



Facultad 6 Centro de Tecnologías de Gestión de Datos
(DATEC)

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

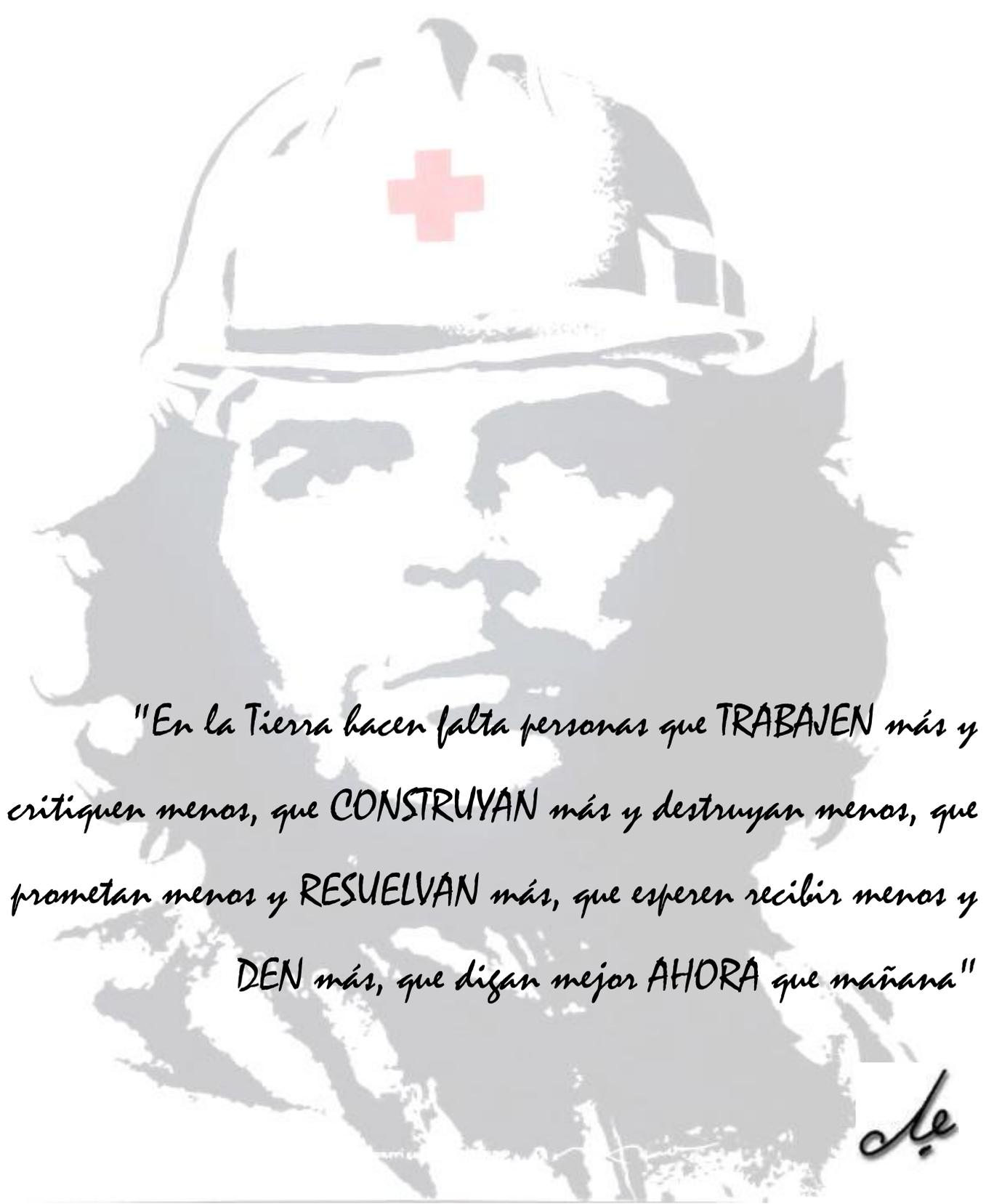
Título: Mercado de datos para el Módulo Visor de Historias
Clínicas del Sistema Integral de Atención Primaria de la Salud.

Autores: Alejandro Queralta Pozo
Osmar Ruano Valdés

Tutores: Ing. Yunier Santana Aldana.
Ing. Yosvany Arrastia Machín.

La Habana, 2012

“Año 54 de la Revolución”



"En la Tierra hacen falta personas que TRABAJEN más y critiquen menos, que CONSTRUYAN más y destruyan menos, que prometan menos y RESUELVAN más, que esperen recibir menos y DEN más, que digan mejor AHORA que mañana"

de

Agradecimientos

Agradecimientos de Alejandro:

A mi madre por ser la persona más especial en mi vida, si su apoyo hubiera sido imposible llegar hasta aquí. Por todo el esfuerzo que ha hecho para que pudiéramos estar aquí.

A mi abuelo Bebito que a pesar de su ausencia siempre me acompañan todas sus enseñanzas y su cariño.

A mi padre que a pesar de las dificultades me ha apoyado en todo lo que ha podido.

A mis dos hermanos Iván y Rafe por ser los amigos que son y lo importante que son en mi vida. Rafe te admiro.

A mis sobrinitas que son la alegría de la casa.

A mi novia por toda la ayuda que me ha dado en estos tres años juntos, por ser amiga compañera y confidente. Por siempre estar cuando la necesito y por los buenísimos momentos que me ha regalado.

A mis suegros Ernesto y Maritza por ser tan preocupados y cariñosos conmigo.

A mi amigo Osmar que a pesar de siempre llevarme la contraria es como otro hermano para mí y sin su ayuda hubiera sido más difícil llegar hasta aquí. Gracias Hermano.

A los amigos que crecieron conmigo y siempre están ahí para las buenas y las malas: Jorgito, David, Joel, Danilo. Alain, Delia, El Rada, Miguelito, El gato, Sergito, El Rafa.

A mis amigos de aquí de la UCI: A Juanita, Yariel, Vladimir, Yan, Yaniel, Silvio, Raúl, Cesar, Ramiro, Victor, Ernesto, Santana, La gorda, Darién, Dausvel, Mijito, Héctor, Janier, El pily, Danay, Liena y a todos los que me han ayudado de veras gracias.

A mis tutores y demás profes que me han ayudado de alguna forma gracias.

Agradecimientos

Agradecimientos de Osmar:

A mi madre por ser la fuerza que me acompaña en los buenos y malos momentos y en honor a ella es que van enmarcados todos mis esfuerzos.

A mi hermana que lo es todo para mí, por brindarme su apoyo y su confianza, sin ti no hubiera llegado hasta aquí.

A mis sobrinos por alegrarme la vida y sé que para ellos soy un ejemplo a seguir.

A mi papa por todo su apoyo y sé que está orgulloso de mi.

Al papa de mi hermana porque solo yo se lo tanto que le debo, nunca voy a poder retribuirte todo lo que incondicionalmente me ha brindado.

A mi abuelo, abuela, tías, tíos y primos.

A mi tía Deysi, a Omniel y Omnielito por ser mis camaroncitos duros aquí en la Habana.

A mi Juana por ayudarme, apoyarme y aguantarme todo este tiempo por brindarme todo el amor y el cariño de mundo, por ser tan única, la mejor de la historia.

A Alejandro por ser mi hermanito y que si no existiera no tuviera con quien fajarme. Gracias mi hermano.

A mis amigos de toda la vida, El rada, Sergito, Medardo, Arito, Timy, Miguelito, Miguel, Osvaldito, Idalberto, Rafa, Marlis, El rafa, El yuco y Alain.

A mi pique de aquí de la UCI que los quiero mucho, Yan, Ramiro, Janier, Mijito, Yariel, Santana, Yaniel, Victor, El vlado, El jota, El pily, Silvio, Raúl, Carlos, Cesar, Dayi, Ismel, Mailen , Yisel, Anabel, Delia, Dauvel, Danay y todos lo que de una forma u otra me han ayudado.

A mis suegros por ser tan preocupados.

A mis tutores, a Monchy y Lisdan.

Declaratoria de Autoría

Declaración de Autoría

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

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

Osmar Ruano Valdés

Alejandro Queralta Poso

Firma del Autor

Firma del Autor

Ing. Yunier Santana Aldana

Ing. Yosvany Arrastia Machín

Firma del Tutor

Firma del Tutor

Datos de Contacto

Nombre y Apellidos: Ing. Yunier Santana Aldana

Correo: ysaldana@uci.cu

Situación laboral: Profesor, Instructor Facultad 6.

Años de graduado: 3 años.

Especialidad de graduación: Ingeniero en Ciencias Informáticas.

Institución a la que pertenece: Universidad de las Ciencias Informáticas (UCI).

Dirección: Carretera San Antonio de los Baños, Torrens, Municipio Boyeros, Ciudad de La Habana, Cuba, Código postal 19370.

Nombre y Apellidos: Ing. Yosvany Arrastia Machín.

Correo: yarrastia@uci.cu

Situación laboral: Profesor, Instructor Facultad 6.

Años de graduado: 3 años.

Especialidad de graduación: Ingeniero en Ciencias Informáticas.

Institución a la que pertenece: Universidad de las Ciencias Informáticas (UCI).

Dirección: Carretera San Antonio de los Baños, Torrens, Municipio Boyeros, Ciudad de La Habana, Cuba, Código postal 19370.

Resumen:

El presente trabajo de diploma se enmarca en el tema de los almacenes de datos, los mercados de datos, y su utilización para apoyar el proceso de toma de decisiones. En el mismo se realiza una investigación detallada. Para escoger la metodología y las herramientas a utilizar, además de especificar las necesidades de información del cliente y el motivo por el cual es necesario el desarrollo del sistema, tomándose como base para el correcto diseño e implementación del Mercado de datos para el Modulo Visor de Historias Clínicas del Sistema Integral de Atención Primaria de la Salud. De igual manera la solución propuesta contiene las políticas de seguridad, respaldo y recuperación de los datos, además de contar con las pruebas realizadas al Mercado de datos. Como resultado final del trabajo se obtiene un mercado de datos poblado y una capa de visualización de los datos.

Palabras claves: mercado de datos, toma de decisiones:

Índice de Contenido

Introducción.....	1
Capítulo 1: Fundamentos teóricos de la Investigación.....	5
1.1 Tecnología para la toma de decisiones.....	5
1.2 Definición de un Almacenes de datos.....	5
1.2.1 Objetivos de un Almacén de datos	6
1.2.2 Ventajas de un Almacén de datos	6
1.2.3 Desventajas de un Almacén de datos.....	7
1.3 Soluciones de almacenes de datos utilizados a nivel mundial	7
1.3.1 Soluciones de almacenes de datos utilizados en Cuba	8
1.4 Definición de Mercado de datos	8
1.4.1 Arquitectura de un Mercado de datos.....	8
1.4.2 Ventajas de la utilización de los Mercados de datos	9
1.5 Modelo multidimensional.....	9
1.5.1 Tipos de Modelados de un Almacén de datos	10
1.6 Modo de Almacenamiento OLAP.....	10
1.6.1 Procesamiento Analítico Multidimensional (MOLAP).....	10
1.6.2 Procesamiento Analítico en Línea Relacional (ROLAP)	11
1.6.3 Procesamiento Analítico Híbrido en Línea (HOLAP)	11
1.7 Integración de datos.....	11
1.8 Inteligencia de Negocios.....	13
1.8.1 Proceso de Inteligencia de Negocio.....	13
1.9 Metodología de desarrollo	14
1.9.1 Metodologías para los almacenes de datos	14
1.10 Herramientas.....	17
1.10.1 Herramientas para el proceso de ETL.....	17
1.10.2 Sistema Gestor de Base Datos.....	17
1.10.3 Herramientas de Modelado.....	18
1.10.4 Administrador de Base de datos.....	19
1.10.5 Herramientas de Inteligencia de negocios	19
Conclusiones del capítulo.....	20

Capítulo 2: Análisis y Diseño del Mercado de datos para Modulo Visor de Historias Clínicas del Sistema Integral de Atención Primaria de la salud.....	22
Introducción	22
2.1 Tema de análisis	22
2.1.1 Reglas del negocio.....	22
2.1.2 Necesidades de los usuarios.....	23
2.1.3 Requisitos de información.....	23
2.1.4 Requisitos funcionales	24
2.1.5 Requisitos no funcionales	25
2.2.1 Descripción textual de los casos de usos	29
2.3 Arquitectura del Mercado de datos.....	31
2.3.1 Hechos.....	32
2.3.2 Dimensiones.....	33
2.4 Matriz BUS.....	34
2.5 Modelo de Datos	35
2.6 Subsistema de integración.....	36
2.7 Roles y permisos	37
2.8 Salva de la base de datos.....	38
Capítulo 3: Implementación del Mercado de datos Visor de Historias Clínicas del Sistema Integral de Atención Primaria de la Salud.....	39
3.1 Implementación del modelo de datos.....	39
3.2. Perfilado de los datos	40
3.3. Transformaciones para la carga de las tablas de dimensiones	40
3.4. Transformación para cargar el hecho	41
3.5. Implementación de los trabajos.....	42
3.6. Implementación del subsistema de visualización	42
3.6.1. Implementación de los cubos OLAP	42
3.7 Navegación de la capa de visualización.....	44
3.8 Creación y administración de los reportes	45
Conclusiones del capítulo.....	46
Capítulo 4: Pruebas del Mercado de datos para el modulo Visor de Historias Clínicas del Sistema Integral de atención Primaria de la Salud.....	47
Introducción	47

4.1 Pruebas.....	47
4.2 Herramientas de pruebas	48
4.3. Caso de Prueba.....	50
Conclusiones del capítulo.....	53
Conclusiones Generales	54
Recomendaciones	55
Referencias Bibliográficas	56
Bibliografía.....	59
Anexos.....	62

Índice de Tablas

Tabla 1: Requisitos funcionales	25
Tabla 2: Tabla de hechos	33
Tabla 3: Tabla de dimensiones	34
Tabla 4: Roles y permisos	38
Tabla 5: Esquemas y tablas del mercado.....	40

Índice de Figuras

Figura 1. Diagrama de casos de uso	29
Figura 2: Diseño de la arquitectura de un Almacén de datos.....	32
Figura 3: Matriz Buz.....	35
Figura 4. Modelo de datos	36
Figura 5. Diseño de las transformaciones para la carga de las dimensiones	37
Figura 6. Diseño de las transformaciones para la carga de los hechos.	37
Figura 7. Transformación para la carga de la dimensión carácter_transfucion.....	41
Figura 8. Transformación para la carga del hech_emergencia	41
Figura 9. Transformación del trabajo para la carga de los hechos	42
Figura 10. Diseño de los cubos OLAP.....	43
Figura 11: Diseño del cubo partos_realizados	43
Figura 12. Diseño del mapa de navegación	44
Figura 13 : Reporte Cantidad de autopsias realizadas en el año 2009.....	46
Figura 14: Gráfico del reporte Autopsias realizadas en el año 2009.....	46

Introducción

En la actualidad en el mundo existe un crecimiento exponencial de los datos en todas las instituciones y empresas, generado por sus propias actividades y en su gran mayoría cuentan con registro *histórico* de actividades que crece con el transcurso de los días. La necesidad del resguardo de información crítica, ha obligado con el transcurso de los años a digitalizar todos estos datos. Es por ello que a nivel mundial todas las instituciones han automatizado todas sus actividades.

