un modelo de datos es un conjunto de estructuras que describen los datos, sus relaciones, su significado y las condiciones que los datos deben cumplir para reflejar correctamente la realidad deseada.

Características (KIMALL RALPH, 2002)

- Es el proceso de analizar los aspectos de interés para una organización y la relación que tienen unos con otros.
- Resulta en el descubrimiento y documentación de los recursos de datos del negocio.
- El modelado hace la pregunta " ¿Qué? " en lugar de " ¿Cómo? ", ésta última orientada al procesamiento de los datos.
- Es una tarea difícil, pero es una actividad necesaria cuya habilidad solamente se adquiere con la experiencia.
- Registrar los requerimientos de datos de un proceso de negocio.
- Permite observar:
 - Patrones de datos.
 - Usos potenciales de los datos.

Modelo entidad-relación

Los diagramas o modelos entidad-relación (denominado por sus siglas, ERD "Diagram Entity relationship") son una herramienta para el modelado de datos de un sistema de información. Estos modelos expresan entidades relevantes para un sistema de información, sus inter-relaciones y propiedades. El ERD está basado en una percepción del mundo real que consta de un conjunto de objetos básicos llamados entidades con sus atributos y de las interrelaciones que existen entre estos objetos. Se desarrolló para facilitar el diseño de Base de Datos permitiendo la especificación de un esquema del universo de discurso que representa la

Capítulo 1. Fundamentación teórica

estructura completa de la misma. Los esquemas del modelo entidad-relación (MER) usan diagramas para representar la estructura natural de los datos, que se nombran diagrama entidad relación, entre estos se pueden encontrar los rectángulos que representan a las entidades, los rombos representan a las interrelaciones que son enlazadas con sus entidades.

Modelo dimensional

El modelado dimensional es una técnica de modelado de datos que permite la visualización de los mismos. Se utilizan para diseñar ADs con la particularidad de que estos van a estar compuestos por hechos, medidas y dimensiones.

Existen tres esquemas que se utilizan para modelar la estructura de un almacén de datos. Ellos son:

- **Esquema de estrella:** Es la arquitectura de almacén de datos más simple. En este diseño del almacén de datos, la tabla de Variables (Hechos) está rodeada por Dimensiones y juntos forman una estructura que permite implementar mecanismos básicos para poder utilizarla con una herramienta de consultas OLAP.

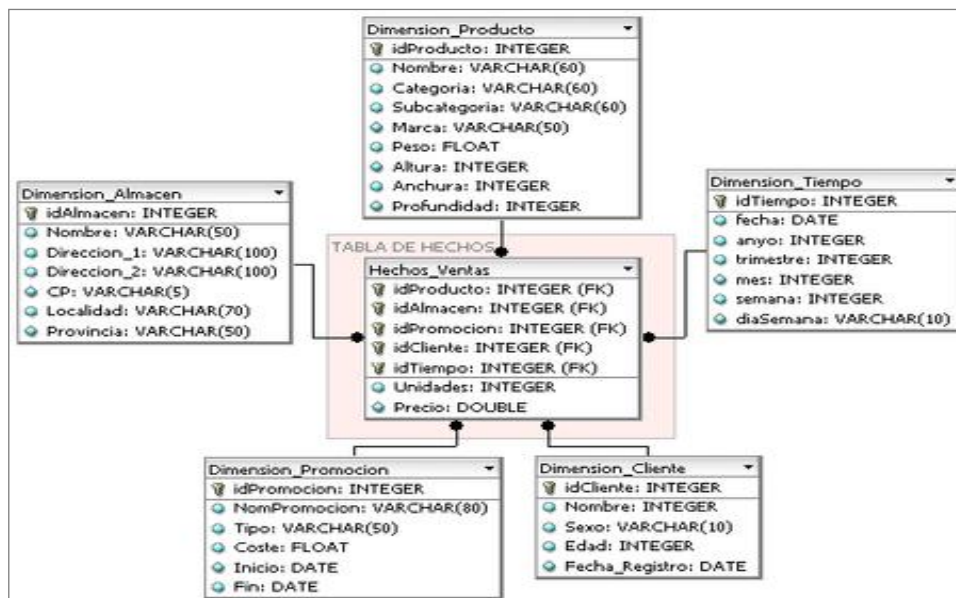


Figura 6. Esquema estrella.

- **Esquema de copo de nieve:** Es parecido al de estrella pero existen jerarquías en las dimensiones. Las tablas de dimensiones pueden estar relacionadas, o sea, existen caminos alternativos en ellas (VEGAS LILIAM, 2008).

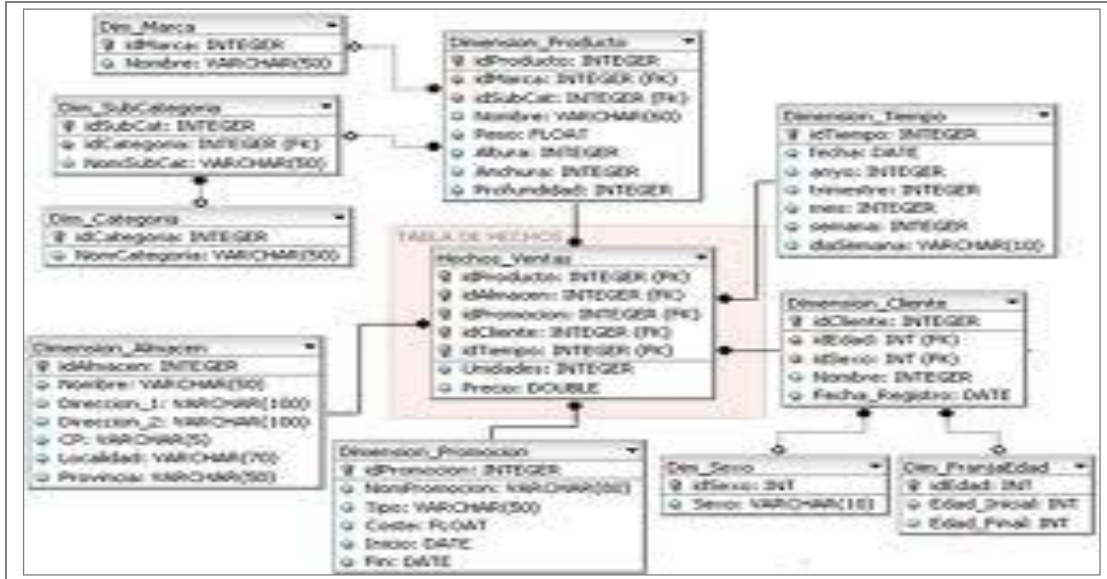


Figura 7. Esquema copo de nieve.

- **Constelación de hechos:** La constelación de hechos es un conjunto de tablas de hechos que comparten algunas tablas de dimensiones.
- **Tablas de hechos:** es la tabla central de un esquema dimensional (en estrella o en copo de nieve) y contiene los valores de las medidas de negocio. Cada medida se toma mediante la intersección de las dimensiones que la definen, dichas dimensiones estarán reflejadas en sus correspondientes tablas de dimensiones que rodearán la tabla de hechos y estarán relacionadas con ella.
- **Tablas de dimensiones:** Las dimensiones de un cubo son atributos relativos a las variables, son las perspectivas de análisis de las variables (forman parte de la tabla de dimensiones). Son catálogos de información complementaria necesaria para la presentación de los datos a los usuarios, como por ejemplo: descripciones, nombres, zonas, rangos de tiempo, etc. Es decir, la información general complementaria a cada uno de los registros de la tabla de hechos.

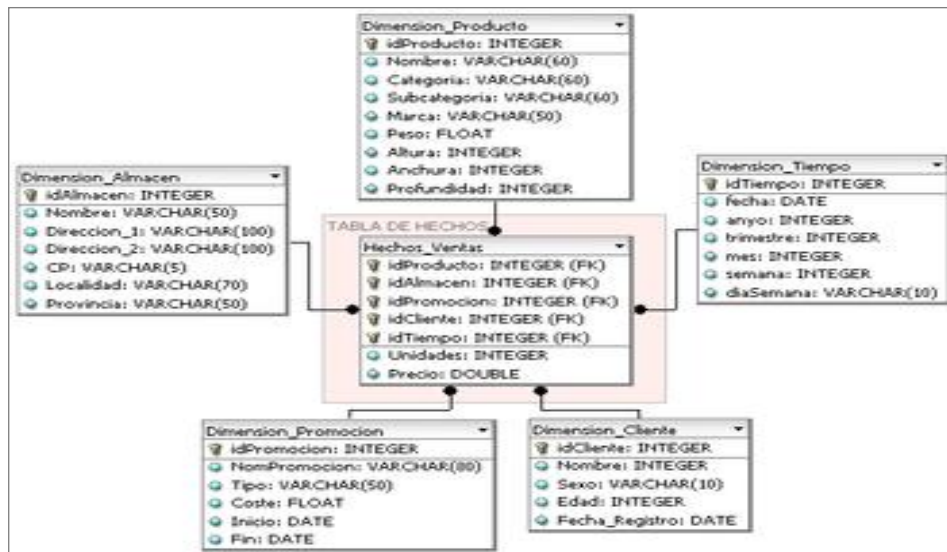


Figura 8. Tablas hechos y dimensiones de Base de Datos.

- **Medidas:** Una medida es un tipo de dato cuya información es usada por los analistas (usuarios) en sus consultas para medir el rendimiento del comportamiento de un proceso o un objeto del negocio. Las medidas candidatas son los datos numéricos, pero no cada atributo numérico es una medida candidata.
- **Indicadores:** Los indicadores son variables que pueden tomar un valor de una determinada unidad de medida y de un determinado tipo de datos.
- **Atributos:** Son criterios utilizados para analizar los indicadores. Se basan, en los datos de referencia de las tablas de dimensiones. En un cubo, los atributos son los ejes del mismo. Son campos o criterios de análisis, pertenecientes a tablas de dimensiones.
- **Jerarquía:** Una jerarquía representa una relación lógica entre dos o más atributos; si poseen una relación “padre-hijo”.

Tienen las siguientes características:

- ✓ Existen varias en un mismo cubo.
- ✓ Tienen dos o más niveles.

- ✓ Relación “1-n” o “padre-hijo” entre atributos consecutivos de un nivel superior y uno inferior.
- ✓ Se pueden identificar cuando existen relaciones “1-n” o “padre-hijo” entre los propios atributos de un cubo.
- **Granularidad:** La granularidad es el nivel de detalle en que se almacena la información. A mayor nivel de detalle, mayor posibilidad analítica. Los datos con granularidad fina (nivel de detalle) podrán ser resumidos hasta obtener una granularidad media o gruesa. No sucede lo mismo en sentido contrario.

1.8 Selección de topología, herramientas de desarrollo.

Topología usada para el desarrollo del presente Mercado de Datos.

La topología o esquema a utilizar para el MDs Salud de la Administración Provincial de Artemisa es la de Estrella, ya que en diseño del almacén de datos, las tablas (Hechos) estarán rodeada por tablas Dimensiones y juntos forman una estructura que permite implementar mecanismos básicos para poder utilizarla con una herramienta de consultas OLAP.

Justificación de las herramientas a utilizar.

En la actualidad se han desarrollado diversas herramientas con el fin de dar un acercamiento a la automatización del diseño, construcción, implementación y mantenimiento de los ADs.

Herramientas de modelado.

Las herramientas de modelado se emplean para la creación de modelos de sistemas que ya existen o que se desarrollarán. Ofrecen gran usabilidad y generación de código.

En la presente investigación se decidió utilizar **Visual Paradigm 6.4** para UML como herramienta de modelado, por ser una herramienta UML profesional que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. El software de modelado

Capítulo 1. Fundamentación teórica

UML ayuda a una más rápida construcción de aplicaciones de calidad a un menor coste. Permite representar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. Genera código para un gran número de lenguajes de programación entre los que se encuentra Java y permite integración con varias herramientas de Java. Además, brinda una versión libre para uso no comercial (SIERRA MARÍA, 2009).

También brinda características como:

- Generación de Base de Datos.
- Ingeniería inversa de Base de Datos desde sistemas gestores de Base de Datos (DBMS) existentes a diagramas de Entidad-Relación.
- Generador de informes para generación de documentación.
- Soporta aplicaciones web.
- Fácil de instalar y actualizar.
- Compatibilidad entre ediciones.

PostgreSQL Code Factory Le permite administrar consultas y scripts SQL por medio de utilidades como el repliegue de código, completado de código y resaltado de sintaxis, construcción visual de la consulta, ejecución de varias consultas simultáneas, ejecución de scripts desde archivos, ver y editar el resultado a través de filtros, habilidad de agrupar y ordenar y exportar datos a 14 formatos diferentes de archivos.

PgAdmin III, 1.10 Es una herramienta de código abierto para la administración de Base de Datos PostgreSQL y derivados (EnterpriseDB Postgres Plus Advanced Server y Greenplum Database). Incluye (REIGNART MARIANO, 2010):

- Interfaz administrativa gráfica.
- Herramienta de consulta SQL (con un EXPLAIN gráfico).
- Editor de código procedural.

- Agente de planificación SQL/shell/batch.
- Administración de Slony-I.

PgAdmin se diseña para responder a las necesidades de la mayoría de los usuarios, desde escribir simples consultas SQL hasta desarrollar Base de Datos complejas. La interfaz gráfica soporta todas las características de PostgreSQL y hace simple la administración. Está disponible en más de una docena de lenguajes y para varios sistemas operativos, incluyendo Microsoft Windows, Linux, FreeBSD, Mac OSX y Solaris. Soporta versiones de servidores 7.3 y superiores. Versiones anteriores a 7.3 deben usar el PgAdmin II.

Herramientas para la integración de datos.

Herramientas para la integración de datos. Las herramientas para la integración de datos son muy útiles para que el proceso de ETL concluya con los resultados esperados, su uso garantiza: ganancias en términos de tiempo y total fiabilidad de los datos.

Pentaho Data Integration 4.0.1 Es de formato abierto y de fácil lectura para los XML que recogen transformaciones, tareas programadas y un repositorio relacional de metadatos ETL.

- Es aplicable a diversos tipos de Base de Datos (SQL server, PostgreSQL, MySQL, Microsoft Access, etc.).
- Posee facilidad para la importación y exportación de datos de un formato a otro cualquiera.
- Su principal fortaleza es la posibilidad que brinda de ser extensible mediante pluggins. (ESPINOSA ROBERTO, 2010)

Pentaho Schema Workbench 3.2.0 Es una interfaz de diseño que permite crear y probar esquemas de cubos Mondrian OLAP visualmente. La Plataforma de BI de Pentaho incrusta el motor de consulta Mondrian, como parte de su arquitectura. Además, permite la ejecución de consultas MDX (CURTO JOSEP, 2008).

Pentaho BI server 3.6.0 La aplicación más conocida de la Plataforma de BI es la Pentaho BI Server que funciona como una red basada en sistema de gestión de informe, el servidor de integración de aplicaciones y un motor de flujo de trabajo ligero (secuencias de acción.) Está diseñado para integrarse fácilmente en cualquier proceso de negocio.

Pentaho Report Designer El Pentaho Report Designer es una herramienta independiente que forma parte de la unidad de reportes de Pentaho (Pentaho Reporting), que simplifica el proceso de generación de reportes, permitiendo a los diseñadores de reportes crear rápidamente informes sofisticados y ricos visualmente basados en el proyecto de reportes de Pentaho JFreeReport.

El diseñador de reportes ofrece un entorno gráfico familiar, con herramientas intuitivas y fáciles de utilizar, y una estructura de reporte bastante acertada y flexible para darle libertad al diseñador de generar reportes que se adapten totalmente a su gusto y necesidad.

Mondrian 3.0.4 En el plano Open Source la herramienta más significativa es Mondrian, la cual es una de las aplicaciones más importantes de la Suite Pentaho BI. Mondrian es un servidor OLAP open source que gestiona la comunicación entre una aplicación OLAP y la Base de Datos con los datos fuente. Es desarrollado en Java/Servlets/JSPs que permite ser instalado en servidores de aplicaciones como JBoss. Entre sus principales características se encuentra la facilidad para el análisis de grandes volúmenes de información que se encuentren almacenados en Base de Datos que soporten JDBC. Mondrian soporta el lenguaje Microsoft's Multidimensional Expressions (MDX). También soporta los APIs: Java OLAP (JOLAP) y XML for Analysis application Programming.

Servidor Web Apache Tomcat Tomcat es un servidor web con soporte de Servlets y JSPs. Es usado como servidor web autónomo en entornos con alto nivel de tráfico y alta disponibilidad. Dado que el mismo fue escrito en Java, funciona en cualquier sistema operativo que disponga de la máquina virtual Java. En él los usuarios disponen de libre acceso a su código fuente y a su forma binaria en los

Capítulo 1. Fundamentación teórica

términos establecidos en la Apache Software Licence. La primera distribución de Tomcat fue la versión 3.0. La versión más reciente es la 7.0.

Conclusiones

En el presente capítulo se realizó un estudio basado en los principales conceptos relacionados con los Almacenes de Datos, la metodología de desarrollo adoptada es la de Kimball, el modo de almacenamiento propuesto es ROLAP. Las herramientas a utilizar son: Pentaho Schema Workbench 3.2.0, Pentaho Analysis server (Mondrian) 3.0.4, Pentaho BI Server, Apache Tomcat 6.0, Pentaho Data Integrator 4.0.1, DataCleaner 1.5.3, PostgreSQL 8.4, PgAdmin III, y Visual Paradigm 6.4.

CAPÍTULO 2. ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS SALUD.

En el presente capítulo se abordan aspectos concernientes a definición de las áreas de análisis, necesidades de los usuarios, reglas del negocio, requisitos de software y de los casos de uso del sistema. De manera general, se refiere al resultado del análisis y el diseño del MD para el Departamento de Salud del CAP de Artemisa.

2.1 Análisis y definición de negocios.

El análisis es un conjunto de reglas clasificadas dispuestas de manera ordenada, mostrando un plan lógico en la unión de las partes, con el fin de crear las bases para levantar una estructura sólida.

Definición del negocio.