A raíz de los avances científicos-tecnológicos alcanzados en el sector de la Informática y las Comunicaciones, se abren nuevas posibilidades, una de ellas es que aplicando las nuevas técnicas a estos registros históricos de datos, los mismos son capaces de arrojar información importante y a la que antes no se tenía acceso. Siendo necesario aclarar que los términos “datos” e “información”, a pesar de parecer similares, son de significados totalmente distintos, ya que los “datos” no son más que un conjunto de observaciones, hechos e imágenes y la “información” nace del proceso de formalización, filtrado, contextualizado y resumen que sufren estos datos para así convertirse en información. (18)

Como consecuencia de estos avances es posible realizar un análisis estadístico de los datos almacenados, dando la posibilidad de una comparación objetiva de los resultados obtenidos por la empresa a través de su comportamiento histórico. Estos análisis posibilitan la extracción de la información que contienen estos registros históricos almacenados, según los temas de análisis que se necesiten conocer, convirtiéndose en una base confiable, para el apoyo del proceso de toma de decisiones. En la actualidad el anhelo de poder contar con sistemas que apoyen el proceso de toma de decisiones ha pasado de ser una alternativa a convertirse en una necesidad. La efectividad de este tipo de sistema se debe en gran medida a que, posibilitan un análisis detallado de un determinado ámbito empresarial, del cual se necesita conocer al máximo su comportamiento histórico, para facilitar la toma de decisiones en este punto, siempre buscando la eficiencia de los servicios o la máxima satisfacción de los clientes.

Cuba no está exenta de estos avances científico-tecnológicos. En cualquier rama de la economía en la que se apliquen este tipo de avances asegura crecimiento económico, una mejoría palpable en los servicios que se prestan y una mayor madurez empresarial. A pesar de las dificultades, el gobierno cubano ha ido introduciendo este tipo de técnicas paulatinamente en dependencia de las necesidades y las posibilidades. La Universidad de las Ciencias Informáticas (UCI), juega un papel protagónico en el avance científico-técnico de Cuba. La UCI, está estructurada por distintos centros que desarrollan diferentes soluciones informáticas. Entre ellos se encuentra el Centro de Informática Medica (CESIM) y

el mismo cuenta con el Sistema Integral de Atención Primaria de la Salud (SIAPS), que satisface problemas como el procesamiento de datos médicos y el almacenamiento de la información resultante de los procesos hospitalarios médico-administrativos, de las distintas áreas que abarca. El sistema antes mencionado, cuenta con un gran cúmulo de datos resultado de las actividades de digitalización que llevan a cabo en el área hospitalaria. En la actualidad presentan diversas dificultades a la hora de realizar el análisis de estos, que imposibilita un análisis adecuado de indicadores viéndose afectado el proceso de toma de decisiones de una manera ágil y efectiva.

Debido a esto se identifica el siguiente **problema de la investigación**: ¿Cómo contribuir a la toma de decisiones en el Módulo Visor de Historias Clínicas del Sistema Integral de Atención Primaria de la Salud?

Definiendo como **objeto de estudio**: Los almacenes de datos e inteligencia de negocios, centrando el **campo de acción**: Mercado de datos y capa de visualización para el Módulo Visor de Historias Clínicas del Sistema Integral de Atención Primaria de la Salud.

Para dar respuesta al problema planteado se determina como **objetivo general** desarrollar un mercado de datos para el Módulo Visor de Historias Clínicas del Sistema Integral de Atención Primaria de la Salud, que apoye el proceso de toma de decisiones.

Derivándose los siguientes **Objetivos Específicos**:

1. Identificar las ventajas y desventajas de las distintas soluciones para el apoyo a la toma de decisiones.
2. Realizar el análisis y diseño del mercado de datos.
3. Implementar el mercado de datos.
4. Realizar pruebas al mercado de datos.

Para dar cumplimiento a los objetivos de la investigación se plantean las siguientes tareas de la investigación:

1. Caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo de almacenes de datos.
2. Levantamiento de requisitos.
3. Descripción de los casos de uso del mercado de datos.
4. Definición de los hechos, las medidas y las dimensiones del mercado de datos.

5. Diseño del modelo de datos.
6. Definición de la arquitectura del mercado de datos.
7. Diseño del subsistema de integración de datos.
8. Diseño del subsistema de visualización.
9. Diseño de los casos de pruebas.
10. Implementación del subsistema de integración.
11. Implementación del subsistema de visualización.
12. Aplicación de las listas de chequeo.
13. Aplicación de los casos de prueba.

Como **resultado esperado** se obtendrá un mercado de datos poblado y una capa de visualización de los datos.

Para cumplir lo antes expuesto el presente Trabajo de Diploma se ha estructurado en cuatro capítulos:

Capítulo 1: Fundamentos teóricos de la investigación

En este capítulo se define todos los aspectos relacionados con el desarrollo de los almacenes de datos, se exponen los conceptos fundamentales que serán tratados a lo largo de la investigación, así como la selección de la metodología, técnicas y herramientas que serán utilizadas.

Capítulo 2: Análisis y diseño del Mercado de datos para el Módulo Visor de Historias Clínicas del Sistema Integral de Atención Primaria de la Salud.

Este capítulo define todo lo referente a el análisis y diseño del mercado de datos para el Módulo Visor de Historias Clínicas del Sistema Integral de Atención Primaria de la Salud: hace referencia al flujo de trabajo de análisis y diseño, en el cual se definen los requerimientos del sistema, se diseña el diagrama de casos de uso del sistema; así como los subsistemas de integración y visualización.

Capítulo 3: Implementación del Mercado de datos para el Módulo Visor de Historias Clínicas del Sistema Integral de Atención Primaria de la Salud.

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

Capítulo 4: Pruebas del Mercado de datos para el Módulo Visor de Historias Clínicas del Sistema Integral de Atención Primaria de la Salud.

Una vez realizado el análisis, diseño e implementación del mercado de datos, se da paso a la validación y prueba de la solución mediante las listas de chequeo, los casos de prueba y las pruebas de aceptación para así verificar que el sistema cumpla con los requerimientos necesarios que garanticen al usuario final del sistema la confiabilidad de los datos cargados en el almacén de datos.

Capítulo 1: Fundamentos teóricos de la Investigación

En este capítulo se definen temas relacionados con el desarrollo de los mercados de datos, se exponen los conceptos fundamentales que serán tratados a lo largo de la investigación e implementación de la solución, así como la selección de la metodología, técnicas y herramientas que serán utilizadas.

1.1 Tecnología para la toma de decisiones

Un almacén de datos se crea al extraer datos desde una o más bases de datos de aplicaciones operacionales. La data extraída es transformada para eliminar inconsistencias y resumir si es necesario y luego ser cargada en el almacén. El proceso de transformar, crear el detalle de tiempo variante, resumir y combinar los extractos de datos, ayudan a crear el ambiente para el acceso a la información Institucional. Este nuevo enfoque ayuda a las personas individuales, en todos los niveles de la empresa, a efectuar su toma de decisiones con más responsabilidad. (1)

1.2 Definición de un Almacenes de datos

El término Datawarehouse fue definido por primera vez por Bill Inmon (3). Su traducción al español Almacén de Datos “*Es una colección de datos orientados por temas, integrados, variables en el tiempo y no volátiles para el apoyo de la toma de decisiones*”. Orientados por tema porque los datos son estructurados según los temas de interés para facilitar su análisis. Integrado debido a que los datos que se introducen en el almacén de datos se obtienen de diversas fuentes de origen y distintos formatos, y es tarea del proceso Extracción, Transformación y Carga, aplicarles las transformaciones necesarias para integrarlos. Variables en el tiempo porque los datos siempre tienen que estar ligados a un instante de tiempo. Y no volátil porque en el almacén los datos pueden ser consultados pero no modificados, por tanto, la información es permanente y la actualización consiste exclusivamente en la incorporación de nuevos datos.

Por su parte, Ralph Kimball quien es una de las personalidades más influyentes en el área, propone otra definición al catalogarlo como: (3) “Una copia de datos transaccionales, específicamente estructurados para la consulta y el análisis”.

Organizado en torno a temas: La información se clasifica en base a los aspectos que son de interés para la empresa.

Integrado: Es el aspecto más importante. La integración de datos consiste en convenciones de nombres, codificaciones consistentes, medida uniforme de variables.

Dependiente del 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).
- La información, una vez registrada correctamente, no puede ser actualizada.

No volátil: El Almacén de Datos sólo permite cargar nuevos datos y acceder a los ya almacenados, pero no permite ni borrar ni modificar los datos. (4)

A pesar de plantear conceptos un poco diferentes, los dos convergen en un mismo punto y es que los almacenes de datos tienen como función fundamental, organizar la información recogida de diferentes fuentes para facilitar su análisis y apoyar el proceso de toma de decisiones.

1.2.1 Objetivos de un Almacén de datos

Los principales objetivos de un Almacén de datos son:

- Proveer una visión única de los clientes en toda la empresa.
- Poner tanta información como sea posible en manos de todos los usuarios.
- Medir el desempeño del trabajo según las medidas requeridas.
- Monitorear el comportamiento de los clientes.
- Analizar los resultados de los reportes en dependencia de
- Apoyar la toma de decisiones en la organización que se utilice.
- Incrementar la precisión de las mediciones.
- Aumentar la productividad.
- Incrementar y distribuir las responsabilidades.(2)

1.2.2 Ventajas de un Almacén de datos

- Transforma datos orientados a las aplicaciones, en información orientada a la toma de decisiones.
- Permite un análisis inmediato de las actividades de la empresa.
- Agilidad en el control de los activos tangibles de la empresa.
- Capacidad de analizar y explorar las diferentes áreas de trabajo.
- Relación total con el cliente.

- Facilidades en la gestión y análisis de recursos.
- Permite establecer una conexión entre los departamentos empresariales que son independientes del resto.
- Reacciona rápidamente a los cambios del mercado.

1.2.3 Desventajas de un Almacén de datos

- Gran inversión que supone este tipo de proyectos.
- La tecnología no se encuentra del todo madura.

1.3 Soluciones de almacenes de datos utilizados a nivel mundial

Desde un inicio, las bases de datos se convirtieron en una herramienta fundamental de control y manejo de las operaciones comerciales. Fue así como en unos pocos años en grandes empresas y negocios existía un considerable número de información almacenada en diferentes fuentes de datos y estas ya habían alcanzado un tamaño considerablemente grande. Con esta gran acumulación de información, los directivos de tales empresas y negocios se dieron cuenta que esta podría tener un fin útil, al estar reflejada la mayoría de sus operaciones comerciales durante los llamados ciclos de negocios propios del mercado. A su vez, los mercados empresariales han experimentado una transformación radical. Las empresas demandan mayor rapidez y eficiencia en la entrega de productos, y mejora en todos los servicios existentes, por lo que 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 y, en definitiva, mayor rapidez a la hora de tomar decisiones. (5)

Ya en 1994 el 90% de las empresas, según la revista Fortune 2000, planeaban implementar un almacén de datos. Ya en 1996 el 90% de las grandes corporaciones consideraba adoptar esta tecnología. A raíz de los acontecimientos anteriores, se desencadenó una revolución en esta esfera, por lo que Hill Hostian, de la empresa Gartner, estimó que para el 2007, el 50% de los proyectos de inteligencia de negocio requerirían de alguien que les proveyera los servicios, para librarse de los obstáculos debidos a falta de personal capacitado y de los recursos.(3)

Las esferas bancarias y de seguros, también han dado pasos concretos en este sentido, con entidades como Banco Galicia en Argentina, Banco de México, Banorte y Banamex en México, Banco París en Francia y el Europa Central Bank perteneciente a la Unión Europea. Todos ellos son ejemplos palpables del avance, y exhiben estudios sobre inflación, población, monetarios y tendencias. El tema estadístico no ha estado ajeno a estas necesidades. En países como México (específicamente en el INEGI (Instituto Nacional de Estadísticas e Informática)), se tiene la información en almacenes de datos, y de esta forma es utilizada para la toma de decisiones a nivel gubernamental. De la misma

manera el Instituto Nacional de Estadística de España es otro ejemplo de entidades que actualmente están utilizando la tecnología.(6)

Muchas empresas de importante nivel profesional y productivo a escala mundial como IBM (*International Business Machines*) que ahora se embarcan en su gran proyecto de la construcción del mayor almacén de datos del mundo, pretende construir un repositorio de datos diez veces mayor que cualquier otro existente, llegando a los 120 petabytes, es decir, 120 millones de Gigabytes. Ellos necesitan esa capacidad de almacenamiento para un supercomputador que se dedicará a realizar simulaciones de fenómenos naturales. (7)

1.3.1 Soluciones de almacenes de datos utilizados en Cuba

En Cuba, algunas instituciones han comenzado a hacer uso de las ventajas de un almacén de datos como es el caso de la corporación CIMEX y la Oficina Nacional de Estadísticas (ONE. Dando la posibilidad a estas instituciones de aprovechar las ventajas que brindan estas tecnologías para el desarrollo de las misma. El almacén de datos perteneciente a la ONE es desarrollado por la Universidad de las Ciencias Informáticas. Las empresas que utilizan almacenes de datos generalmente son aquellas que manejan grandes volúmenes de datos relativos a clientes, marketing, transacciones y operaciones, por lo que es muy difícil encontrar pequeñas y medianas empresas que utilicen estos sistemas. (6)

1.4 Definición de Mercado de datos

Un mercado de datos es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. (19)

1.4.1 Arquitectura de un Mercado de datos

Los mercados de datos que a su vez pueden ser almacenes de datos presentan la siguiente arquitectura:

- **Sistemas Operacionales:** es la fuente de datos donde se nutre el almacén. Pueden estar organizados en archivos indexados, bases de datos jerárquicas o en sistemas de base de datos relacionales.
- **Extracción, Transformación y Carga:** el primer paso es la extracción de los datos de los sistemas operacionales; luego se realizan las transformaciones necesarias para que solo quede la información que el cliente desee y en el formato correspondiente; por último esta información “limpia” es cargada al almacén.

- **Creación de los metadatos:** es el proceso de crear la información sobre los datos que van a ser almacenados. Describe cómo se almacenan, se integran y se transforman estos datos.
- **Acceso de usuarios finales:** los usuarios finales acceden a la información del DWH a través de una interfaz gráfica. Estas pueden incluir *software* de consultas, generadores de reportes, procesamientos analíticos en línea, entre otras funciones, dependiendo de lo que desee el cliente.

1.4.2 Ventajas de la utilización de los Mercados de datos

- Acelera las consultas al reducir la cantidad de datos a recorrer.
- Estructura los datos para su adecuado acceso por una herramienta.
- Divide los datos para imponer estrategias de control de acceso.
- Segmenta los datos en diferentes plataformas de hardware.
- Permite el acceso a los datos mediante un gran número de herramientas del mercado, logrando así la independencia de estas.

1.5 Modelo multidimensional

El proceso del modelado de datos busca transformar modelos de datos orientados a procesos en modelos de datos orientados a negocios, los modelos dimensionales. Todo modelo dimensional es compuesto de una tabla con una llave compuesta de varias partes, llamado Tabla de hechos y un grupo de tablas menores, llamadas Tablas de dimensiones.(14)

Modelos de datos como conjuntos de medidas descritas por dimensiones.

- Adecuado para resumir y organizar datos, por ejemplo hojas de cálculo.
- Enfocado para trabajar sobre datos de tipo numérico.
- Más simple: Más fácil de visualizar y entender que el modelado E/R.

Tabla de Hechos: Es la tabla central en un esquema dimensional. Es en ella donde se almacenan las mediciones numéricas del negocio. Estas medidas se hacen sobre el grano, o unidad básica de la tabla.

Tabla de Dimensiones: Estas tablas son las que se conectan y alimentan a la Tabla de Hechos, almacena un conjunto de valores que están relacionados a una dimensión particular.

Medidas: Atributo numérico de un hecho. Representan el comportamiento del negocio relativo a las dimensiones.

Indicadores: Son sumalizaciones efectuadas sobre algún hecho, que serán incluidos en algún cubo multidimensional, con el fin de analizar los datos almacenados en el DW. El valor que estos adopten estará condicionado por los atributos/jerarquías que se utilicen para analizarlos.

Atributos: Constituyen los criterios de análisis que se utilizarán para analizar los indicadores dentro de un cubo multidimensional. Los mismos se basan, en su gran mayoría, en los campos o datos de referencia de las tablas de dimensiones. Dentro de un cubo multidimensional, los atributos son los ejes del mismo.

Granularidad: Representa el nivel de detalle al que se desea almacenar la información sobre el negocio que se esté analizando. Por ejemplo, los datos referentes a ventas o compras realizadas por una empresa, pueden registrarse día a día, en cambio, los datos pertinentes a pagos de sueldos o cuotas de socios, podrán almacenarse a nivel de Mes.

1.5.1 Tipos de Modelados de un Almacén de datos

El modelo multidimensional incluye tres maneras de modelar el sistema según su complejidad:

- **Esquema Estrella:** Su estructura base es similar: una tabla central y un conjunto de tablas que la atienden radialmente, de este se deriva su nombre del hecho que su diagrama forma una estrella, con puntos radiales desde el centro.
- **Esquema copo de nieve:** Es un esquema derivado del esquema de estrella, las tablas de dimensiones se ramifican en más puntas.
- **Constelación de hechos:** La constelación de hechos es un conjunto de tablas de hechos que comparten algunas tablas de dimensiones.

1.6 Modo de Almacenamiento OLAP

Los sistemas OLAP (Procesamiento Analítico en Línea) son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejos. Existen diferentes tipos de almacenamientos OLAP entre ellos MOLAP, HOLAP y ROLAP. (20)

1.6.1 Procesamiento Analítico Multidimensional (MOLAP)

La arquitectura MOLAP usa bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos

multidimensionalmente. Utiliza una arquitectura de dos niveles: Las bases de datos multidimensionales y el motor analítico donde la base de datos multidimensional es la encargada del manejo, acceso y obtención del dato.

La arquitectura MOLAP requiere unos cálculos intensivos de compilación. Lee datos precompilados, y tiene capacidades limitadas de crear agregaciones dinámicamente o de hallar ratios que no se hayan precalculados y almacenados previamente.

1.6.2 Procesamiento Analítico en Línea Relacional (ROLAP)

La arquitectura ROLAP, accede a los datos almacenados en un almacén de datos para proporcionar los análisis OLAP. La premisa de los sistemas ROLAP es que las capacidades OLAP se soportan mejor contra las bases de datos relacionales.

La arquitectura ROLAP es capaz de usar datos precalculados si estos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del almacén de datos, y soporta técnicas de optimización de accesos para acelerar las consultas. Estas optimizaciones son, entre otras, particionado de los datos a nivel de aplicación, soporte a la desnormalización y uniones múltiples.

1.6.3 Procesamiento Analítico Híbrido en Línea (HOLAP)

Combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado.

Después de realizarse un análisis de los modos de almacenamiento de datos mencionados anteriormente, se llegó a la conclusión de que el más adecuado a la solución es ROLAP, porque está orientado a bases de datos relacionales, que será la utilizada en el desarrollo de la investigación.

1.7 Integración de datos

Todas las compañías buscan poder contar con la información adecuada en el momento preciso, de mezclar las diversas necesidades de información presente en diferentes fuentes (bases de datos, ficheros), y de proveer una panorámica unificada de esta información para el usuario. Precisamente este es el objetivo de la integración de datos, permitir el desarrollo rápido de nuevas aplicaciones que requieren información de múltiples fuentes.

Según lo expuesto por Laura Haas (8) “La integración de datos es un proceso que cuenta con cuatro tareas principales: la comprensión, la normalización, especificación y ejecución”.

- **Comprensión:** La primera tarea de la integración de datos es analizar y entender la fuente. Durante esta tarea, el integrador puede buscar relaciones entre los datos y sus significados.
- **Normalización:** Esta tarea aprovecha el trabajo de la tarea Comprensión para determinar el mejor método de integración, la forma de limpiar o reparar los datos.
- **Especificación:** En esta tarea se producen los artefactos que controlarán la ejecución. Las técnicas y tecnologías utilizadas para obtener las especificaciones están íntimamente vinculadas a la elección del motor de ejecución. La especificación es parte de la configuración de un motor de integración para hacer la integración deseada.
- **Ejecución:** Aquí es donde realmente sucede la integración de datos y puede ser lograda a través de la Materialización, Federación o Indexación.

ETL este término viene de inglés de las siglas (Extract-Transform-Load) que significan Extraer, Transformar y Cargar y se refiere a los datos en una empresa. ETL es el proceso que organiza el flujo de los datos entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un almacén de datos, reformatear, limpiar y cargar los mismos. Es importante destacar que el éxito del almacén de datos está dado por el amplio estudio del negocio partiendo del conocimiento obtenido de las fuentes de datos externas y de los responsables de su mantenimiento, no del simple copiado de la información. (21)

ETL es un proceso donde se desarrollan las siguientes tareas:

Extracción: Se basa en las necesidades y requisitos del usuario, se exploran las diversas fuentes de Transacción y Procesamiento en línea (OLTP) que se tengan a disposición, y se extrae la información que se considere relevante al caso. Una vez que los datos son seleccionados y extraídos, se guardan en un almacenamiento intermedio, lo cual permite, entre otras ventajas:

- Manipular los datos sin interrumpir ni paralizar los OLTP, ni tampoco el Almacén de datos.
- Almacenar y gestionar los metadatos que se generarán en los procesos ETL.
- Facilitar la integración de las diversas fuentes, internas y externas.

Transformación: Esta función es la encargada de convertir aquellos datos inconsistentes en un conjunto de datos compatibles y congruentes, para que puedan ser cargados en el DW. Estas acciones se llevan a cabo, debido a que pueden existir diferentes fuentes de información, y es vital conciliar un formato y forma única, definiendo estándares, para que todos los datos que ingresarán al DW estén integrados.

Los casos más comunes en los que se deberá realizar integración, son los siguientes:

- Codificación.

- Medida de atributos.
- Convenciones de nombramiento.
- Fuentes múltiples.

Además de lo antes mencionado, esta función se encarga de realizar el proceso de Limpieza de Datos (Data Cleaning).

Carga: Este proceso es el responsable de cargar la estructura de datos del DW con:

- Aquellos datos que han sido transformados y que residen en el almacenamiento intermedio.
- Aquellos datos de los OLTP que tienen correspondencia directa con el depósito de datos.

Se debe tener en cuenta, que los datos antes de moverse al almacén de datos, deben ser analizados con el propósito de asegurar su calidad, ya que este es un factor clave, que no debe dejarse de lado.

Los ETL, son los encargados de realizar dos tareas bien definidas:

- Carga Inicial (Initial Load).

Actualización, mantenimiento o refresco periódico (siempre teniendo en cuenta un intervalo de tiempo predefinido para tal operación).

1.8 Inteligencia de Negocios

Actualmente, en las actividades diarias de cualquier organización, se generan datos como producto secundario, que son el resultado de todas las transacciones que se realizan. Es muy común, que los mismos se almacenen y administren a través de sistemas transaccionales en bases de datos relacionales. (18).

Definición de Inteligencia de negocios

Se puede describir la Inteligencia de Negocios (BI), como un concepto que integra por un lado el almacenamiento y por el otro el procesamiento de grandes cantidades de datos, con el principal objetivo de transformarlos en conocimiento y en decisiones en tiempo real, a través de un sencillo análisis y exploración.

La definición antes expuesta puede representarse a través de la siguiente fórmula:

$$\text{Datos} + \text{Análisis} = \text{Conocimiento. (18)}$$

1.8.1 Proceso de Inteligencia de Negocio

Para comprender cómo es que las organizaciones pueden crear inteligencia de sus datos y brindar a los usuarios finales el acceso a la información deseada. A continuación se muestran las cinco fases:

Fase 1: Dirigir y Planear.

Fase 2: Recolección de Información.

Fase 3: Procesamiento de Datos.

Fase 4: Análisis y Producción.

Fase 5: Difusión. Finalmente. (18)

Beneficios

Entre los beneficios más importantes que BI proporciona a las organizaciones, vale la pena destacar los siguientes:

- Reduce el tiempo mínimo que se requiere para recoger toda la información relevante de un tema en particular, ya que la misma se encontrará integrada en una fuente única de fácil acceso.
- Automatiza la asimilación de la información, debido a que la extracción y carga de los datos necesarios se realizará a través de procesos predefinidos.
- Proporciona herramientas de análisis para establecer comparaciones y tomar decisiones.
- Cierra el círculo que hace pasar de la decisión a la acción.
- Permite a los usuarios no depender de reportes o informes programados, porque los mismos serán generados de manera dinámica.
- Posibilita la formulación y respuesta de preguntas que son claves para el desempeño de la empresa.
- Permite acceder y analizar directamente los indicadores de éxito.
- Se pueden identificar cuáles son los factores que inciden en el buen o mal funcionamiento de la empresa.
- Permitirá predecir el comportamiento futuro con un alto porcentaje de certeza, basado en el entendimiento del pasado.
- El usuario podrá consultar y analizar los datos de manera sencilla e intuitiva. (18)

1.9 Metodología de desarrollo

“La metodología es la filosofía que sustenta todo el proceso investigativo y a los métodos en particular” (22). Es un conjunto de pasos y procedimientos que deben seguirse que favorecen a los desarrolladores a dividir las etapas de un proyecto, qué tareas se llevan a cabo en cada etapa, qué restricciones deben aplicarse, qué técnicas y herramientas se emplean y cómo se controla y gestiona un proyecto.

1.9.1 Metodologías para los almacenes de datos

Existen diferentes metodologías para el desarrollo de los almacenes de datos, entre las que se destacan la metodología de Inmon y la metodología Kimball. Ralph Kimball y Bill Inmon son dos de las personalidades referentes y más influyentes en el área de los Almacenes de Datos. Ralph Kimball es un especialista reconocido a nivel mundial en el diseño de los almacenes de datos y creador del enfoque Multidimensional; mientras que Bill Inmon es el creador del término almacenes de datos y considerado como el padre de la disciplina. La principal diferencia que existe entre ambas

metodologías está basada en la forma de enfrentar el problema. La visión de Inmon se basa en un enfoque descendente top-down, la cual propone construir primero el almacén de datos y a partir de este los mercados de datos, plantea la creación de un repositorio de datos corporativo como fuente de información consolidada, persistente, histórica y de calidad. Al ser construido descendentemente los mercados de datos se nutren del almacén de datos corporativo, convirtiéndose en un complejo empresarial de bases de datos relacionales.

Mientras Inmon afirma que la creación de una base de datos relacional con una leve normalización necesita ser la base para los mercados de datos. Por lo que no se crean los mercados de datos directamente desde los Sistemas de Procesamiento de Transacciones en Línea (OLTP) a través de un área de ensayo. En lugar de ello, se crean a partir de la arquitectura relacional de los datos corporativos.

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. Esta es eficaz y conduce a una solución completa en un corto período de tiempo. Se puede empezar de cero y dar al usuario una primera información sobre sus datos en cuestión de días. Además, tiene abundante documentación y se puede encontrar una respuesta a casi todas las preguntas que se puedan tener.

Como característica principal posee una arquitectura ascendente bottom-up. Plantea que se debe crear por cada departamento un conjunto de mercados de datos independientes orientados a los temas que estén relacionados con él.

Metodología a usar para el desarrollo del Almacén de datos

Como metodología a emplear en el trabajo tenemos la **Propuesta de Metodología para el desarrollo de Soluciones de Almacenes de Datos e Inteligencia de Negocio (DW&BI)**. Esta es una adaptación hecha por el Centro de Gestión de Datos (DATEC). La misma propone una metodología para este tipo de soluciones, adaptada a su proceso de desarrollo, el cual está basado en Líneas de Productos de Software, y a los lineamientos de calidad exigidos por la Universidad de Ciencias Informáticas, realizando un estudio del Expediente de Proyecto y de cada uno de los artefactos propuestos en el expediente oficial para establecer cuáles eran las adaptaciones que debían realizarse. Esta propuesta también buscaba entre sus retos adaptarse al modo de producción establecido por el departamento DATEC.

Características principales

- Crea los conceptos de hechos y dimensiones, lo que indudablemente es muy eficaz en el proceso de la toma de decisiones y proporciona mayor agilidad en el proceso de desarrollo.
- Propone ir construyendo el almacén de datos a través de la construcción de los mercados de datos departamentales, lo que constituye una estrategia buena y coincide con la división lógica de las empresas, entidades, organismos, etc.
- Existe abundante documentación sobre la misma, la respuesta a todas las dudas y preguntas que puedan surgir se pueden encontrar en la web, a través de los servicios que brindan el grupo creador de la metodología.
- Es una metodología madura y reconocida por el resto de la comunidad dedicada al tema. Tiene bien definida las etapas, actividades, artefactos y roles.

Como complemento a la misma y fortaleciendo la etapa del levantamiento de requisitos; se tomó lo planteado por Leopoldo Zenaido Zepeda Sánchez en su Tesis de Doctorado, orientando así el trabajo a los Casos de Uso y se logra estar más alineado con las tendencias y normas de la Universidad, teniendo en cuenta las características de la Universidad de las Ciencias Informáticas (UCI) y DATEC.

Ciclo de vida de la Metodología para el desarrollo:

- Estudio Preliminar o Planeación
- Requerimientos
- Arquitectura y Diseño
- Implementación
- Prueba
- Despliegue
- Soporte y Mantenimiento
- Gestión y Administración del Proyecto

1.10 Herramientas

1.10.1 Herramientas para el proceso de ETL

Las herramientas de desarrollo son la piedra angular en el proceso de desarrollo de un almacén de datos. De la selección correcta de ellas depende que se logre el éxito en la implementación del sistema.

Pentaho Data Integration 4.2.1

Para el desarrollo del proceso ETL se propone el Pentaho Data Integration en su versión 4.2 debido a que es una herramienta de código abierto, con una fácil lectura para los XML que recogen transformaciones, tareas programadas y un repositorio relacional de metadatos ETL. Aplicable a diversos tipos de bases de datos que permite la importación y exportación de datos de un formato a otro cualquiera además de su fortaleza ser extensible mediante pluggins.

PDI incluye cuatro herramientas: (23)

- **Spoon:** Herramienta gráfica que posibilita el diseño de las transformaciones y trabajos. Posee opciones para previsualizar y testear los elementos desarrollados. Es la principal herramienta de trabajo de PDI y con la que se construirán y validarán los procesos ETL.
- **Pan:** Es un motor de transformación de datos que realiza muchas funciones tales como lectura, manipulación, y escritura de datos hacia y desde varias fuentes de datos [23]. Herramienta que permite la ejecución de las transformaciones diseñadas en spoon (bien desde un fichero o desde el repositorio). Permite desde la línea de comandos preparar la ejecución mediante scripts.
- **Kitchen:** Similar a Pan, pero para ejecutar los trabajos diseñados por Spoon en xml o en un catálogo de BD.
- **Carte:** es un pequeño servidor web que permite la ejecución remota de transformaciones y trabajos.

1.10.2 Sistema Gestor de Base Datos.

Un Sistema Gestor de base datos (DBMS por sus siglas en inglés) está conformado por un conjunto de programas que permiten crear y dar soporte a una base de datos, asegurando su integridad, confidencialidad y seguridad. Ejemplos de gestores de bases de datos son:

- PostgreSQL
- SQLite

- DB2 Express-C
- Apache Derby
- Microsoft SQL Server
- Sybase ASE Express Edition
- Oracle Express Edition 10

PostgreSQL 9.1

- Es un sistema de gestión de bases de datos objeto-relacional, distribuido bajo licencia BSD y con su código fuente disponible libremente. Es el sistema de gestión de bases de datos de código abierto más potente del mercado.
- Utiliza un modelo cliente/servidor y usa multiprocesos en vez de multihilos para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto y el sistema continuará funcionando. (16)

Ventajas del PostgreSQL v 9.1

- Diseñado para ambientes de alto volumen.
- Multiplataforma
- Extensible.
- Ahorros considerables en costos de operación.
- Mejor soporte que los proveedores comerciales.
- Tamaño de base de dato ilimitado. (16)

1.10.3 Herramientas de Modelado

Las herramientas CASE (Ingeniería de Software Asistida por Ordenador) son de gran ayuda para el desarrollo de un software. Estas son un grupo de programas que utilizan las personas que intervienen en el desarrollo de un software, como es el caso de los diseñadores, desarrolladores, analistas, entre otros, durante las fases del desarrollo del software (análisis, diseño, implementación), para agilizar y facilitar el trabajo, ya que dichas herramientas

proveen de métodos, técnicas y utilidades que ayudan al perfeccionamiento del desarrollo de sistemas de información, de forma total o parcial.