La Dirección de Salud tiene el control de toda la información con carácter relevante que se maneja en cada uno de los departamentos que componen el mismo. La información estadística se encuentra almacenadas en los modelos que se trabajan, los cuales se llenan por especialistas de cada departamento y posteriormente revisados por el director de cada departamento para evitar la mayor cantidad de errores posibles.

Temas de análisis.

Los temas de análisis a desarrollar en la solución son:

- Análisis de los indicadores en el departamento de Hospitales.
- Análisis de los indicadores en el departamento de Medicamentos del sector Drogas.
- Análisis de los indicadores en el departamento de Medicamentos del sector Farmacia.

- Análisis de los indicadores en el departamento de Medicamentos del sector Fármaco-Epidemiología.
- Análisis de los indicadores en el departamento de Medicamentos del sector GAP.
- Análisis de los indicadores en el departamento de PAMI.
- Análisis de los indicadores en el departamento de Servicios Ambulatorios.

Reglas del negocio.

Las reglas del negocio se identificaron en el levantamiento de información y en el análisis de las fuentes. Dichas reglas son una entrada fundamental para los procesos de diseño del almacén, ETL y BI.

- Para los campos en blanco, se le introduce el valor 0.
- Los campos predefinidos se comprueban con los nomencladores.
- No debe permitirse cargar información de una fecha anterior cuando se carga una actual.
- No debe permitirse cargar información de la misma fecha una vez que esta haya sido cargada.
- Una vez cargados los datos en el almacén, no pueden existir campos nulos.

Necesidades de los usuarios.

Se definen como necesidades de los usuarios el análisis de los indicadores del departamento de Salud de la AP de Artemisa, tanto del año en curso como de años anteriores. Los datos se brindarán en un período de tiempo variable: mensuales o anuales.

Para ver los aspectos se tuvieron en cuenta para el levantamiento de información desarrollado durante la entrevista ver [Anexo 1](#).

Especificación de los Requisitos.

Requisitos Informativos.

Los requisitos informativos son la cantidad de reportes internos de cada sistema. En el mercado de datos para la Dirección General de Salud del CAP de Artemisa se detectaron 53 requisitos los cuales se dividieron por temas de análisis, a continuación como ejemplos se presentan 2 requisitos informativos:

- Obtener del modelo Plan Quirúrgico: el plan total de operaciones, el plan de operaciones mayores, el plan de operaciones menores, el plan de operaciones electivas, el plan de operaciones ambulatorias, el plan de operaciones de ingresados, el plan de operaciones urgentes, el plan de operaciones de mínimo acceso, el plan de operaciones menores electivas, el plan de operaciones menores urgentes por hospitales y por año.
- Obtener del modelo Plan de Operaciones: el total de operaciones realizadas y la diferencia entre estas y el plan de cumplimiento por cada uno de los hospitales y por año.

Para conocer todos se puede hacer referencia al artefacto Especificación de Requisitos.

Requisitos Funcionales.

RF1. Autenticar usuario: Permite el acceso a la aplicación y definir las actividades según el rol.

RF2. Adicionar usuario: Permitir adicionar usuarios al sistema.

RF3. Eliminar usuario: Permitir eliminar un usuario existente del sistema.

RF4. Asegurar la limpieza de los datos: Verificar que la información sea correcta.

RF5. Asegurar la transformación y carga de los datos desde distintas fuentes: Permitir realizar la transformación a los datos extraídos, para luego ser cargados en el mercado de datos.

RF6. Asegurar la extracción de los datos: Permitir realizar la extracción de los datos de los archivos fuentes.

RF7. Adicionar los reportes: Permitir adicionar nuevos reportes al sistema.

RF8. Eliminar reporte: Permitir eliminar un reporte existente en el sistema.

RF9. Adicionar rol: Permitir adicionar nuevos roles al sistema.

RF10. Eliminar rol: Permitir eliminar un rol existente en el sistema.

RF11. Configurar vistas de análisis: Permitir configurar vistas de análisis.

RF12. Asegurar la existencia de vistas materializadas para los reportes más comunes: Permitir configurar vistas materializadas.

Requisitos no funcionales.

Requisitos de Usabilidad.

Por las peculiaridades de los sistemas de almacenamiento y análisis de datos el usuario final debe tener conocimiento en el trabajo con las herramientas de integración de datos.

- Dominar conocimientos básicos sobre base de datos.
- Tener conocimiento en análisis y tratamiento de la información.
- Los usuarios deben recibir capacitación de las herramientas utilizadas en la confección de las Tablas de Salida y cómo interpretar la información de los mismos.
- Tener al menos una persona que domine el trabajo con la herramienta “Pentaho Data Integration” para realizar o modificar las transformaciones para la integración de datos.

Requisitos de Fiabilidad.

1. La integridad de los datos que se manejan es fundamental para el éxito de la solución de almacenamiento y análisis de datos.

Requisitos de Eficiencia.

1. La eficiencia del sistema constituye el elemento que aporta la calidad requerida a las funcionalidades solicitadas por los clientes.

Utilización de recursos:

1. El sistema debe permitir la abundancia de usuarios conectados simultáneamente sin que se afecte el tiempo de respuesta.

2. El tiempo de respuesta debe ser en tiempo real.

Capacidad:

1. En el proceso de integración solamente estará conectado un usuario que tendrá la tarea de vigilar el proceso de integración de datos.

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

1. Se dispondrá de una Guía de ayuda sobre componentes del sistema, presentadas en el sitio de la comunidad de Pentaho.

Componentes Comprados.

1. Las herramientas utilizadas está licenciada bajo GPL, y la misma es Open Source y totalmente gratis. Es ampliamente reutilizable y extensible.

Interfaz.

1. Los requerimientos de interfaz de usuario se centran en la presentación de la información de cara al cliente.

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

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

4. Los gráficos serán con los colores establecidos por la Dirección de Salud de la Administración Provincial de Artemisa ajustándose a los estándares establecidos de un buen diseño.

Interfaces de usuario.

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

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

3. Los gráficos serán con los colores establecidos por la dirección de Salud de la Administración Provincial de Artemisa ajustándose a los estándares establecidos de un buen diseño.

Interfaces Hardware.

1. En el proceso de integración es necesario la utilización de una memoria mínima de 2 GB para el proceso de transformación.

2. Se debe contar de un área de almacenamiento intermedio de 500 GB mínimos.

3. Para la visualización y la inteligencia de negocio se necesita una memoria de 2 GB.

4. Las estaciones de trabajo (pc clientes) deben contar con impresoras (para garantizar la impresión de las tablas de salida).

Interfaces Software.

1. Se debe disponer de la instalación de la herramienta BIServer.

2. Debe existir un navegador asociado al sistema operativo que se escoja para lograr que las interfaces web de las tablas de salida puedan visualizarse.

3. El perfilado de datos se realizará con el JavaScript, Pentaho, herramienta libre especializada en estas funcionalidades.

Interfaces de Comunicación.

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

Requisitos de Licencia.

1. Las herramientas a utilizar se encuentran bajo licencia GPL.

2.2 Diseño y propuesta del sistema.

El diseño es un proceso mediante el que se traducen los requisitos en una representación del software.

Propuesta del sistema.

Al concluir el análisis y de acuerdo a las necesidades de los usuarios se propone realizar un sistema que brinde las siguientes funciones:

- Lograr un análisis de la información con el propósito de la proyección de la información para la toma de decisiones.
- Donde se centralicen los datos de los sistemas fuentes, guardarlos durante períodos de tiempo extensos permitiendo el acceso a datos históricos por años lo que proporciona un nivel de análisis basado en experiencias.
- Proporcionar al usuario una interfaz consolidada única para los datos, que hace más fácil el trabajo con las consultas para la toma de decisiones.

Definición de los casos de uso.

Un caso de uso es una secuencia de transacciones que son desarrolladas por un sistema en respuesta a un evento que inicia un actor sobre el propio sistema.

Tabla 1. Actores del Sistema.

Actor	Descripción
Usuario	Analiza las actividades en el Departamento de de Salud y se autentica en el sistema.
Programador de ETL	Realiza los procesos de ETL.
Administrador	Administra los usuarios, rol y reportes del sistema.

Diagrama de Casos de Uso Informativos.