Visual Paradigm for UML 8.0

Visual Paradigm para UML en su versión 8.0 es 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 UML ayuda a una más rápida construcción de aplicaciones de calidad, mejores y a un menor coste. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. La herramienta UML CASE también proporciona abundantes tutoriales de UML, demostraciones interactivas de UML y proyectos UML.(9)

1.10.4 Administrador de Base de datos

PgAdmin 1.14.1

Es uno de los más populares administradores de base de datos por sus características en el mundo. Esto se debe en gran medida a que es una aplicación de código abierto y multiplataforma. PgAdmin está diseñado para responder a las necesidades de todos los usuarios, desde la escritura de simples consultas SQL, hasta para crear bases de datos complejas. Su interfaz gráfica soporta todas las características de PostgreSQL y facilita la administración. La aplicación también incluye un resaltado de sintaxis SQL editor, un editor de código del lado del servidor, una gente de la programación de SQL/ lote/ Shell de trabajo, soporte para el motor de replicación Slony-Iy mucho más. La conexión con el servidor se puede hacer a través de TCP/ IPoUnix Domain Sockets y posee encriptado SSL para la seguridad. No necesita controladores adicionales para comunicarse con el servidor de bases de datos. PgAdmin es desarrollado por una comunidad de expertos de PostgreSQL en todo el mundo y está disponible en más de una docena de idiomas. Es Software Libre publicado bajo la licencia de PostgreSQL. (10)

1.10.5 Herramientas de Inteligencia de negocios

Mondrian OLAP Server 3.0.4: Es una de las herramientas más significativas de código abierto, la cual es una de las aplicaciones más importantes de la plataforma Pentaho Business Intelligence. Mondrian es un servidor OLAP código abierto que gestiona la comunicación entre una aplicación OLAP y la base de datos con los datos fuente. Es desarrollado en Java/Servlets/JSPs. 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 bases de

datos que soporten JDBC. Mondrian soporta el lenguaje Microsoft's Multidimensional Expressions (MDX). También soporta los APIs: Java OLAP (JOLAP) y XML para el análisis de aplicaciones programadas.

Pentaho BI Server 3.8: Esta herramienta es la encargada de orquestar todos los procesos de inteligencia de negocio. Configura el servidor de inteligencia de negocio (BI server) para poder realizar la explotación del ODS. Sobre esta plataforma se definen las áreas de interés que poseen los usuarios para la preparación de reportes, consultas dinámicas y la realización del análisis OLAP (Pentaho Corporation 2005).

Schema Workbench 3.2: Esta es una herramienta para el desarrollo del esquema del modelo estrella en XML desarrollada en Java. Este programa recientemente publicado (2007) entrega todas las facilidades para poder realizar el modelo lógico del cubo OLAP al cual se le realizarán las consultas. Este programa se conecta directamente con la base de datos para así poder diseñar los cubos OLAP que se requieren para que el usuario final pueda visualizar los indicadores. Luego, el archivo generado se utiliza para definir la estructura de cubo en Jasper. (Pentaho Corporation 2005)

Apache Tomcat 6.0: Es un servidor web y de aplicaciones, robusto y altamente eficiente, que gestiona solicitudes y respuestas HTTP. Es gratis, fácil de instalar y se puede ejecutar en máquinas con pocos recursos. Está implementado con tecnología Java bajo la licencia de Apache 2, por lo que funciona en cualquier sistema operativo que se encuentre la máquina virtual de Java lo que permite a los usuarios realizar reportes.

1.10.6 Herramienta para el perfilado de los datos

DataCleaner 1.5.4: es una aplicación Open Source para el análisis, perfilado, transformación y limpieza de datos. Estas actividades ayudan a administrar y controlar la calidad de los datos. Datos de alta calidad es la clave para hacer que los datos útiles y aplicables a cualquier empresa moderna. Data-Cleaner es la alternativa gratuita al software de gestión de datos maestros (MDM) metodologías, almacenamiento de datos (DW) proyectos, la investigación estadística, la preparación para la extracción, transformación y carga (ETL) y más actividades. (11)

Conclusiones del capítulo

Capítulo 1: Fundamentos teóricos de la Investigación

En el presente capítulo se abordaron los principales conceptos relacionados con los almacenes de datos y mercado de datos. Se plantean los conceptos fundamentales a dominar para dar solución al problema planteado. Después del estudio realizado se seleccionó la metodología de desarrollo a utilizar además de las herramientas adecuadas según las necesidades del problema a resolver.

Capítulo 2: Análisis y Diseño del Mercado de datos para Modulo Visor de Historias Clínicas del Sistema Integral de Atención Primaria de la salud.

Introducción

En este capítulo se describe el negocio, donde se realiza el análisis y diseño del MD, se identifican los requisitos funcionales, no funcionales y los de información, las necesidades de los usuarios, las reglas del negocio, se especifican los actores, se describen los CU. En el diseño se obtiene el modelo de datos. Además, se definen las dimensiones del MD y sus hechos, así como la matriz bus que los relaciona.

2.1 Tema de análisis

La identificación de los temas de análisis es definen importancia para el desarrollo del Mercado de datos. Las áreas de una entidad o empresa cuentan con un alto nivel de datos, que poseen objetivos o metas propias. Para la construcción de la propuesta se define como tema de análisis la información que se genera en las historias clínicas dentro de los centros hospitalarios.

2.1.1 Reglas del negocio

Las reglas del negocio describen políticas, normas, operaciones, definiciones y restricciones especiales que se deben cumplir en algunas actividades de los procesos diarios de la organización y son de vital importancia para alcanzar los objetivos misionales. Forman parte de transformaciones que se le deben realizar a ciertos datos para obtener a partir de ellos otros datos y como son presentados los mismos en el subsistema de visualización.

RN1-Los identificadores de los indicadores no pueden estar repetidos.

RN2-Los identificadores de los indicadores no pueden ser nulos.

RN3-Los identificadores de las dimensiones no pueden ser nulos.

RN4-En una misma dimensión no pueden existir identificadores repetidos.

RN5-Las medidas de todos los hechos deben poseer valores mayores o iguales a cero, nunca un número negativo.

RN6-En una misma dimensión no pueden existir códigos repetidos.

RN7-Las cadenas en los nombres de los indicadores no pueden ser alfanuméricas.

RN8-Los códigos deben ser alfanuméricos o numéricos.

RN9- Las fechas deben estar entre un rango real de vida del paciente.

2.1.2 Necesidades de los usuarios

Para el desarrollo del análisis en el proceso del negocio es necesario conocer lo que necesitan los usuarios, así como la implicación de los mismos en el proyecto, aumentando así las probabilidades de satisfacer sus necesidades. Luego de un análisis del negocio que tiene como objeto el Modulo Visor de Historias Clínicas del Sistema Integral de Atención Primaria de la Salud, se define como necesidad de los usuarios el análisis que estos necesiten hacer con los datos acumulados, dentro de dicho modulo. Teniéndose como base este análisis sobre el cual se derivaran los requisitos de información.

Requerimientos

2.1.3 Requisitos de información

Los requisitos de información son las principales informaciones que deben estar disponibles al realizar los análisis sobre los datos. Constituyen una entrada fundamental para el proceso de inteligencia del negocio y futuros reportes. Los requisitos de información identificados son los siguientes:

1. Cantidad de consultas especializadas dado un servicio, en un período de tiempo, la entidad que lo realiza y si tuvo solicitud de análisis de laboratorio.
2. Cantidad de Transfusiones efectuadas en un periodo de tiempo, el carácter de la transfusión y el componente sanguíneo.
3. Cantidad de emergencias en un periodo de tiempo, los niveles de gravedad, condición de los pacientes y el servicio de emergencia requerido.
4. Cantidad de estudios autopsia en un periodo de tiempo, tipo de muerte y el departamento solicitante.
5. Cantidad de hospitalizaciones efectuadas dado la especialidad por la cual se hospitaliza, en un periodo de tiempo y la evolución del paciente
6. Cantidad de intervenciones quirúrgicas efectuadas en un periodo de tiempo, tipo de intervención y la especialidad por la cual se interviene quirúrgicamente.
7. Cantidad de egresados en un periodo de tiempo y el tipo de egreso.

8. Cantidad de fallecidos en un periodo de tiempo, la edad promedio de los fallecidos y tipo de fallecidos.
9. Cantidad de partos realizados en un periodo de tiempo, la condición del niño al nacer, el tipo de parto realizado y el peso promedio de los recién nacidos.

2.1.4 Requisitos funcionales

Los requisitos funcionales son capacidades o condiciones que el sistema debe cumplir y su proceso de obtención es importante en el desarrollo del MD, ya que de la calidad de los mismos depende el éxito del producto. Para el desarrollo de los procesos de integración se identificaron los siguientes requisitos funcionales:

No	Requisitos funcionales	Descripción
1	Autenticar usuario	Permite autenticar un usuario en el sistema.
2	Adicionar rol	Permite adicionar un rol en el sistema.
3	Eliminar rol	Permite eliminar un rol en el sistema.
4	Visualizar rol	Permite visualizar un rol en el sistema.
5	Modificar rol	Permite modificar un rol existente en el sistema.
6	Adicionar usuario	Permite adicionar un usuario en el sistema.
7	Eliminar usuario	Permite eliminar un usuario existente en el sistema.
8	Visualizar usuario	Permite buscar un usuario en el sistema.
9	Modificar usuario	Permite modificar un usuario existente en el sistema.
10	Adicionar reporte	Permite adicionar un reporte en el sistema.
11	Visualizar reporte	Permite buscar un reporte en el sistema.
12	Eliminar reporte	Permite eliminar un reporte existente en el sistema.
13	Modificar reporte	Permite modificar un reporte existente en el sistema.
14	Extraer datos de la fuente	Permite extraer los datos de la fuente.

15	Realizar transformación de los datos	Permite realizar la transformación de los datos de la fuente.
16	Cargar los datos al mercado de datos.	Permite realizar la carga de los datos hacia el mercado de datos.
17	Personalizar reporte	Permite mostrar las propiedades de los reportes.
18	Exportar a otro formato	Permite exportar el reporte a otro formato (Pdf, Excel).

Tabla1: Requisitos funcionales

2.1.5 Requisitos no funcionales

Usabilidad

RNF 1. Cumplir con las pautas de diseño de las interfaces.

El sistema debe tener una interfaz gráfica uniforme a través del mismo incluyendo pantallas, menús y opciones.

RNF 2. Mostrar los mensajes, títulos y demás textos que aparezcan en la interfaz del sistema en idioma español.

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

RNF 3. Diseñar un reporte del mercado de datos de manera sencilla y ágil.

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

RNF 4. Navegar en los reportes del mercado de datos de manera ágil.

El usuario podrá realizar los cruces y agrupaciones de variables en los reportes del mercado de datos de manera dinámica en la misma pantalla de trabajo. Esto permite agilizar la navegabilidad de los usuarios en un reporte.

Confiabilidad

RNF 5. Garantizar la persistencia de la información.

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

RNF 6. Asegurar la disponibilidad del sistema

El sistema debe estar disponible durante el horario de trabajo. En caso de fallo, la recuperación del servicio no deberá ser de un período de tiempo muy prolongado.

Restricciones de diseño

RNF 7. Utilizar los lenguajes de programación definidos durante la investigación.

Como lenguaje dentro del sistema gestor de base de datos para la programación en el almacén de datos se utilizará PL/pgSQL. En la implementación de los procesos de integración de datos se utilizará el lenguaje JavaScript. También se hará uso del lenguaje MDX para realizar las consultas.

RNF 8. Utilizar el Sistema Gestor de Base de Datos definido durante la investigación.

El gestor de base de datos que se utilizará es PostgreSQL y como interfaz de administración de dicho gestor PgAdmin.

RNF 9. Utilizar la herramienta de integración de datos definida durante la investigación.

Para el proceso de integración de datos se usará la herramienta Pentaho Data Integration.

RNF 10. Utilizar las herramientas para la implementación de la capa de inteligencia de negocios definidas durante la investigación.

De la suite Pentaho, se usarán los siguientes componentes:

- **SchemaWorkbench:** herramienta gráfica que se utiliza para construir el esquema multidimensional que soportará la creación de los reportes multidimensionales.
- **Pentaho BI Server:** servidor que se encarga de visualizar los reportes, tableros de control digital, controlar el acceso a la información y unificar en una solución de inteligencia de negocios el uso de las demás herramientas que componen la suite.

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

Para el uso de las herramientas anteriores se requiere la instalación de la máquina virtual de java (Java Virtual Machine 6.0).

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

RNF 11. Confección de un manual de usuario.

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

Interfaz

RNF 12. Acceso al sistema.

El usuario deberá acceder a la aplicación mediante el protocolo HTTP, usando preferiblemente el navegador web Firefox 2.0 en adelante.

Interfaces de usuario

RNF 13. Garantizar una interfaz amigable al usuario.

El sistema debe tener una interfaz amigable y sencilla de utilizar, teniendo en cuenta que los usuarios finales no son personas adiestradas en el campo de la informática.

Hardware

RNF 14. Proporcionar características mínimas de hardware a las PC clientes.

Los PC clientes deben contar con los siguientes requerimientos de hardware:

- Sistema operativo Linux.
- 1 GB RAM.
- 1 Procesador 2.0 Hz

RNF 15. Proporcionar características mínimas de hardware a los servidores.

Los servidores deben contar con los siguientes requerimientos de hardware:

- Sistema operativo Linux.

- 1 GB RAM.
- 1 Procesador Intel Core 2 Duo.
- 80 GB de disco

2.1.6 Descripción de los actores del sistema

El sistema contará con tres actores:

1. Analista, quien será el encargado de analizar las necesidades de información.
2. Administrador de ETL, quien llevará a cabo las funciones ETL.
3. Administrador del sistema, que incluye las funciones del analista y además es el encargado de gestionar los usuarios, roles, permisos y reportes.

2.2 Diagrama de Caso de Uso del sistema

El diagrama de caso de uso del sistema es una representación de todos los actores del sistema, los casos de uso y las relaciones que existen entre ellos. Para la confección del diagrama, se agruparon los 25 requisitos funcionales del sistema en 16 casos de uso; y los 13 requisitos de información en 10 casos de uso de información.