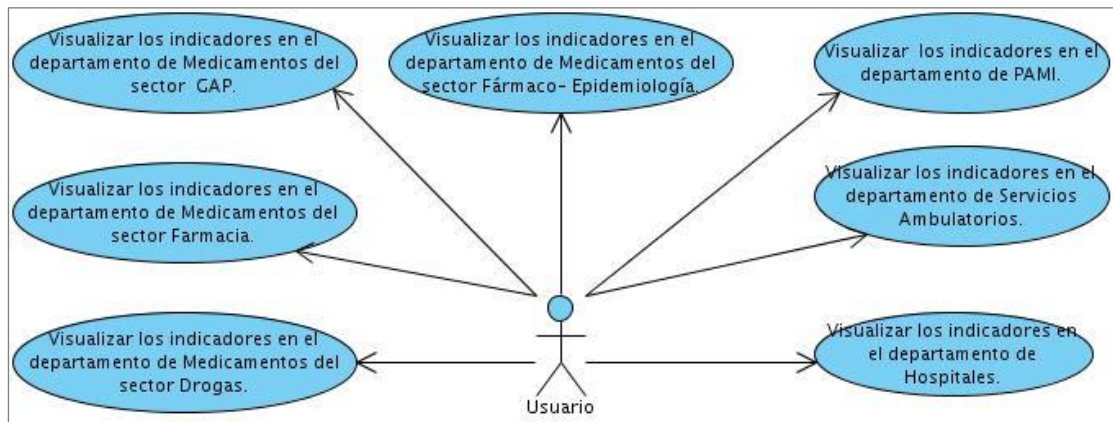
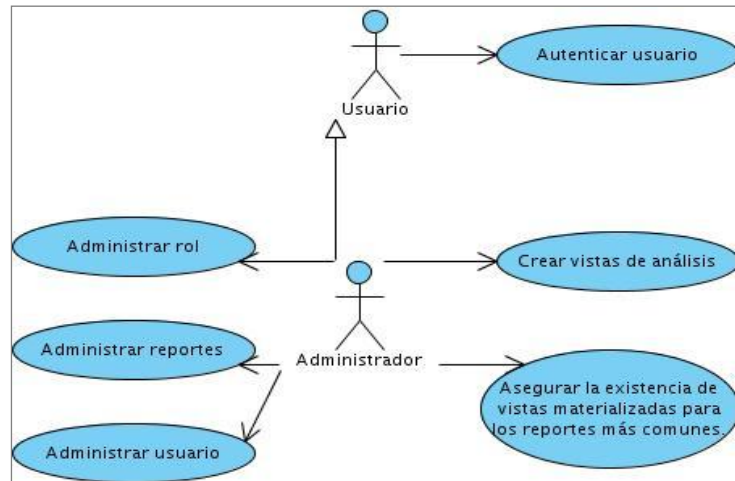


Figura 9. Casos de Uso Informativos.

Diagrama de Casos de Uso Funcionales.



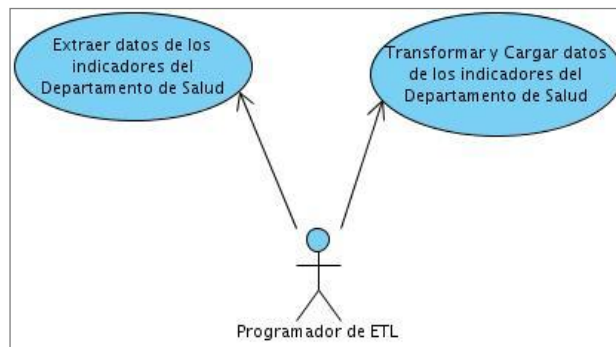


Figura 10. Casos de Uso Funcionales.

Especificación de los Casos de Uso.

Casos de Uso Informativos.

Visualizar los indicadores en el departamento de Medicamentos en el sector Farmacia.

Tabla 2. Descripción de Casos de Uso.

Caso de Uso:	Visualizar los indicadores en el departamento de Medicamentos en el sector Farmacia.	
Actores:	Rol o Grupo (Usuario)	
Resumen:		
Precondiciones:	Compleitud del almacén. Carga de los Datos.	
Referencias		
Prioridad		
Flujo Normal de Eventos		
Sección "Sector Farmacia"		
Acción del Actor	Respuesta del Sistema	
Entra al sistema.	Muestra opciones de reportes.	

Prototipo de Interfaz	
Flujos Alternos	
Acción del Actor	Respuesta del Sistema
Selecciona reportes del departamento de Medicamentos en el sector Farmacia.	Muestra tipos de reportes del departamento de Medicamentos en el sector Farmacia.
Prototipo de Interfaz	
Poscondiciones	Disponibilidad de opciones (cruces de variables) de reportes relacionados con el sector de Farmacia.

Continuación de las tablas de descripción de los casos de uso dirigirse al artefacto Especificación de Casos de Uso.

2.3 Dimensiones y tablas de hechos.

Dimensiones.

Las dimensiones poseen algunas características que prevalecen, entre ellas, por ejemplo la definición de jerarquías entre sus atributos, con el objetivo de plasmar explícitamente la forma en que se puede consolidar.

Tabla 3. Dimensión dim_hospital.

Nombre del atributo	Descripción	Tipo de Dato	Política de actualización
hospital_nombre	Nombre del hospital.	varchar(30)	anual
hospital_codigo	Código del hospital.	integer	anual

Continuación de las tablas de dimensión dirigirse al artefacto Especificación de Modelo Dimensional.

Tablas de hechos.

Las tablas de hechos son las que almacenan las medidas numéricas.

Tabla 4. Tabla de hechos.

Tablas de hechos	Medidas
hech_plan_quirurgico_anual	total_operaciones
	operaciones_mayores
	operaciones_electivas
	operaciones_ambulatorias
	operacion_ingresados
	operaciones_urgentes
	operacion_min_acceso
	operaciones_menores
	operacion_menor_electiva
	operacion_menor_urgente

Continuación de las tablas de hechos dirigirse al artefacto Especificación de Modelo Dimensional.

Matriz bus o matriz dimensional.

La matriz bus, es la relación que existe entre las dimensiones y los hechos del MD. Se define como la habilidad para describir y seguir la vida tanto de una dimensión

como de un hecho, la cual permite determinar el impacto que provocaría un cambio durante el desarrollo del sistema.

Tabla 5. Matriz Bus.

Hechos	Dimensiones									
	D1	D2	D3	D13	D20	D5	D6	D7	D8	D9
hech_plan_quirurgico_anual	x	x								
hech_plan_operaciones	x		x							
hech_act_quirurgica_preliminar	x		x							
hech_actividad_quirurgica_mensual	x		x							

Continuación de la Matriz Dimensional dirigirse al artefacto Especificación de Modelo Dimensional.

2.3 Modelo de datos.

El modelo de datos está conformado por las dimensiones, medidas, y hechos identificados para el departamento de Salud de la AP de Artemisa.

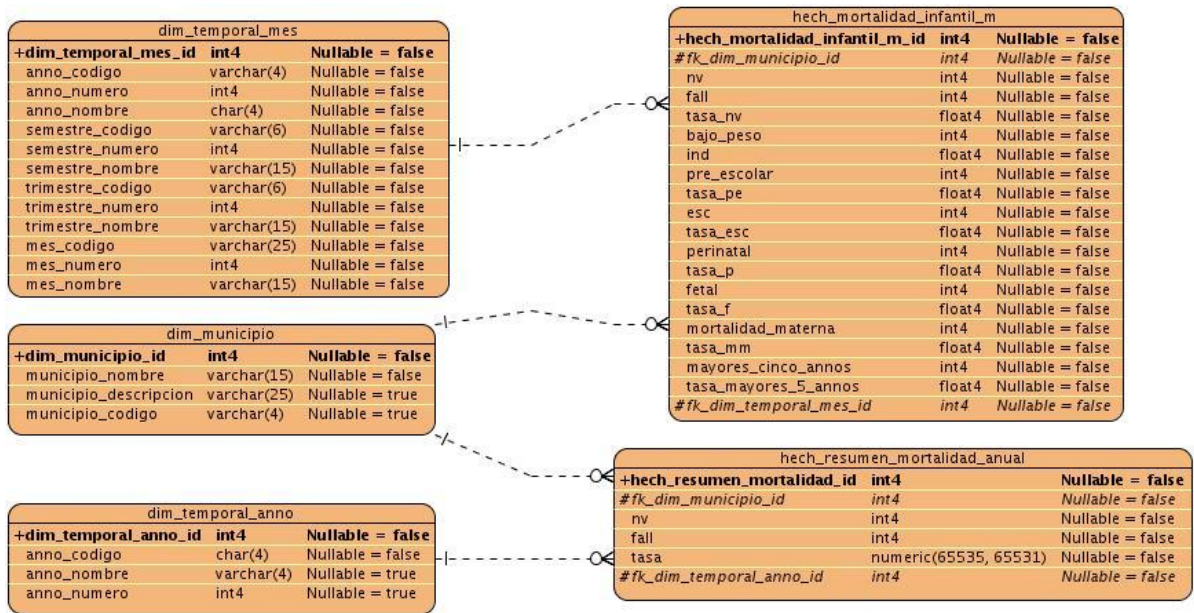


Figura 11. Tablas de hechos del departamento de Programa de Atención Materno Infantil.

Continuación del Modelo de Datos dirigirse al artefacto Especificación de Modelo Dimensional.

2.4 Esquema de seguridad.

La seguridad para el MD Salud, se rige fundamentalmente por los roles y permisos que los usuarios poseen en su interacción con la base de datos y la aplicación.

Para la interacción de los usuarios con la base de datos se definió el rol de:

Tabla 6. Seguridad en la base de datos.

Actor	Permisos
Administrador	Tiene total acceso a la base de dato del sistema.

Tabla 7. Seguridad en la aplicación.

Roles	Permisos
Administrador	Gestiona el sistema de información para el MD de Salud.
Usuario	Tiene acceso de solamente lectura al MD de Salud según el rol que le corresponda. Visualiza los reportes.

Conclusiones

En el presente capítulo quedaron explícitas las reglas del negocio, así como los requisitos según las necesidades de los usuarios. Se diseñó el modelo de caso de uso del sistema donde se representa la relación de los casos de usos y los actores de la aplicación. Se definió el modelo de datos en el cual se muestra la interacción de los hechos y las dimensiones identificadas.