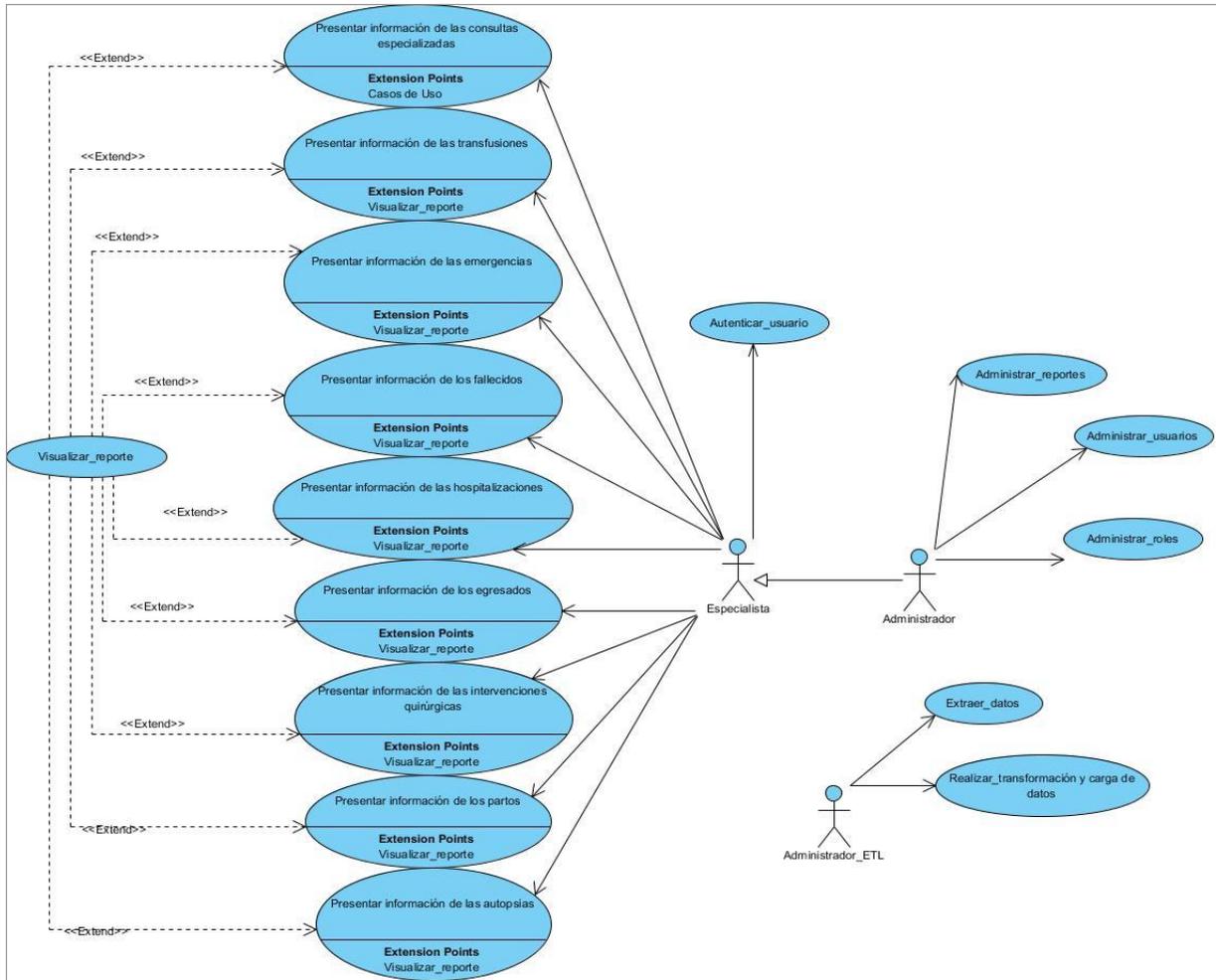


Figura 1. Diagrama de casos de uso

2.2.1 Descripción textual de los casos de usos

Se identificaron 16 casos de uso:

1. Autenticar usuarios: Realiza la autenticación de los usuarios en el sistema.
2. Administrar reportes: Elimina, inserta y modifica los reportes que se visualizan.
3. Administrar roles: Elimina, inserta y modifica los roles que interactúan con el sistema.
4. Administrar usuario: Gestiona los usuarios del sistema.
5. Realizar transformación y carga de datos: Realiza la carga y transformación de los datos.

6. Extraer datos: Realiza la extracción de los datos.
7. Presentar Información de las consultas especializadas: Analiza la información referente a las consultas especializadas en un rango de fecha, la especialidad de consulta requerida, el tipo de hospital que se desea analizar y el grupo etario deseado.
8. Presentar Información de las transfusiones: Analiza la información referente a las transfusiones realizadas en un rango de fecha, el carácter de la transfusión, el componente sanguíneo y el grupo etario deseado.
9. Presentar Información de las emergencias: Analiza la información referente a las emergencias en un rango de fecha, el nivel de gravedad de las emergencias atendidas, la condición de egreso de los pacientes atendidos de emergencia, el servicio de emergencia prestado, el tipo de hospital que atendió la emergencia y el grupo etario deseado.
10. Presentar Información de los fallecidos: Analiza la información referente a los fallecidos en un rango de fecha, el sexo de los fallecidos y el grupo etario deseado.
11. Presentar Información de las hospitalizaciones: Analiza la información referente a las hospitalizaciones en un rango de fecha, el servicio de hospitalización prestado, la evolución del paciente los pacientes hospitalizados y el grupo etario deseado.
12. Presentar Información de los egresados: Analiza la información referente a los egresados en un rango de fecha, el tipo de egreso y el grupo etario deseado.
13. Presentar Información de las intervenciones quirúrgicas: Analiza la información referente a las intervenciones quirúrgicas en un rango de fecha, el servicio de intervención quirúrgica prestado, el tipo de intervención quirúrgica y el grupo etario deseado.
14. Presentar Información de los partos: Analiza la información referente a los partos realizados en un rango de fecha, el tipo de parto realizado, la condición de los recién nacidos, el peso de los mismo, el sexo y el grupo etario deseado.
15. Presentar Información de las autopsias: Analiza la información referente a las autopsias realizadas en un rango de fecha y el departamento solicitante.

16. Visualizar Reporte: Visualiza los reportes.

2.3 Arquitectura del Mercado de datos

El proceso de desarrollo de un software requiere de la definición previa de una arquitectura. Esta representa la guía para el diseño y construcción de un sistema. La arquitectura de este proceso está conformada por algunos elementos necesarios para la implementación de la solución.

Primeramente se extraen los datos de los sistemas fuentes que en este caso se encuentran en una base de datos, posteriormente los datos son modificados o transformados al formato correspondiente según las necesidades de información del cliente y una vez limpia la información se carga al almacén de donde a través de una interfaz gráfica el usuario puede acceder a la información visualizando los reportes.

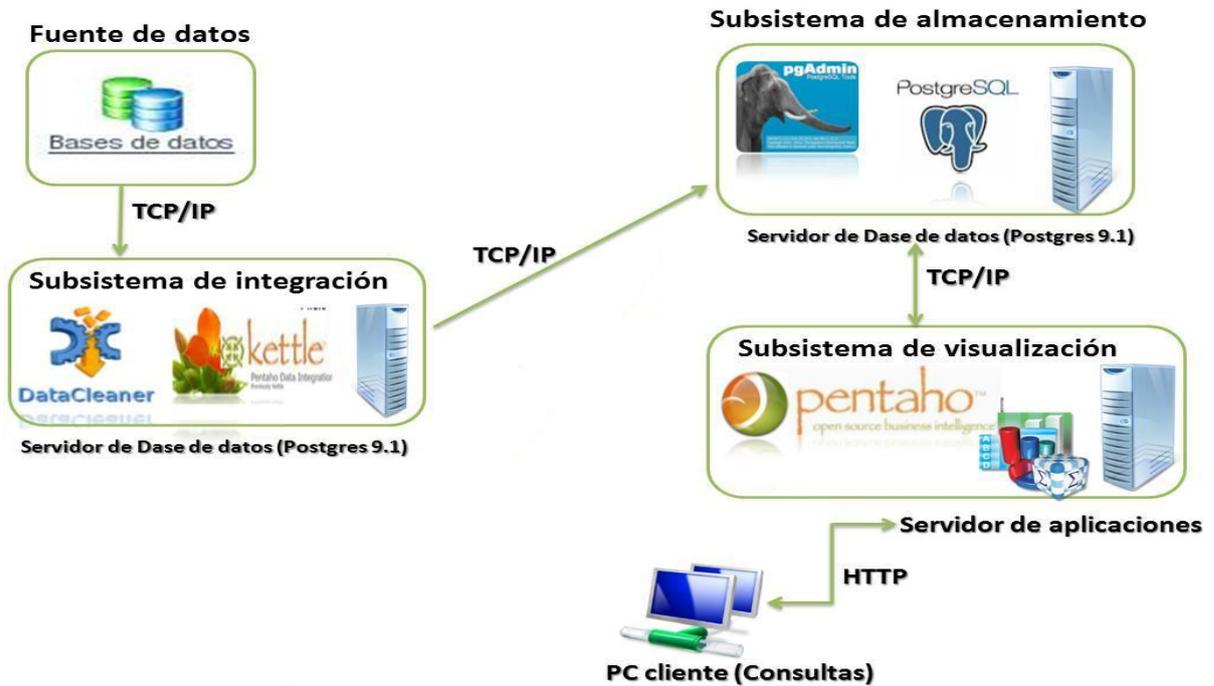


Figura 2: Diseño de la arquitectura de un Almacén de datos

2.3.1 Hechos

Se definieron 9 tablas de hechos:

Hechos	Descripción
hech_autopsias_realizadas	Este hecho encierra toda la información referente a las autopsias realizadas
hech_consultas_especializadas	Este hecho contiene toda la información referente a las consultas especializadas
hech_emergencias	Este hecho encierra toda la información referente a las emergencias atendidas
hech_egresados	Este hecho encierra toda la información referente a los egresados
hech_hospitalizaciones	Este hecho contiene toda la información referente a las hospitalizaciones
hech_fallecidos	Este hecho encierra toda la información referente los fallecidos

hech_partos_realizados	Este hecho encierra toda la información referente a los partos realizados
hech_transfusiones	Este hecho encierra toda la información referente a las transfusiones
hech_intervenciones_quirúrgicas	Este hecho contiene toda la información referente a las intervenciones quirúrgicas.

Tabla 2: Tabla de hechos

2.3.2 Dimensiones

Se definieron las siguientes tablas de dimensiones:

Dimensiones	Descripción
dim_rango_fecha	Esta es la dimensión más común e importante en los diseños de un Mercado de datos ya que define una línea de tiempo específica.
dim_especialidad_consulta	Esta dimensión especifica los tipos de especialidad de consulta existentes.
dim_niveles_gravedad	Esta dimensión especifica los niveles de gravedad.
dim_tipo_hospital	Esta dimensión describe los tipos de hospitales
dim_servicio_emergencia	Esta dimensión especifica los tipos de servicios de emergencias posibles.
dim_departamento_solicitante	Esta dimensión especifica el nombre del que solicita una autopsia.
dimsexo	Esta dimensión especifica el sexo de los pacientes
dim_condicion_nino	Esta dimensión expone la condición de los niños al nacer.
dim_servicio_intervencion	Esta dimensión expone el nombre de todos los servicios de intervenciones quirúrgicas que se prestan.
dim_tipo_intervencion	Esta dimensión especifica los tipos de intervenciones quirúrgicas
dim_tipo_egreso	Esta dimensión especifica los tipos de egreso.

dim_servicio_hospitalizacion	Esta dimensión especifica los distintos tipos servicios de hospitalización.
dim_caracter_transfucion	Esta dimensión especifica el carácter de las transfusiones.
dim_peso	Esta dimensión agrupa los pesos de los recién nacidos y define qué tipo de recién nacido es atendiendo al peso al nacer.
dim_edad	Esta dimensión especifica los grupos de edades según la clasificación dada en el negocio.
dim_evolucion_paciente_hospitalizado	Esta dimensión especifica la evolución de los pacientes que han sido hospitalizados.
dim_componente_sanquineo	Esta dimensión especifica todos los componentes sanguíneos de las transfusiones realizadas
dim_condicion_egreso	Esta dimensión especifica la condición con la que egresan los pacientes al hospital
dim_tipo_parto	Esta dimensión especifica los tipos de partos que se realizan

Tabla 3: Tabla de dimensiones

2.4 Matriz BUS

La Matriz Bus es la representación de la relación que existe entre los hechos y las dimensiones. 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.

Dimensiones/Hechos	hech_transfusiones	hech_egresados	hech_hospitalizacion	hech_fallecidos	hech_intervencion_quirurgica	hech_partos_realizados	hech_autopsias_realizadas	hech_emergencias	hech_consulta_especializada
dim_rango_fecha	X	X	X	X	X	X	X	X	X
dim_tipo_hospital								X	X
dim_condicion_egreso								X	
dim_especialidad_consulta									X
dim_condicion_ninno						X			
dim_servicio_intervencion					X				
dim_tipo_intervenciones					X				
dim_tipo_egreso		X							
dim_servicio_hospitalizacion			X						
dim_caracter_transfusion	X								
dim_tipos parto						X			
dim_componente_sanguineo	X								
dim_evolucion_paciente_hospitalizado			X						
dim_edad				X					
dim_clasificacion_segun_peso						X			
dimsexo						X			
dim_departamento_solicitante							X		
dim_servicio_emergencia								X	
dim_niveles de gravedad								X	

Figura 3: Matriz Buz

2.5 Modelo de Datos

El modelo de datos está conformado por las dimensiones, medidas, y hechos identificados para el área de Ciencia y técnica. Por las necesidades del negocio existe un modelo que relaciona las dimensiones definidas y las medidas que se han detallado hasta el momento. En él se ve la relación que existe entre las dimensiones y las tablas de hechos.

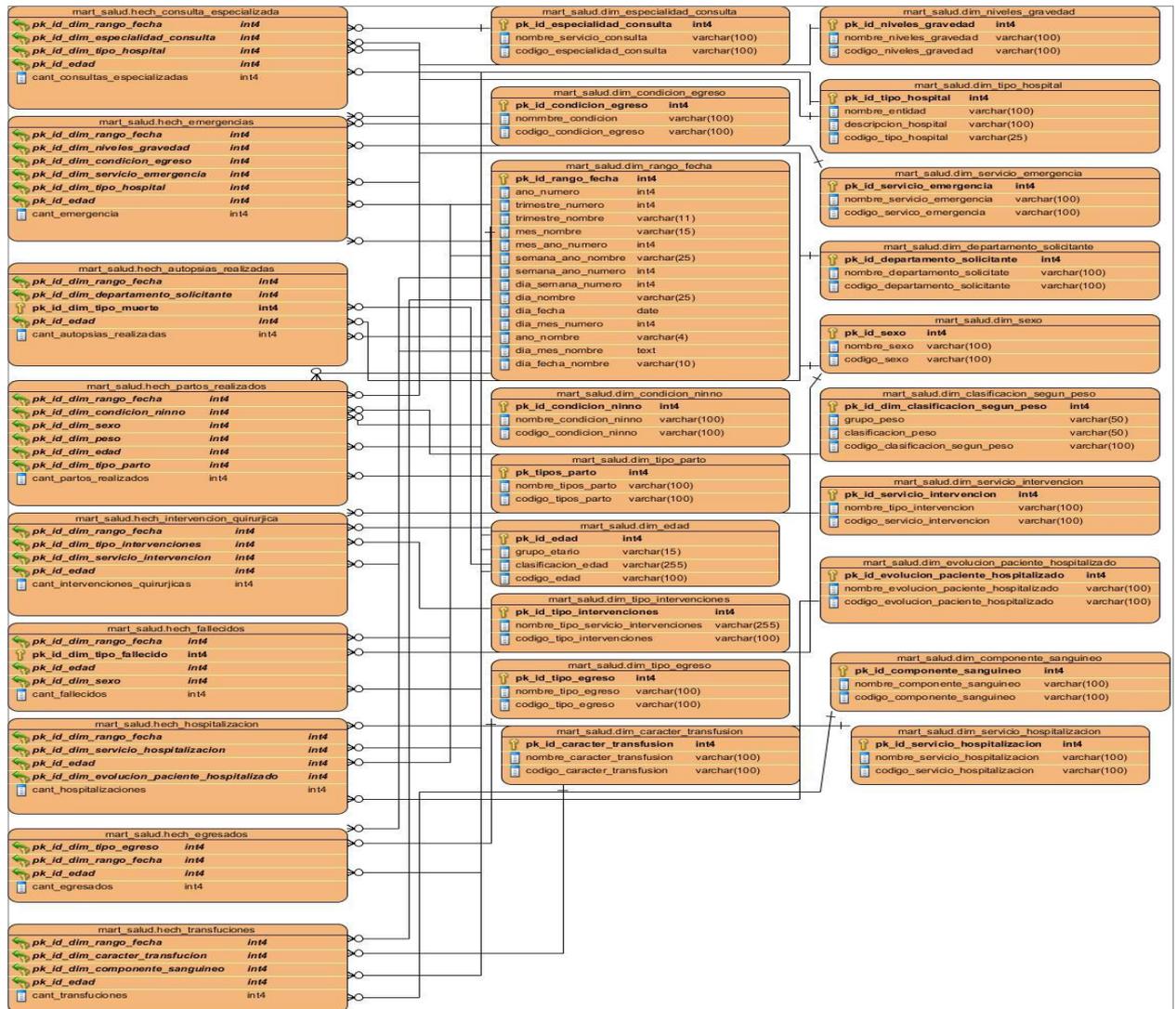


Figura 4. Modelo de datos

2.6 Subsistema de integración

Como parte de la capa de integración se realiza el diseño de las transformaciones el cual puede variar a la hora de hacer la extracción, transformación y la carga de los datos.