CAPÍTULO 3. IMPLEMENTACIÓN DEL MERCADO DE DATOS SALUD.

En el presente capítulo se realizarán todas las transformaciones del MD Salud perteneciente a la AP de Artemisa, proceso que deja toda la información lista para la siguiente fase Inteligencia de Negocios, en la cual se realizará la implementación del modelo de datos, los cubos OLAP, los reportes y la política de seguridad de los usuarios, dándole solución a los requisitos del sistema.

3.1 Implementación de la base de datos.

El modelado de datos es uno de los elementos más importantes a la hora de iniciar el desarrollo de cualquier proyecto. La verdadera esencia de una aplicación reside en esta estructura. Cuando se diseña el modelo dimensional, el mismo se transforma a un modelo físico, del cual se genera el script de la base de datos, y es allí donde se evidencian las relaciones que existen entre las diferentes tablas, y a la vez determina si el proyecto va a cumplir con su verdadero objetivo.

Estructuras de los datos.

Para que el trabajo con los datos del MD sea de una manera óptima es necesario organizarlos de una forma correcta. Estas estructuras son denominadas: esquemas y tablas.

Esquemas.

Los esquemas en una base de datos, son una forma de organizar la información contenida en la misma. Los usuarios solamente tendrán acceso a aquellos que su rol se lo permita. Dentro de los esquemas se pueden encontrar funciones, operadores y tipos de datos que facilitarán su implementación.

En el presente trabajo se definieron dos esquemas:

Esquema dimensiones: contiene las tablas de las dimensiones generales del AD, y de ellas utilizamos las que se necesiten para implementar el MD.

Esquema mart_salud: contiene todas las tablas de hechos y las dimensiones

propias propuestas en el MD.

Tablas.

El diseño del modelo de datos, consta de 61 tablas en total, 17 dimensiones y 44 hechos. En la figura 8 se muestra utilizando la herramienta pgAdmin la estructura física del MD.



Figura 12. Almacén.

3.2 Usuarios y privilegios.

Con el objetivo de una mayor seguridad en la base de datos, es necesario definir los usuarios y roles, para con esto, poder realizar su función como trabajador del sistema.

Los roles establecidos son:

Programador de ETL: Su función se basa en la realización de los procesos de ETL en la interacción con la BD Privilegios:

Administrador: Se le es asignado los privilegios de Select, Insert, Update, Delete, Refresh y Trigger de los datos almacenados en el MD.

3.3 Implementación del subsistema de integración de datos. Arquitectura del subsistema de integración.

El desarrollo de un software no se puede iniciar hasta no tener bien definida una arquitectura. Esta no es más que un grupo de patrones que sirven de guía para la elaboración del mismo. Para la integración de los datos, la arquitectura queda de la siguiente forma:

- Fuente de Datos: Son archivos de extensión dbf o xls que contienen la información.
- Área temporal: Es el punto intermedio entre la fuente de datos y el MD. Es donde se realiza la integración y transformación de los datos.
- Mercado de Datos: Donde son cargados los datos para su futuro análisis.

Extracción de los Datos.

Se obtiene toda la información de las distintas fuentes tanto internas como externas. Se cargan los datos de los archivos dbf o excel, para el área temporal y así adaptarlos al modelo relacional que se ha establecido. Estos archivos contienen toda la información referente al MD Salud, la cual será almacenada en las tablas de hecho.

Transformación y Limpieza.

Una vez terminado el proceso de extracción, se realiza la limpieza de los datos provenientes de las diferentes fuentes, porque los mismos pueden ser incoherentes, tener errores o estar incompletos. Con esto se busca obtener datos precisos, completos, y lo más accesibles posibles. Después del proceso de limpieza se lleva a cabo la integración de los datos con el propósito de eliminar problemas de redundancia e identificar las fuentes de datos más fiables. La transformación y limpieza es de gran importancia porque en esta etapa es donde se garantiza el resultado final de cómo se van a mostrar los datos, se aplican las

reglas del negocio; y se detectan otras posibles deficiencias de la fuente y se corrigen.

Carga.

Es donde los datos son cargados al MD, organizados y actualizados, para que puedan ser usados por el cliente de forma satisfactoria. En el presente trabajo se elaboró un área temporal, donde van a ser llevados los datos, y una vez estando allí serán limpiados y transformados para posteriormente ser cargados al MD.

En la figura un ejemplo de las transformaciones implementadas que abarca todos los procesos de la integración de datos: la extracción, limpieza, transformación y carga de los mismos.

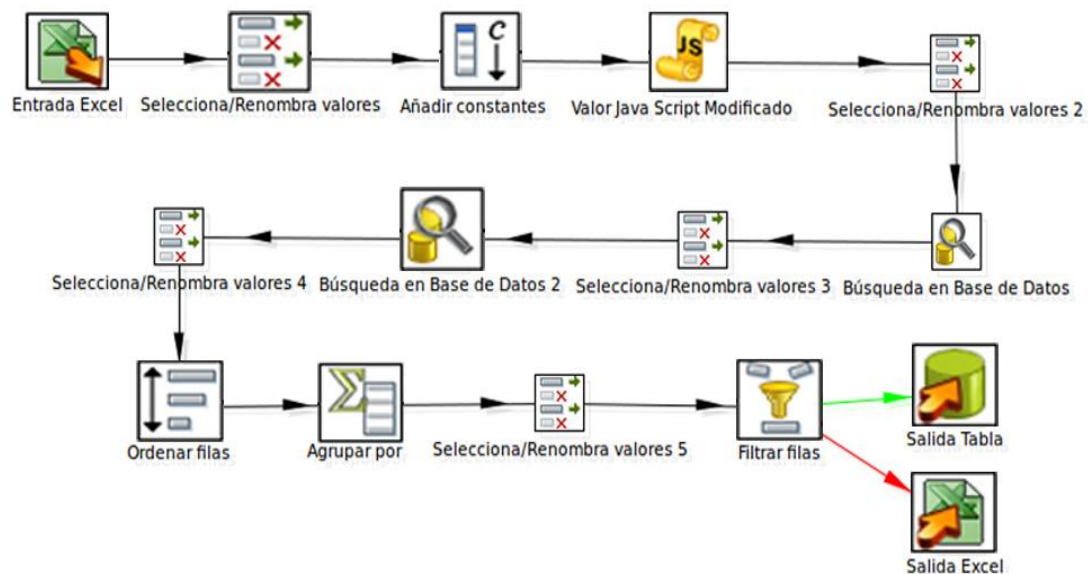


Figura 13. Transformación: Carga desde un Excel.

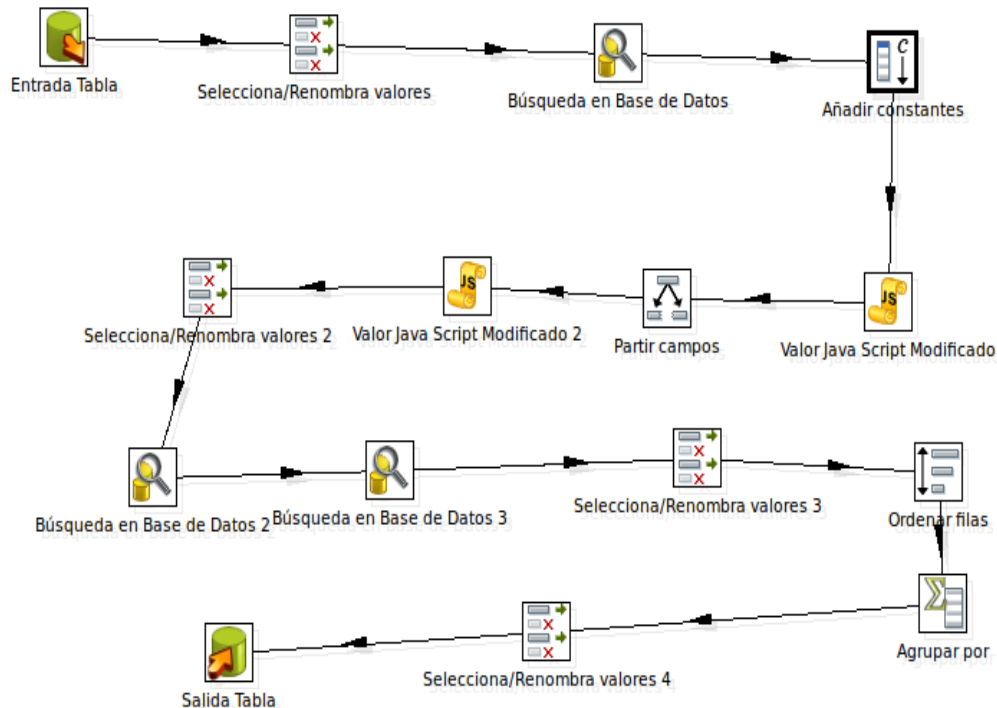


Figura 14. Transformación: Carga desde una base de datos.

Implementación de los trabajos.

Una vez que la conexión al MD se encuentra en perfecto estado, se procede a la carga del mismo, y el trabajo (job en inglés) es la forma en que se realiza la carga de los datos hacia el MD.

Para la correcta realización de un job, se debe tener bien definido cuáles son las dimensiones estáticas y cuáles no, pues en un job solamente se cargan las dimensiones que pueden tomar valores nuevos o cambiar los que tenían anteriormente.