Diseño de las transformaciones

Diseño para la carga de las dimensiones: Los datos se cargan de la fuente de datos, luego se seleccionan los indicadores y se adicionan a la base de datos. La figura muestra el diseño de la transformación para la carga de las dimensiones.

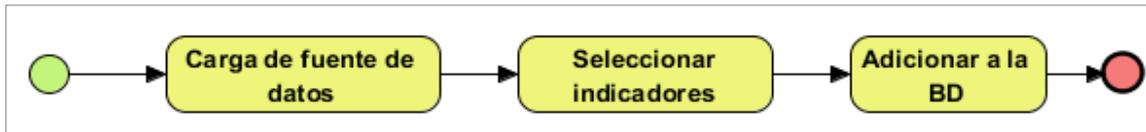


Figura 5. Diseño de las transformaciones para la carga de las dimensiones

Diseño para la carga de los hechos: Una vez ejecutada dicha transformación, se valida un grupo de restricciones definidas por el analista. En caso de no cumplirse una de estas restricciones, los datos se almacenan en un fichero Excel para darle tratamiento; en caso contrario, todos los datos se insertan en la tabla que tiene como salida dicha transformación. La figura muestra el diseño de la transformación para la carga de las dimensiones.

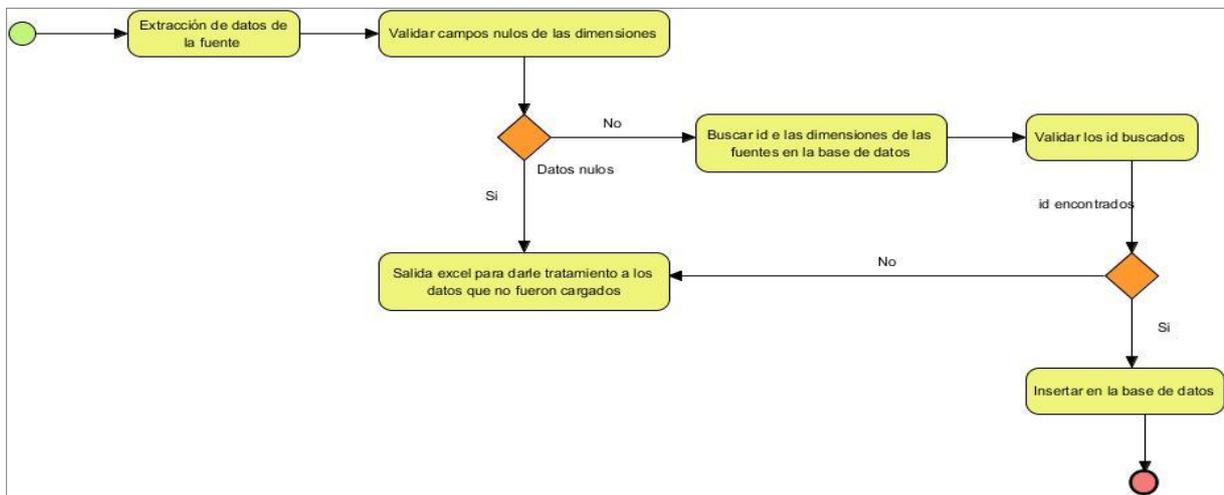


Figura 6. Diseño de las transformaciones para la carga de los hechos.

2.7 Roles y permisos

Para el acceso al Mercado de datos se define un usuario por cada uno de los roles existentes en el sistema, con el objetivo de garantizar que cada usuario opera en el sistema según los permisos que se le definan al rol (Ver tabla 1).

Roles	Permisos	
	Lectura	Escritura
Sobre la Base de Datos		
Administrador	X	..
Administrador ETL	X	X

Especialista	X	..
Sobre la aplicación		
Administrador	X	X
Administrador ETL	X	..
Especialista	X	..

Tabla 4: Roles y permisos

2.8 Salva de la base de datos

Para garantizar la persistencia de la información se realizará un respaldo total de los datos del Mercado, la política de respaldo y recuperación que se empleará está condicionada por tres puntos fundamentales:

- **Periodicidad de salvas del sistema:** Estas se realizarán mensual pero se brinda la posibilidad de que puedan ser quincenal, trimestral, semestral y anual. Además de que se verifica que exista una copia de toda la información almacenada.
- **Tablas involucradas:** Todas las tablas que forman el Mercado de datos ya sean las tablas de hechos y las Tablas de dimensiones.
- **Salvas existentes:** Las salvas se realizan con el mismo período que es recopilada la información.

Conclusiones del capítulo

Se identificaron 20 requisitos de información agrupados por tema de análisis, 26 funcionales y 16 no funcionales del sistema, los cuales fueron agrupados en nueve casos de uso para la elaboración del diagrama de caso de uso del sistema.

Se confeccionó el modelo de datos dimensional donde se refleja la relación existente entre las nueve tablas de hechos y las diecinueve tablas de dimensiones. Se describió el diseño de cada uno de los subsistemas que conforman el Mercado de datos. Se establecieron las políticas de seguridad que deben cumplir los empleados de la institución.

Capítulo 3: Implementación del Mercado de datos Visor de Historias Clínicas del Sistema Integral de Atención Primaria de la Salud.

Introducción: En este capítulo se procederá a la implementación de la solución propuesta, se implementará el modelo de datos, se poblará la base de datos a través del *script* del mercado, se realizarán las transformaciones y la carga de los datos. Además de que se implementaran y serán visualizados los reportes candidatos.

3.1 Implementación del modelo de datos

Para una correcta organización del Mercado de datos, se implementaron las tablas del modelo de datos en un esquema: “mart_salud” el cual contiene todas las tablas de hechos y dimensiones pertenecientes al Mercado de datos.

Esquema	Tablas
mart_salud	dim_rango_fecha
mart_salud	dim_caracter_transfusión
mart_salud	dim_clasificación_según_peso
mart_salud	dim_componente_sanguíneo
mart_salud	dim_condición_egreso
mart_salud	dim_condición_ninno
mart_salud	dim_departamento_solicitante
mart_salud	dim_edad
mart_salud	dim_especialidad_consulta
mart_salud	dim_evolución_paciente_hospitalizado
mart_salud	dim_niveles_gravedad
mart_salud	dim_servicio_emergencia
mart_salud	dim_servicio_hospitalización
mart_salud	dim_servicio_intervención
mart_salud	dimsexo
mart_salud	dim_tipo_egreso
mart_salud	dim_tipo_hospital
mart_salud	dim_tipo_intervenciones

mart_salud	dim_tipo_parto
mart_salud	hech_autopsias_realizadas
mart_salud	hech_consulta_especializada
mart_salud	hech_egresados
mart_salud	hech_emergencias
mart_salud	hech_fallecidos
mart_salud	hech_hospitalizacion
mart_salud	hech_intervencion_quirurgica
mart_salud	hech_partos_realizados
mart_salud	hech_transfusiones

Tabla 5: Esquemas y tablas del mercado

3.2. Perfilado de los datos

El perfilado de datos se realiza con el objetivo de conocer el estado en que se encuentran los datos de la fuente origen y así puede supervisar la calidad de los mismos con el fin de garantizar que la información es útil y aplicable a una situación de negocio. Implementación de las transformaciones

3.3. Transformaciones para la carga de las tablas de dimensiones

Para poblar las tablas de dimensiones se validaron los datos acorde a las reglas de negocio definidas para el proceso extracción, carga y transformación de los datos. Después se procedió a insertar los datos en las tablas de dimensiones. A continuación se muestran una transformación de la carga de la dimensión carácter de transfusión. (Ver Figura 7)

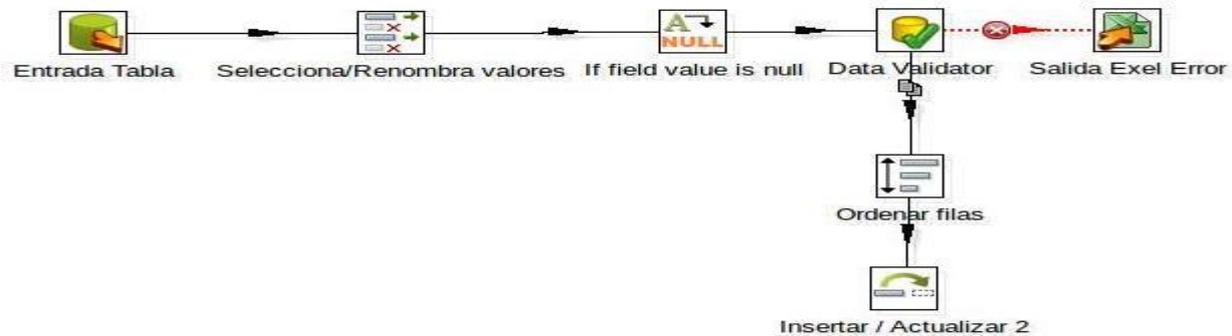


Figura 7. Transformación para la carga de la dimensión carácter_transfucion

3.4. Transformación para cargar el hecho

Para cargar los datos en las tablas de hecho se validó que los mismos estuvieran de acorde a las reglas de negocio definidas para el proceso extracción, carga y transformación, después se procedió a insertar los datos en las tablas de dimensiones. A continuación se muestran una transformación de la carga del hecho emergencia (Ver Figura 8)

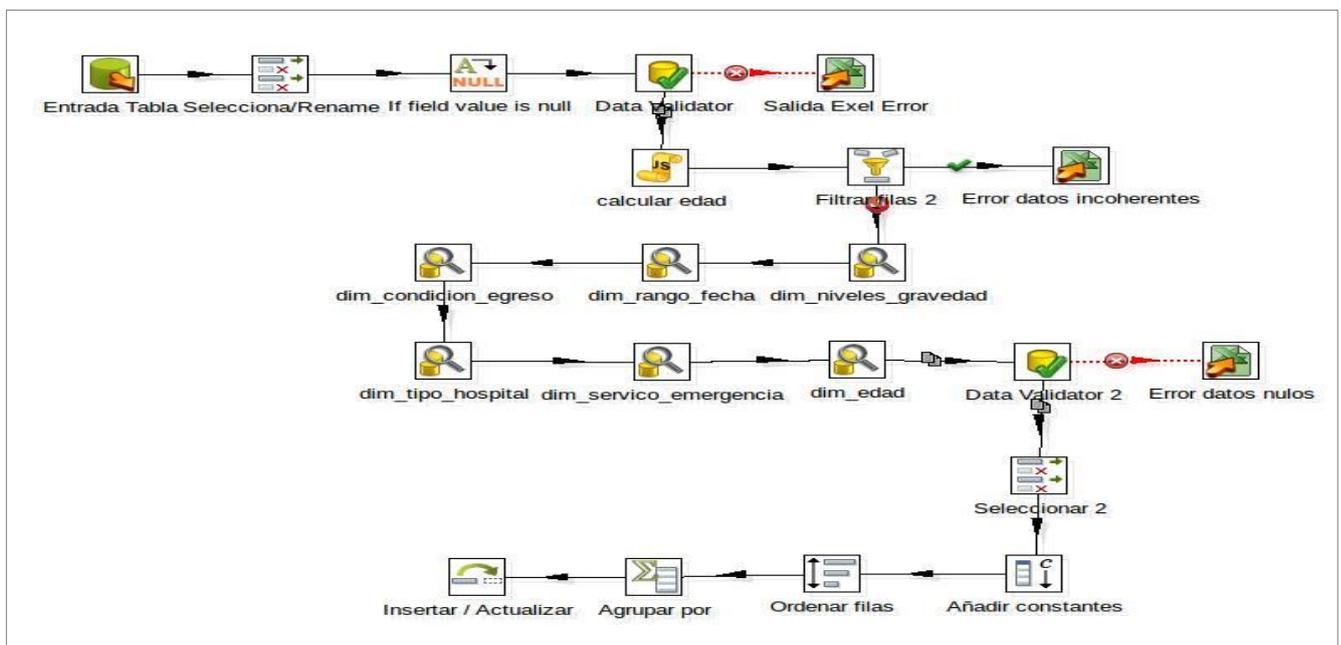


Figura 8. Transformación para la carga del hecho emergencia

3.5. Implementación de los trabajos

Con el propósito de definir una secuencia lógica para la ejecución de las transformaciones se crea un trabajo principal para la verificación de las conexiones hacia a la fuente y el Mercado de datos, posteriormente debido a la dependencia que poseen las tablas de hechos con las tablas de dimensiones, se hace necesario ejecutar primero las transformaciones correspondientes a las dimensiones y posteriormente las relacionadas con los hechos. (Ver figura 9)

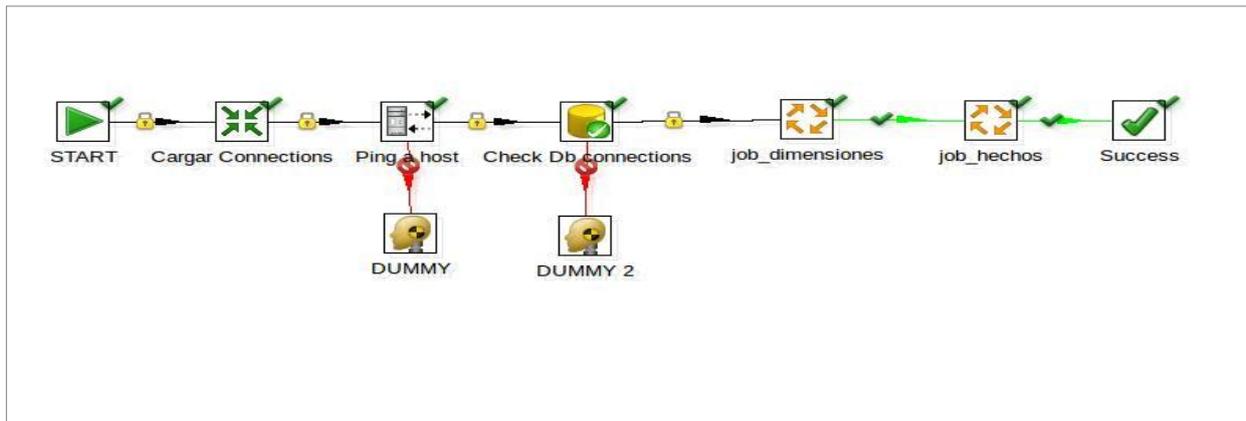


Figura 9. Transformación del trabajo para la carga de los hechos

3.6. Implementación del subsistema de visualización

3.6.1. Implementación de los cubos OLAP

Para la implementación de los cubos, se definieron las dimensiones, las medidas y los niveles de jerarquía de cada dimensión. Se creó el esquema Salud y se implementó un cubo por cada una de las 9 tablas de hechos incluyendo dentro de cada cubo la dimensión relacionada con el mismo (Ver figura 10).

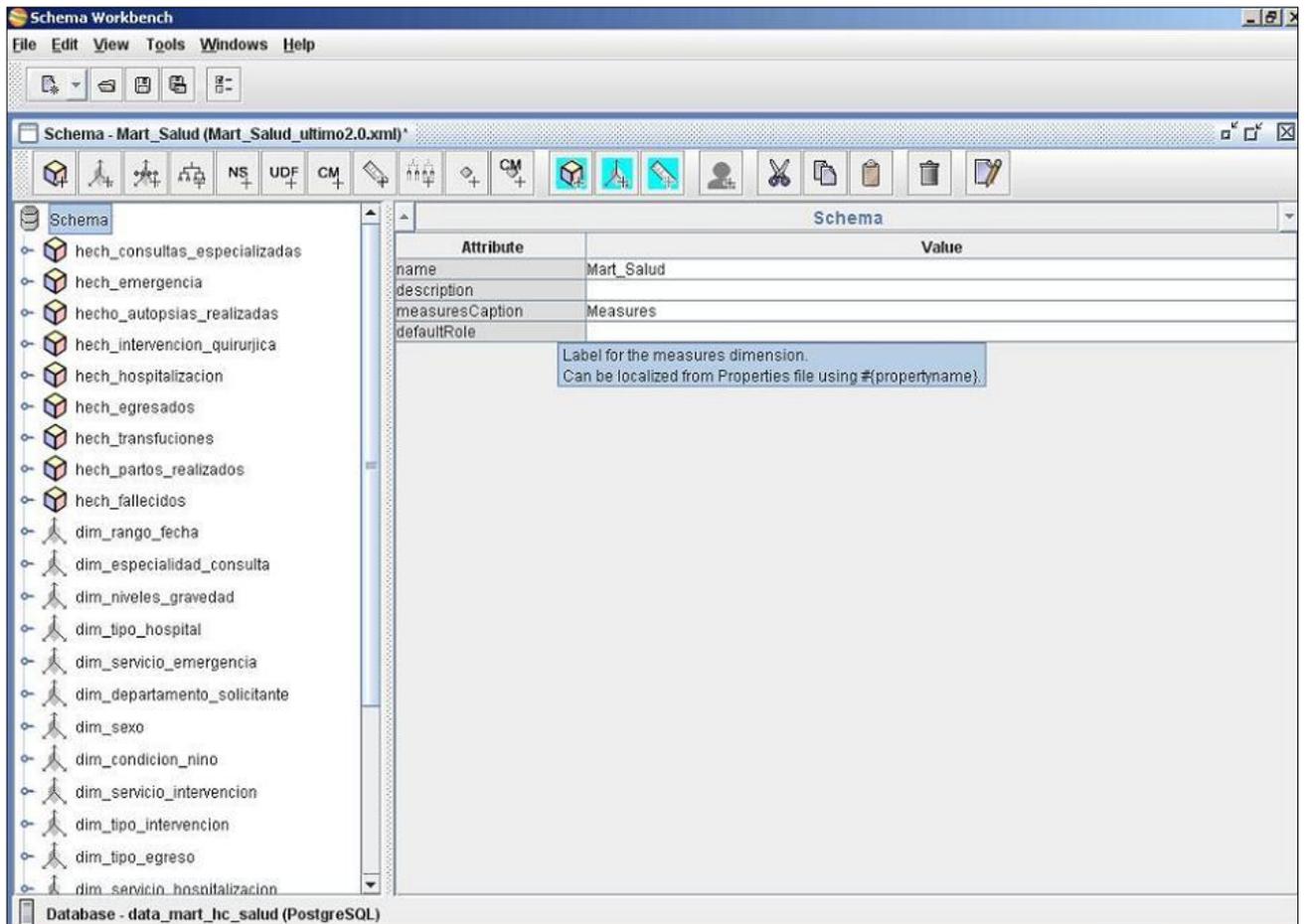


Figura 10. Diseño de los cubos OLAP

A continuación se muestra la estructura del hecho `hech_partos_realizados`, que contiene la cantidad de partos realizados, los cuales se pueden analizar desde varias perspectivas de análisis, como se muestra en la siguiente figura:



Figura 11: Diseño del cubo `partos_realizados`

3.7 Navegación de la capa de visualización

La estructura de navegación de la capa de visualización está organizada de la siguiente manera: un área de análisis general, un área de análisis específica para el área de salud y nueve libros de trabajo incluidos en esta última área de análisis. (Ver Figura 12).



Figura 12. Diseño del mapa de navegación

Descripción del Área General (A.A.G)

A.A.G Sistema integral de atención primaria de la salud (SIAPS): Agrupa toda la información de todos los Mercados de datos perteneciente al Sistema Integral de Atención Primaria de la Salud.

Descripción del Área de Análisis (A.A)

A.A Visor de historias clínicas: Agrupa toda la información referente a los indicadores de las historias clínicas y contiene los libros de trabajo donde se encuentran los reportes.

Descripción de los Libros de Trabajo (L.T)

L.T Autopsias realizadas: agrupa la información perteneciente a las autopsias realizadas, contiene cuatro reportes que permiten realizar un análisis general de los datos.

L.T Consultas especializadas: agrupa la información referente a las consultas especializadas, contiene doce reportes que permiten realizar un análisis general de los datos.

L.T Egresados: agrupa la información perteneciente a los egresados atendidos, contiene doce reportes que permiten realizar un análisis general de los datos.

L.T Emergencias: agrupa la información referente a las emergencias atendidas, contiene ocho reportes que permiten realizar un análisis general de los datos.

L.T Fallecidos: agrupa la información perteneciente a los fallecidos, contiene cuatro reportes que permiten realizar un análisis detallado de los datos.

L.T Hospitalizaciones: agrupa la información referente a las hospitalizaciones, contiene doce reportes que permiten realizar un análisis detallado de los datos.

L.T Intervenciones quirúrgicas: agrupa la información perteneciente a las intervenciones quirúrgicas, contiene doce reportes que permiten realizar un análisis detallado de los datos.

L.T Partos Realizados: agrupa la información referente a los partos realizados, contiene diez reportes que permiten realizar un análisis detallado de los datos.

L.T Transfusiones: agrupa la información referente a las transfusiones realizadas, contiene cuatro reportes que permiten realizar un análisis detallado de los datos.

3.8 Creación y administración de los reportes

En la implementación de los reportes se utilizaron consultas mdx, las cuales son administradas por el Pentaho BI Server. Se crearon 72 reportes en total. A continuación se muestra el reporte “Autopsias realizadas en el año 2009”, el cual devuelve la estructura de la cantidad de autopsias realizadas en ese período, los departamentos que solicitaron dichas autopsias, se muestra además el entorno de trabajo de la herramienta y como se muestran los datos después de ejecutado el reporte.

Rango de Fecha	Medidas	Departamento solicitante			
		Departamentos	Apoyo Diagnóstico y tratamiento	Estudios especiales	Emergencia y Medicina Crítica
2009	Cantidad de autopsias realizadas	88	10	16	14
Trimestre 1	Cantidad de autopsias realizadas	10		2	2
Trimestre 2	Cantidad de autopsias realizadas	33	4	4	6
Trimestre 3	Cantidad de autopsias realizadas	23	4	5	4
Trimestre 4	Cantidad de autopsias realizadas	22	2	5	2

Figura 13: Reporte Cantidad de autopsias realizadas en el año 2009

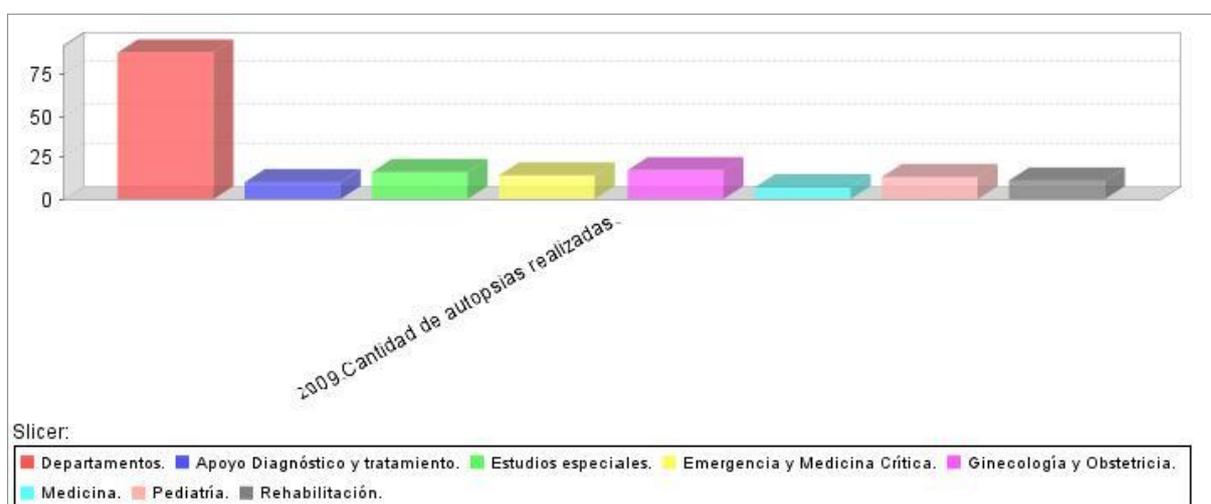


Figura 14: Gráfico del reporte Autopsias realizadas en el año 2009.

Conclusiones del capítulo

En el presente capítulo se describieron los elementos de implementación para la construcción del Mercado de datos para el modulo Visor de Historias Clínicas del Sistema Integral de atención primaria de la Salud. Quedó definida la estructura de los datos a partir del modelo físico, contando con un esquema mart_salud, se realizó el proceso ETL, se implementaron los cubos OLAP, quedando definidos nueve cubos y diecinueve dimensiones, se desarrolló el subsistema de visualización determinándose nueve libros de trabajos y finalmente se implementaron todos los requisitos de información.

Capítulo 4: Pruebas del Mercado de datos para el modulo Visor de Historias Clínicas del Sistema Integral de atención Primaria de la Salud.

Introducción

Luego de haber concluido con el análisis, diseño e implementación del Mercado de datos para el modulo Visor de Historias Clínicas del Sistema Integral de atención Primaria de la Salud, se dio paso a la validación y prueba de la solución mediante las listas de chequeo y los casos de prueba, para así verificar que el sistema cumpla con los requerimientos necesarios, que garanticen al usuario final del sistema la confiabilidad de los datos cargados en el almacén de datos.

4.1 Pruebas

El modelo V es un método para comprobar el buen funcionamiento de los almacenes de datos, proviene del principio que establece que los procedimientos utilizados para probar si la aplicación cumple las especificaciones ya deben haberse creado en la fase de diseño. Es una representación de dos cascadas enfrentadas y relacionadas, con su vértice en la codificación como punto en común. Propone una cascada a la izquierda, con las actividades relacionadas al desarrollo y una a la derecha con las actividades del aseguramiento de la calidad del software. Mediante este modelo se describe a un nivel muy alto de abstracción las fases del ciclo de desarrollo en las que se involucra la prueba. (25)

Este modelo cuenta con los siguientes tipos de prueba:

La prueba unitaria es la prueba que valida que las unidades individuales está trabajando correctamente (26).

La pruebas de integración que se realiza después de las pruebas unitarias y antes de las pruebas del sistema. Su objetivo es identificar defectos en las interfaces y las interacciones entre los componentes del sistema, ya que se unen para formar grandes subsistemas progresivamente. Dado que los componentes del sistema son identificados y diseñados durante el diseño físico, la mayoría de la información utilizada en la planificación de pruebas de integración deriva del diseño físico (28).

La prueba de sistema es una técnica para verificar si un sistema completo cumple con las especificaciones acordadas. El propósito de la prueba de sistema es detectar discrepancias entre el comportamiento del sistema construido y su especificación (29).

Pruebas de aceptación: son pruebas funcionales, pero vistas directamente desde el cliente. Son aquellas pruebas que demuestran al cliente que la funcionalidad está terminada y funciona correctamente.

4.2 Herramientas de pruebas

Listas de chequeo

Es un listado de preguntas, en forma de cuestionario que sirve para verificar el grado de cumplimiento de determinadas reglas establecidas, con un fin determinado y se deben tener en cuenta desde el inicio del proyecto. (24)

Ventajas del uso de las listas de chequeo.

- Ayudan a organizar y entender la documentación de los proyectos.
- Sirven de guía para la elaboración del documento de los proyectos.
- Respaldan cada paso que se da en el desarrollo de cualquier proyecto.

La lista de chequeo contiene diferentes indicadores a evaluar los cuales se encuentran distribuidos en tres secciones fundamentales:

- **Estructura del documento:** abarca todos los aspectos definidos por el expediente de proyecto o el formato establecido por el proyecto.
- **Indicadores definidos:** abarca todos los indicadores a evaluar durante la etapa.
- **Semántica del documento:** contempla todos los indicadores a evaluar respecto a la ortografía, redacción y demás.

La lista de chequeo que se utilizó fue:

Estructura del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios

crítico	1. ¿El entregable contiene las secciones obligatorias de la plantilla estándar definida para el expediente de proyecto?	0			
crítico	2. ¿El alcance del proyecto describe correctamente los datos de las dimensiones y hechos del mercado de datos?	0			
crítico	3. ¿El objetivo expresa correctamente el propósito del documento?	0			
	4. ¿Se hace un uso adecuado del control del documento?	0			
	5. ¿En la sección de acrónimos se definen todos los acrónimos utilizados en el documento?	0			
	6. ¿En el entregable, la definición de las variables se hace correctamente?	0			
	7. ¿Existe una adecuada correspondencia entre las variables definidas y las descripciones que tienen estas variables?	0			
	8. ¿En el entregable se crea una hoja por cada variable definida?	0			
	9. ¿Queda registrado en el entregable todos los posibles valores que van a tener las variables definidas?	0			

Indicadores definidos en el desarrollo

Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
	1. ¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?	0			
Semántica del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
crítico	1. ¿Se ha identificado errores ortográficos en el entregable?	0			
crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?	0			
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?		NP		

4.2. Caso de Prueba

El caso de prueba específica la forma de probar un sistema incluyendo las entradas, salidas y resultados esperados, así como bajo qué condiciones debe probarse el sistema. Se realizó un caso de prueba por caso de uso de información. (27)

Ejemplo para el caso de uso de información “Cantidad de autopsias realizadas dando un rango de fecha y el departamento solicitante”..

Escenario	Descripción	Variables de Entrada	Variables de Salida	Respuesta del sistema	Flujo central
EC 1.1 Autopsias	Muestra la cantidad de	dim_departamento_solicitante	cant:__autopsias_realizadas	El sistema muestra todas	Se autentifica Se entra al sistema

realizadas en el 2009	autopsias realizadas por el departament o que la solicita y el año.	dim_rango_fecha		las variables disponibles para el análisis, ubicados en las filas y las columnas que pueden ser visualizadas para cada reporte	Se selecciona el A.A Salud Se selecciona el L.T Autopsias Realizadas En la parte inferior izquierda se selecciona el reporte deseado En el área de trabajo se visualiza el reporte
Escenario	Descripción	Variables de Entrada	Variables de Salida	Respuesta del sistema	Flujo central
EC 1.2 Autopsias realizadas en el 2010	Muestra la cantidad de autopsias realizadas por el departament o que la solicita y el año.	dim_departamento_solicitante dim_rango_fecha	cant:__autopsias_realizadas	El sistema muestra todas las variables disponibles para el análisis, ubicados en las filas y las columnas que pueden ser visualizadas para cada reporte	Se autentifica Se entra al sistema Se selecciona el A.A Salud Se selecciona el L.T Autopsias Realizadas En la parte inferior izquierda se selecciona el reporte deseado En el área de trabajo se visualiza el reporte
Escenario	Descripción	Variables de Entrada	Variables de Salida	Respuesta del sistema	Flujo central

EC 1.3 Autopsias realizadas en el 2011	Muestra la cantidad de autopsias realizadas por el departamento o que la solicita y el año.	dim_departamento_solicitante	cant:__autopsias_realizadas	El sistema muestra todas las variables disponibles para el análisis, ubicados en las filas y las columnas que pueden ser visualizadas para cada reporte	Se autentifica Se entra al sistema Se selecciona el A.A Salud Se selecciona el L.T Autopsias Realizadas En la parte inferior izquierda se selecciona el reporte deseado En el área de trabajo se visualiza el reporte
Escenario	Descripción	Variables de Entrada	Variables de Salida	Respuesta del sistema	Flujo central
EC 1.4 Autopsias realizadas en el 2011	Muestra la cantidad de autopsias realizadas por el departamento o que la solicita y el año.	dim_departamento_solicitante dim_rango_fecha	cant:__autopsias_realizadas	El sistema muestra todas las variables disponibles para el análisis, ubicados en las filas y las columnas que pueden ser visualizadas para cada reporte	Se autentifica Se entra al sistema Se selecciona el A.A Salud Se selecciona el L.T Autopsias Realizadas En la parte inferior izquierda se selecciona el reporte deseado En el área de trabajo se visualiza el reporte

Conclusiones del capítulo

En este capítulo se realizó las pruebas al Mercado de datos Módulo Visor de Historias Clínicas. Se describió el desarrollo y la aplicación de una lista de chequeo y de un caso de prueba. Se evaluaron los resultados de los procesos de análisis e integración de datos. La evaluación final fue satisfactoria, resultado que demuestra la calidad del análisis e integración de datos realizado.