En la siguiente figura se muestra un ejemplo de un trabajo que se realizó en el presente trabajo.

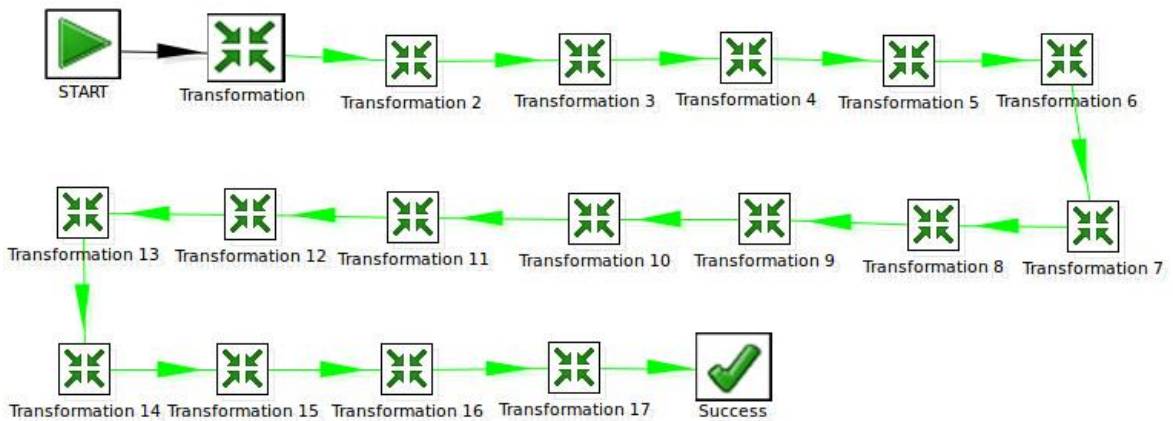


Figura 15. Trabajo general.

3.4 Implementación del subsistema de visualización de datos. Cubos OLAP.

La razón de usar OLAP para las consultas es la rapidez de respuesta y de poder agrupar los datos para garantizar un mejor análisis, obteniéndose los datos más importantes entre toda la información que posee el MD.

Para la implementación del módulo de reportes OLAP, es necesaria la creación de los cubos multidimensionales los cuales se realizan utilizando la herramienta Pentaho Schema Workbench, la misma permite generar un fichero de configuración XML. En este fichero de esquema se pueden definir las dimensiones, los niveles de jerarquía de dimensiones, los hechos y conexión con el almacén que sirve los datos para el cubo OLAP.

Se modelaron 41 cubos multidimensionales, en los mismos se especificaron las dimensiones y las medidas. La siguiente imagen muestra el diseño utilizando la herramienta Pentaho Schema Workbench de unos de los cubos modelados para el departamento de Servicios Ambulatorios al modelo de Servicios y equipos rotos a su correspondiente tabla `hech_servicios_y_equipos_rotos`, el cual está formado por sus dimensiones, y medidas.

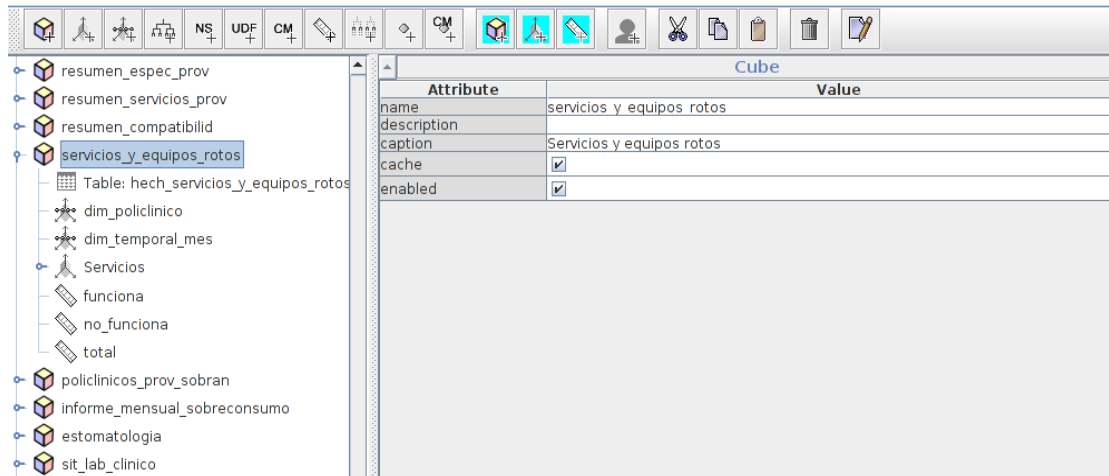


Figura. 16 Cubo OLAP.

Report Designer.

Con el objetivo de generar reportes que necesitan ser entregados para un posterior análisis se realizaron una serie de reportes en dicha herramienta, los cuales fueron efectuados a las tablas de hechos con su correspondiente vista materializada. Un ejemplo es el que se le realizó al modelo de Actividad quirúrgica mensual Otros indicadores, a continuación se muestra cómo quedaría después de su diseño.

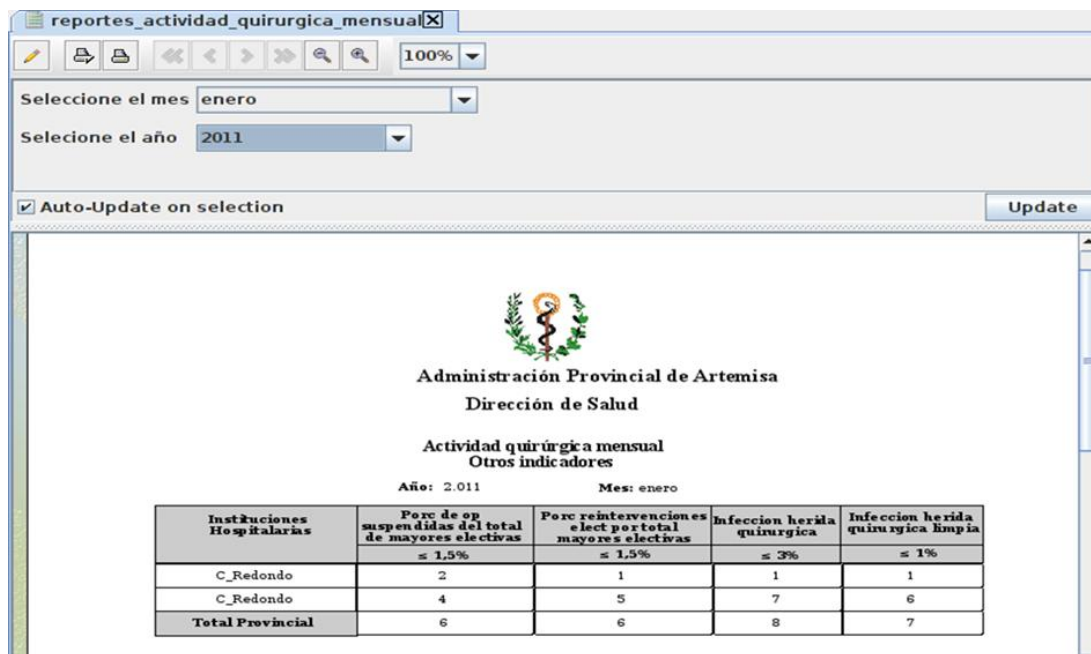


Figura. 17 Reportes.

Navegación de la capa de visualización.

El mapa de navegación es una representación gráfica donde se encuentra organizada toda la información. El MD para la Dirección de Salud tiene un siguiente mapa de navegación compuesto por 7 áreas de análisis, cada una con sus respectivos libros de trabajo y sus reportes.

Las 7 áreas de análisis son:

- Área de Análisis: Análisis de los indicadores del Departamento de Hospitales.
- Área de Análisis: Análisis de los indicadores del Departamento de Medicamentos sector Drogas.
- Área de Análisis: Análisis de los indicadores del Departamento de Medicamentos sector Farmacia.
- Área de Análisis: Análisis de los indicadores del Departamento de Medicamentos sector Fármaco – Epidemiología.
- Área de Análisis: Análisis de los indicadores del Departamento de Medicamentos sector GAP.
- Área de Análisis: Análisis de los indicadores del Departamento de PAMI.
- Área de Análisis: Análisis de los indicadores del Departamento de Servicios Ambulatorios.

Para ver la totalidad del mapa de navegación dirigirse al artefacto Arquitectura de Información del expediente del proyecto.

Cuando se accede al reporte deseado se pueden observar las vistas de análisis generadas, la siguiente figura muestra la vista de análisis correspondiente al cubo perteneciente al hech_servicios_y_equipos_rotos.

	Tiempo														
	2011														
	enero														
	Medidas														
	Funciona					No_funciona					Total				
	Servicio			Servicio		Servicio		Servicio			Servicio		Servicio		
Policlínico	EKG	US	RX	EKG	US	RX	EKG	US	RX	EKG	US	RX	EKG	US	RX
A_Sansaric	6,00	5,00	6,00	3,00	2,00	2,00	9,00	7,00	8,00	6,00	5,00	6,00	3,00	2,00	:

Figura. 18 Vista de análisis en el BI Server.

3.5 Configurar la seguridad de los usuarios.

El desarrollo de la implementación del subsistema de visualización del MD para la Dirección de Salud se crearon 6 usuarios y roles los cuales tienen diferentes permisos de acceso a la información, proporcionando una mayor seguridad al sistema los cuales son:

Administrador: tiene todos los permisos de la aplicación y posee el usuario administrador del sistema.

Director General: tiene permiso de solo lectura, posee el usuario Director_General del sistema y tiene acceso a todos los departamentos.

Director del Departamento de Medicamentos: tiene permiso de solo lectura, posee el usuario Director_del_Departamento_de_Medicamentos del sistema y tiene acceso a su departamento.

Director del Departamento de PAMI: tiene permiso de solo lectura, posee el usuario Director_del_Departamento_de_PAMI del sistema y tiene acceso a su departamento.

Director del Departamento de Hospitales: tiene permiso de solo lectura, posee el usuario Director_del_Departamento_de_Hospitales del sistema y tiene acceso a su departamento.

Director del Departamento de Servicios Ambulatorios: tiene permiso de solo lectura, posee el usuario `Director_del_Departamento_de_Servicios_Ambulatorios` del sistema y tiene acceso a su departamento.

Conclusiones

En el presente capítulo se obtuvo la estructura física del MD, se diseñó el esquema multidimensional con sus respectivos cubos OLAP para agrupar los datos y así facilitar su análisis posteriormente, se identificaron las áreas de análisis, los libros de trabajo y los reportes candidatos, se implementaron y visualizaron los reportes.

CAPÍTULO 4. VALIDACIÓN MEDIANTE CASOS DE PRUEBA.

En el presente capítulo se aplicarán los casos de prueba diseñados para comprobar el buen funcionamiento del mercado de datos, comprobándose que el resultado obtenido sea el esperado de acuerdo a las necesidades de los clientes de la Dirección de Salud de Artemisa.

4.1 Pruebas.

Las pruebas son un conjunto de actividades en las cuales un sistema o componente es ejecutado bajo unas condiciones o requerimientos específicos, donde los resultados son observados y registrados para dar una evaluación de algún aspecto del sistema o componente que es evaluado y determinar la calidad del mismo. En ellas se describen como comprobar cada versión operacional del sistema durante la integración del mismo, describiendo también como hacer las pruebas al sistema, verificando que todos los requerimientos hayan sido implementados, determinado los defectos del mismo.

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

4.2 Diseño de los casos de prueba.

El diseño de los caso de prueba se realiza para determinar si los requisitos o los caso de uso son completamente satisfactorios. Para el mercado de datos Salud se diseñaron 7 casos de prueba, uno por cada caso de uso, a continuación se muestra el diseño de caso de prueba correspondiente al caso de uso “Visualizar los indicadores en el departamento de PAMI (Programa de Atención Materno Infantil)”, para ver todos los caso de pruebas dirigirse a los artefactos Casos de Prueba desarrollados para cada caso de uso.

Capítulo 4. Validación mediante casos de pruebas

Tabla 8. Diseño del caso de prueba “Visualizar indicadores en el departamento de PAMI (Programa de Atención Materno Infantil)”.

Escenario	Descripción	Perfiles de análisis	Indicadores a medir	Respuesta del sistema	Flujo central
EC1.1 PAMI	Muestra los reportes de los indicadores de las PAMI.	Municipio Temporal_mes	nv fall tasa_nv bajo_peso ind pre_escolar tasa_pe esc tasa_esc perinatal tasa_p fetal tasa_f mortalidad_materna tasa_mm mayores_cinco_anos tasa_mayores_5_anos	El sistema muestra los reportes de los indicadores de las PAMI.	Se abre la aplicación. Se autentifica. Se entra al sistema. En el lateral izquierdo donde aparece el navegador se selecciona las áreas de análisis de A.A Análisis de los indicadores del Departamento PAM. Se selecciona el libro de trabajo de correspondiente a lo que se desea buscar. En la parte inferior izquierda se selecciona el reporte deseado. En el área de trabajo se visualiza la tabla correspondiente al reporte y la vistas correspondientes a esta área.

Estrategias de pruebas para el plan de pruebas.

Con el objetivo de comprobar la calidad del proceso de implementación, en el cual, se recogen precondiciones y medidas que facilitan la identificación de resultados

Capítulo 4. Validación mediante casos de pruebas

verídicos y eficientes se muestra a continuación algunos ejemplos de las pruebas establecidas para la capa de visualización:

Tabla 9. Prueba de acceso a la información por roles o niveles de usuario.

Descripción	Precondiciones	Medidas
Se verifica el acceso y privilegios que posee cada usuario o rol de la Dirección de Salud, definido con respecto al sistema.	Acceso restringido a la información, solo se puede ver la información necesaria del mercado Salud 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 o no cumple con los privilegios de acceso.

Capítulo 4. Validación mediante casos de pruebas

Tabla 10. Prueba de Consistencia de los Datos.

Descripción	Precondiciones	Medidas
Se inserta un número finito de datos a través de proceso de ETL en las tablas de cada hecho ya sea capital, presupuesto, etc. 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.

Capítulo 4. Validación mediante casos de pruebas

Tabla 11. 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 Salud.	Estructuras sólidas que respondan a las necesidades de información que se definieron en el negocio.	Eficiente: Las estructuras cumplen con el alcance de información previsto. Parcial: Las estructuras cumplen en un 80% con el alcance de información previsto. Deficiente: Las estructuras cumplen menos del 80% con el alcance de información previsto.

Conclusiones

En el presente capítulo con el objetivo de mejorar la organización y la ejecución de las pruebas se diseñaron y aplicaron 7 casos de pruebas para validar la calidad del mercado de datos Salud obteniendo resultados satisfactorios. También se estableció el plan de pruebas para verificar la calidad del proceso de implementación.

Conclusiones:

- ✓ Mediante la sistematización de la bibliografía accesible se pudo comprobar que existe suficiente fundamentación teórica relacionada con la tecnología de almacén de datos, se realizó un estudio de los sistemas de almacenes de datos a nivel nacional y mundial así como de las metodologías y herramientas para el desarrollo de los almacenes de datos.
- ✓ En el diagnóstico realizado se detectaron problemas en el proceso de toma de decisiones en la dirección de salud y lo que hizo factible la búsqueda de soluciones a partir de la determinación los requisitos funcionales para el desarrollo del almacén de datos.
- ✓ Durante la implementación del almacén de datos se identificaron las metodologías y herramientas que se ajustaban a los requerimientos del producto a elaborar teniendo en cuenta que estas fueran no privativas, las versiones más actualizadas a las que se puede acceder en el momento de la investigación y acorde a los requerimientos técnicos del lugar en que se implantará el sistema.
- ✓ Se validaron las transformaciones a partir de la ejecución correcta de las mismas, se mostraron las vistas y reportes en correspondencia con las necesidades del cliente por lo cual en los casos de uso resultó avalado el producto de forma satisfactoria.

Recomendaciones

Para el presente trabajo de diploma se recomienda:

- Adicionar nuevas formas de visualización de la información y aplicar otras técnicas de inteligencia de negocios.

Referencias

CURTO JOSEP. Mondrian y su ecosistema. Última actualización: 2008. Disponible en: <http://www.beyenetwork.es/view/8560>.

ESPINOSA ROBERTO. Pentaho Data Integration. Última actualización: 2010. Disponible en: <http://churriwifi.wordpress.com/2010/06/01/comparativa-talend-vs-kettle-pdi/>.

FALCÓN YOLANDA Y LEYVA REYNALDO. Mercado de Datos Estadístico de Inmigración y Extranjería para el Departamento de Turismo y Comercio de la Oficina Nacional de Estadísticas. Habana. UCI, 2010.

HERNÁNDEZ ASNIOBY. Documento de Arquitectura del Sistema Almacén de datos para la ONE. Cuba, Habana, 2009.

HERNANDO ROBERTO. Almacenes de datos (Datawarehouse). Última actualización: 2007. Disponible en: <http://www.rhernando.net/modules/tutorials/doc/bd/dw.html>.

HERRERA CRISTIAN. Todo lo que querías saber sobre DatawareHouse (I). Última actualización:

2007. Disponible en: <http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=datawarehouse#2.4.4.No%20vol%C3%A1til|outline>.

ISAITH WILLIAN. Datawarehouse. Última actualización: 2007. Disponible en: http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.

ITATÍ PAOLA. Tesis de Información. Última actualización: 2010. Disponible en: <https://docs.google.com/viewer?a=v&q=cache:dCGluOH-PIJ:200.45.54.90/depar/areas/informatica/SistemasOperativos/PaolaMonog.pdf>

KIMBALL RALPH. The Data Warehouse Lifecycle Toolkit. EUA : Wiley Publishing Inc, 2002.

PADRÓN LIUDMILA. Características de un almacén de datos. Última actualización: 2006. Disponible en: <http://www.mailxmail.com/curso-almacenes-datos-importancia-estandar/caracteristicas-almacen-datos>.

SIERRA MARÍA. UML Editor with BPMN, Visual Paradigm Última actualización: 2009. Disponible en: <http://www.visual-paradigm.com/product/?favor=vpuml>.

REIGNART MARIANO. PgAdmin Última actualización: 2010. Disponible en: <http://www.arpug.com.ar/trac/wiki/PgAdmin>.

RODRIGUEZ PEDRO. ORACLE - TecnoXML. Última actualización: 2009. Disponible en: <http://tecnoxml.wikidot.com/oracle>.

VEGAS LILIAM. La Inteligencia del negocio Última actualización: 2008. Disponible en: <https://docs.google.com/viewer?a=v&q=cache:yyTWOEAwnUJ:www.redciencia.info.ve/memorias/ProyProsp/trabajos/l3.doc>

Bibliografías

CURTO JOSEP. Mondrian y su ecosistema. Última actualización: 2008. Disponible en: <http://www.beyenetwork.es/view/8560>.

ESPINOSA ROBERTO. Pentaho Data Integration. Última actualización: 2010. Disponible en: <http://churriwifi.wordpress.com/2010/06/01/comparativa-talend-vs-kettle-pdi/>.

FALCÓN YOLANDA Y LEYVA REYNALDO. Mercado de Datos Estadístico de Inmigración y Extranjería para el Departamento de Turismo y Comercio de la Oficina Nacional de Estadísticas. Habana.UCI, 2010.

HERNÁNDEZ ASNIOBY. Documento de Arquitectura del Sistema Almacén de datos para la ONE. Cuba, Habana, 2009.

HERNANDO ROBERTO. Almacenes de datos (Datawarehouse). Última actualización: 2007. Disponible en: <http://www.rhernando.net/modules/tutorials/doc/bd/dw.html>.

HERRERA CRISTIAN. Todo lo que querías saber sobre DatawareHouse (I). Última actualización: 2007. Disponible en: <http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=datawarehouse#2.4.4.No%20vol%C3%A1til|outline>.

ISAITH WILLIAN. Datawarehouse. Última actualización: 2007. Disponible en: http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.

ITATÍ PAOLA. Tesis de Información. Última actualización: 2010. Disponible en: <https://docs.google.com/viewer?a=v&q=cache:dCGluOH-PIJ:200.45.54.90/depar/areas/informatica/SistemasOperativos/PaolaMonog.pdf>

KIMBALL RALPH. The Data Warehouse Lifecycle Toolkit. EUA : Wiley Publishing Inc, 2002.

PADRÓN LIUDMILA. Características de un almacén de datos. Última actualización: 2006. Disponible en: <http://www.mailxmail.com/curso-almacenes-datos-importancia-estandar/caracteristicas-almacen-datos>.

SIERRA MARÍA. UML Editor with BPMN, Visual Paradigm Última actualización: 2009. Disponible en: <http://www.visual-paradigm.com/product/?favor=vpuml>.

REIGNART MARIANO. PgAdmin Última actualización: 2010. Disponible en: <http://www.arpug.com.ar/trac/wiki/PgAdmin>.

RODRIGUEZ PEDRO. ORACLE - TecnoXML. Última actualización: 2009. Disponible en: <http://tecnoxml.wikidot.com/oracle>.

VALLEJOS SOFÍA "Minería de Datos", Trabajo de Adscripción para la Licenciatura en Sistemas de Información, Universidad Nacional del Nordeste Facultad de Ciencias 27.Exactas, Naturales y Agrimensura, 2006 Argentina. Disponible en: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/Mineria_Datos_Vallejos.pdf.

VEGAS LILIAM. La Inteligencia del negocio Última actualización: 2008. Disponible en: <https://docs.google.com/viewer?a=v&q=cache:yyTWOEAwnUJ:www.redciencia.info.ve/memorias/ProyProsp/trabajos/l3.doc>

ZORRILLA MARTA Data Warehouse y OLAP. Universidad de Cantabria: s.n., 2007.

ZEPEDA SÁNCHEZ LEOPOLDO Tesis. [En línea] junio de 2008. [Citado el: 23 de noviembre de 2010.] <http://tesis.com.es/autores/zepeda-sanchez-leopoldo-zenaido/>.

Anexos

Anexo 1

Tabla 12. Entrevista realizada a especialistas de la Administración Provincial de Artemisa.

Autor	Reynel Iglesias Aldecoa	Fecha	12/10/2011
Lugar	Dirección de Salud	Hora Inicio	9:00 am
Proyecto/Grupo	Dirección de Salud	Hora Terminación	10:30 am
Asunto	Levantamiento de información		
Asistentes	Orelvi Gázquez Martínez		
Ausentes			
Descripción	<p>La especialista nos explica cuáles son sus salidas trimestrales, así como el análisis que se realiza sobre cada indicador.</p> <p>Se le realizan preguntas como:</p> <ul style="list-style-type: none"> ¿Cuáles son las prioridades de la entidad que deben resolverse? ¿Cómo obtiene habitualmente los datos? ¿Cuál es la información o los tipos de reportes que actualmente se puede obtener en el negocio? ¿Con qué frecuencia se obtiene información? ¿Qué cantidad de información en general se maneja? ¿Qué criterios se tienen en cuenta para analizar los indicadores? 		

Anexo 2

Validación de calidad

Aval de Calidad de Software


El grupo de Calidad de Software del Centro de Desarrollo de la Facultad Regional "Mártires de Artemisa" conformado por:

- **Asesora de Calidad:** Ing. Maidel Ojeda Cruz
- **Asesor de Tecnología:** Ing. Domma Moreno Dager
- **Especialista de Calidad:** Ing. Yenisleydi Rodríguez Martínez


emite el presente **Aval de Calidad de Software** en colaboración con los especialistas del Centro de Desarrollo a: Geidy Acosta Méndez, como resultado satisfactorio de su desempeño en la tareas asociadas al proyecto: Almacén de Datos para la Administración Provincial de Artemisa.

Para emitir el presente aval se valoraron un conjunto de elementos evaluados de manera individual teniendo en cuenta los parámetros de calidad de software del proyecto. A continuación se presenta los resultados en cada uno de los aspectos valorados:

Elementos evaluados	Resultado
Validar la coincidencia entre requisitos informativos y estructuras de visualización.	Satisfactorio
Validar el correcto funcionamiento de las estructuras de visualización.	Satisfactorio
Validar el cumplimiento de la arquitectura de la información.	Satisfactorio
Validar estructura y funcionamiento de los metadatos y reportes ad-hoc.	Satisfactorio
Validar estructura y funcionamiento de los esquemas de mondrian.	Satisfactorio
Verificar que los reportes brindaran información coherente.	Satisfactorio
Validar el funcionamiento interno de las vistas materializadas.	Satisfactorio
Validar el correcto funcionamiento de los procesos de ETL.	Satisfactorio
Validar el tratamiento de errores en los procesos de ETL.	Satisfactorio
Generación de todos los artefactos de la metodología Kimball.	Satisfactorio
Cumplimiento de las plantillas establecidas para cada artefacto.	Satisfactorio
Ortografía, concordancia y redacción de cada uno de los artefactos.	Satisfactorio
Efectividad de los casos de pruebas definidos en el proyecto	Satisfactorio


Ing. Maidel Ojeda Cruz
Asesora de Calidad de SW




Msc. Yamila Vigil Regalado
Directora Centro de Desarrollo

Aval de Calidad de Software


El grupo de Calidad de Software del Centro de Desarrollo de la Facultad Regional "Mártires de Artemisa" conformado por:

- **Asesora de Calidad:** Ing. Maidel Ojeda Cruz
- **Asesor de Tecnología:** Ing. Domma Moreno Dager
- **Especialista de Calidad:** Ing. Yenisleydi Rodríguez Martínez


emite el presente **Aval de Calidad de Software** en colaboración con los especialistas del Centro de Desarrollo a: Reynel Iglesias Aldecoa, como resultado satisfactorio de su desempeño en la tareas asociadas al proyecto: Almacén de Datos para la Administración Provincial de Artemisa.

Para emitir el presente aval se valoraron un conjunto de elementos evaluados de manera individual teniendo en cuenta los parámetros de calidad de software del proyecto. A continuación se presenta los resultados en cada uno de los aspectos valorados:

Elementos evaluados	Resultado
Validar la coincidencia entre requisitos informativos y estructuras de visualización.	Satisfactorio
Validar el correcto funcionamiento de las estructuras de visualización.	Satisfactorio
Validar el cumplimiento de la arquitectura de la información.	Satisfactorio
Validar estructura y funcionamiento de los metadatos y reportes ad-hoc.	Satisfactorio
Validar estructura y funcionamiento de los esquemas de mondrian.	Satisfactorio
Verificar que los reportes brindaran información coherente.	Satisfactorio
Validar el funcionamiento interno de las vistas materializadas.	Satisfactorio
Validar el correcto funcionamiento de los procesos de ETL.	Satisfactorio
Validar el tratamiento de errores en los procesos de ETL.	Satisfactorio
Generación de todos los artefactos de la metodología Kimball.	Satisfactorio
Cumplimiento de las plantillas establecidas para cada artefacto.	Satisfactorio
Ortografía, concordancia y redacción de cada uno de los artefactos.	Satisfactorio
Efectividad de los casos de pruebas definidos en el proyecto	Satisfactorio


Ing. Maidel Ojeda Cruz
Asesora de Calidad de SW




Msc. Yamila Vigil Regalado
Directora Centro de Desarrollo