Conclusiones Generales

Al culminar el desarrollo de la presente investigación se llegaron a las siguientes conclusiones generales:

- La correcta selección de la metodología, herramientas y tecnologías utilizadas, permitieron el desarrollo de todo el proceso de construcción de la solución.
- Se realizó el análisis y diseño del mercado de datos, identificándose los requisitos que dieron solución a las necesidades de los usuarios.
- La implementación de los subsistemas de integración y de visualización permitieron obtener como resultado un mercado de datos poblado y funcional, con información disponible para ser consultada por parte de los usuarios, apoyando el proceso de toma de decisiones.
- Se realizaron distintos tipos de pruebas que permitieron verificar la calidad del producto a diferentes niveles, arrojándose algunas no conformidades que fueron solucionadas.

Recomendaciones

Para el presente trabajo de diploma se recomienda:

- Realizar un profundo estudio acerca de técnicas de optimización, que puedan ser aplicadas al proceso de extracción, transformación y carga e Inteligencia de Negocios desarrollado.
- Realizar la integración de otras fuentes de datos que complementen la información del Mercado de datos.

Referencias Bibliográficas

1. **Domínguez, Antonio Blasco.** Programacion. *Data Warehousing*. [En línea] [Citado el: 26 de noviembre de 2011.] http://www.programacion.com/articulo/data_warehousing_201.
2. **Bressán, Griselda E.** *Objetivos del Data Warehouse*. [En línea] <http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MineriaDatosBressan.htm#Objetivos%20del%20Data%20Warehouse..>
3. **Inmon, W. H.** Building the Data Warehouse. s.l. : QED Press/John Wiley,, 1992.
4. **Casales, María Evelia Cabrera.** Data Warehouse (Almacenes de Datos). 2009.
5. **Padrón, Liudmila Torres.** *Almacenes de Datos*. [En línea] <http://www.monografias.com/trabajos31/almacenes-datos/almacenes-datos.shtml>.
6. **Osorio, Yolanda Falcón Rodríguez y Reinaldo Leyva.** Mercado de Datos Estadístico de Inmigración y Extranjería para el Departamento de Turismo y Comercio de la Oficina Nacional de Estadísticas.
7. **Velasco, JJ.** Alt1040. *IBM construye mayor Almacén Datos mundo*. [En línea] [Citado el: 26 de agosto de 2011.] <http://alt1040.com/2011/08/ibm-construye-mayor-almacen-datos-mundo>.
8. **Haas, Laura.** The theory and practice of information integration.
9. Freedownloadmanager. [En línea] [Citado el: 23 de Noviembre de 2011.] www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_%5Bcuenta_de_Plataforma_de_Java_14715_p.
10. Pgadmin. [En línea] [Citado el: 22 de Noviembre de 2011.] www.pgadmin.org/.
11. [En línea] [Citado el: 25 de Noviembre de 2011.] <http://datacleaner.eobjects.org/>.
12. **Ross, Ralph Kimball y Margy.** The Data Warehouse Toolkit. New York: Wiley Computer Publishing, 2002. Second Edition.
13. **Dario, Ricardo Bernabeu.** Metodología propia para la Construcción de un Data Warehouse. Córdoba, Argentina: s.n., 21 de Abril de 2009.

14. **Wolff, Carmen Gloria.** Modelamiento multidimensional.
15. Metodología para las Soluciones de Almacenes de Datos e Inteligencia de Negocios en CENTALAD. 2009: s.n.
16. PostgreSQL. [En línea] [Citado el: 24 de noviembre de 2011.] http://www.postgresql.org.es/sobre_postgresql.
17. *Pentaho BI Suite Enterprise Edition.* [En línea] 2005. [Citado el: 24 de Noviembre de 2011.] <http://www.pentaho.com/index.html>.
18. Darío, Ing. Bernabéu Ricardo. DATA WAREHOUSING: Investigación y Sistematización de Conceptos. Córdoba, Argentina: s.n., 2009.
19. Sinnexus. [En línea] Copyright 2007 - 2011. http://www.sinnexus.com/business_intelligence/datamart.aspx.
20. OLAP: Definición << Portal BI, [En línea] [Citado el: 24 de noviembre de 2011.] <http://portal-bi.cl/site/noticias/2011/olap-definicion>
21. Business Intelligence - Almacenes de Datos - ETL. (s.f.). [En línea] [Citado el: 24 de noviembre de 2011.] http://etl-tools.info/es/bi/proceso_etl.htm.
22. Sánchez, L.Z.Z., *Metodología para el Diseño Conceptual de Almacenes de Datos.* . 2008, Valencia
23. MEDINA MUSTELIER, Doris. Técnicas de extracción, transformación y carga de datos del Sistema de Información Nacional de Seguridad Ciudadana en la República Bolivariana de Venezuela. Tesis (Ingeniero en Ciencias Informáticas) Ciudad de La Habana. Universidad de las Ciencias Informáticas.
24. Martín Casal, M., *"la Técnica de las Cheskliten"*. 1989.
25. CICLO DE VIDA DEL SOFTWARE. [En línea] (2008). [Citado el: 24 de noviembre de 2011.]. Disponible en: < <http://es.kioskea.net/contents/genie-logiciel/cycle-de-vie.php3> />
26. UNIT TESTING EXAMPLE CONCEPTS AND FRAMEWORKL. [en línea] (2008). [Citado el: 24 de noviembre de 2011] Disponible en:< <http://maurizistorani.wordpress.com/>>
27. PRUEBAS DE SOFTWARE [en línea] (2009). [Citado el: 24 de noviembre de 2011.] Disponible en: <<http://www.slideshare.net/aracelij/pruebas-de-software/>>.

Referencias Bibliográficas

28. Kristin Daniel. INTEGRATION TESTING [En línea] (2000). [Citado el: 24 de noviembre de 2011.] Disponible en: <http://www.sei.cmu.edu/intro/process/technqs/q_it.htm />.
29. Kristin Daniel. FUNCTIONAL SYSTEM TESTING [En línea] (2000). [Citado el: 24 de noviembre de 2011.] Disponible en: <http://www.sei.cmu.edu/intro/process/technqs/q_fst.htm />

Bibliografía

1. **Domínguez, Antonio Blasco.** Programación. *Data Warehousing*. [En línea] [Citado el: 26 de noviembre de 2011.] http://www.programacion.com/articulo/data_warehousing_201.
2. **Bressán, Griselda E.** *Objetivos del Data Warehouse*. [En línea] <http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MineriaDatosBressan.htm#Objetivos%20del%20Data%20Warehouse..>
3. **Inmon., W. H.** Building the Data Warehouse. s.l. : QED Press/John Wiley,, 1992.
4. **Casales, María Evelia Cabrera.** Data Warehouse(Almacenes de Datos). 2009.
5. **Padrón, Liudmila Torres.** *Almacenes de Datos*. [En línea] <http://www.monografias.com/trabajos31/almacenes-datos/almacenes-datos.shtml>.
6. **Osorio, Yolanda Falcón Rodríguez y Reinaldo Leyva.** Mercado de Datos Estadístico de Inmigración y Extranjería para el Departamento de Turismo y Comercio de la Oficina Nacional de Estadísticas.
7. **Velasco, JJ.** Alt1040. *IBM construye mayor Almacen Datos mundo*. [En línea] [Citado el: 26 de agosto de 2011.] <http://alt1040.com/2011/08/ibm-construye-mayor-almacen-datos-mundo>.
8. **Haas, Laura.** The theory and practice of information integration.
9. Freedownloadmanager. [En línea] [Citado el: 23 de Noviembre de 2011.] www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_%5Bcuenta_de_Plataforma_de_Java_14715_p.
10. Pgadmin. [En línea] [Citado el: 22 de Noviembre de 2011.] www.pgadmin.org/.
11. [En línea] [Citado el: 25 de Noviembre de 2011.] <http://datacleaner.eobjects.org/>.
12. **Ross, Ralph Kimball y Margy.** The Data Warehouse Toolkit. New York : Wiley Computer Publishing, 2002. Second Edition.
13. **Dario, Ricardo Bernabeu.** Metodología propia para la Construcción de un Data Warehouse. Córdoba, Argentina : s.n., 21 de Abril de 2009.

14. **Wolff, Carmen Gloria.** Modelamiento multidimensional.
15. Metodología para la Soluciones de Almacenes de Datos e Inteligencia de Negocios en CENTALAD. 2009 : s.n.
16. PostgreSQL. [En línea] [Citado el: 24 de noviembre de 2011.] http://www.postgresql.org.es/sobre_postgresql .
17. *Pentaho BI Suite Enterprise Edition.* [En línea] 2005. [Citado el: 24 de Noviembre de 2011.] <http://www.pentaho.com/index.html>.
18. Dario, Ing. Bernabeu Ricardo. DATA WAREHOUSING:Investigación y Sistematización de Conceptos. Córdoba, Argentina : s.n., 2009.
19. Sinnexus. [En línea] Copyright 2007 - 2011. http://www.sinnexus.com/business_intelligence/datamart.aspx.
20. OLAP: Definición << Portal BI, Recuperado el 8/05/2011, disponible en <http://portal-bi.cl/site/noticias/2011/olap-definicion>
21. Hanik, F. (2007). INTRODUCTION TO APACHE TOMCAT 6.
22. Kimball, Ralph. Designing the Operational Data: Revista DM Review. The Data Warehouse Toolkit. s.l: WILEY PUBLICHING, 1996.
23. Apache Tomcat. [En línea] [Citado el: 20 de 10 de 2011.] <http://tomcat.apache.org/>.
24. MOLAP, ROLAP, HOLAP. [En línea] 2010. [Citado el: 20 de 10 de 2011.] http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx.
25. Mondrian - El servidor OLAP Open Source. [En línea] [Citado el: 20 de 10 de 2011.] <http://pentaho.almacen-datos.com/mondrian.html>.
26. Data Integration: Using ETL. EAI and EII Tools to Create an Integrate Enterprise. [En línea] [Citado el: 20 de 10 de 2011.] <http://www.bi-bestpractices.com/view-articles/4737>.
27. Pentaho Mondrian Documentación. [En línea] [Citado el: 20 de 10 de 2012.] <http://mondrian.pentaho.com/documentation/index.php>.

28. Sobre PostgreSQL. [En línea] [Citado el: 20 de 10 de 2011.] http://www.postgresql-es.org/sobre_postgresql.
29. Business Intelligence - Almacenes de Datos - ETL. (s.f.). [En línea] [Citado el: 24 de noviembre de 2011.] http://etl-tools.info/es/bi/proceso_etl.htm.
30. Sánchez, L.Z.Z., Metodología para el Diseño Conceptual de Almacenes de Datos. . 2008, Valencia.
31. MEDINA MUSTELIER, Doris. Técnicas de extracción, transformación y carga de datos del Sistema de Información Nacional de Seguridad Ciudadana en la República Bolivariana de Venezuela. Tesis (Ingeniero en Ciencias Informáticas) Ciudad de La Habana. Universidad de las Ciencias Informáticas.
32. Martín Casal, M., "la Técnica de las Chesklisen". 1989.
33. CICLO DE VIDA DEL SOFTWARE. [En línea] (2008). [Citado el: 24 de noviembre de 2011.]. Disponible en: < <http://es.kioskea.net/contents/genie-logiciel/cycle-de-vie.php3> />
34. UNIT TESTING EXAMPLE CONCEPTS AND FRAMEWORK. [en línea] (2008). [Citado el: 24 de noviembre de 2011] Disponible en:< <http://maurizistorani.wordpress.com/>>
35. PRUEBAS DE SOFTWARE [en línea] (2009). [Citado el: 24 de noviembre de 2011.] Disponible en: <<http://www.slideshare.net/aracelij/pruebas-de-software/>>.
36. Kristin Daniel. INTEGRATION TESTING [En línea] (2000). [Citado el: 24 de noviembre de 2011.] Disponible en: <http://www.sei.cmu.edu/intro/process/technqs/q_it.htm />.
37. Kristin Daniel. FUNCTIONAL SYSTEM TESTING [En línea] (2000). [Citado el: 24 de noviembre de 2011.] Disponible en: <http://www.sei.cmu.edu/intro/process/technqs/q_fst.htm>

Anexos

Anexo 1: Especificación del Caso de Uso: Presentar información de las transfusiones

Objetivo	Presentar información de las transfusiones efectuadas dado un rango de fecha, el carácter de la transfusión y el componente sanguíneo.	
Actores	Analista: (Inicia) validando los datos de un usuario en el sistema	
Resumen	El CU inicia cuando el usuario accede al sistema como analista, cuando finaliza el sistema muestra los reportes relacionados con los pacientes.	
Complejidad	Media	
Prioridad	Opcional <i>[Crítico, Secundario, Auxiliar, Opcional]</i>	
Precondiciones	Mercado de datos poblado Usuario autenticado	
Postcondiciones	Se muestran los datos	
Flujo de eventos		
Flujo básico <Nombre del flujo básico>		
	Actor	Sistema
1.	Se autentica en el sistema	
2.		El sistema verifica la autenticación
3.	Selecciona el AAG SIAPS	
4.		Muestra las AA contenidas dentro del AAG Salud
5.	Selecciona el AA Visor de Historias Clínicas	
6.		Muestra los LT contenidos dentro del AA Visor de Historias Clínicas
7.	Selecciona el LT Trasmisiones	
8.		Muestra los reportes contenidos dentro

		del LT Donaciones y transfusiones
9.	Selecciona el reporte que desea analizar	
10.		Muestra la información contenida en el reporte seleccionado y brinda opciones al actor para hacer cambios al reporte durante su análisis. Ir al CU Visualizar reportes. Finaliza el CU.
Flujos alternos		
Nº Evento < datos son incorrectos >		
	Actor	Sistema
1.		Muestra un mensaje de error y regresa al paso 1 del Flujo Normal de eventos
Opciones de reportes de < Presentar Información de donaciones y transfusiones >		
Perspectivas de análisis		Posibles resultados
		Medidas
		Periodicidad
Rango de fecha		Cantidad de transfusiones
Carácter de transfusión		
Componente sanguíneo		Semanal o mensual
Prototipo de interfaz de usuario		
Relaciones	CU Incluidos	
	CU Extendidos	<i>Visualizar Reporte</i>
Requisitos no	RNF3. Agilizar el acceso a los reportes del almacén de datos mediante la	

funcionales	distribución de la información por áreas de análisis. RNF4. Asegurar la disponibilidad del sistema.
Asuntos pendientes	<i>[Posibles mejoras al caso de uso.]</i>

Anexo 2: Consulta MDX correspondiente al reporte transfusiones por departamento solicitud.

SELECT NON EMPTY Hierarchize(Union({[dim_departamento_solicitante].[Departamentos]},
[dim_departamento_solicitante].[Departamentos].Children)) **ON COLUMNS, NON EMPTY**
{[dim_rango_fecha].[2009]} **ON ROWS** from [hecho_autopsias_realizadas]

Anexo 3: Lista de chequeo correspondiente al artefacto Mapa lógico.

Estructura del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
crítico	1. ¿El alcance del proyecto describe correctamente los datos de las dimensiones y hechos del mercado de datos?	0			
crítico	2. ¿El objetivo expresa correctamente el propósito del documento?	0			
crítico	3. ¿Existe una adecuada correspondencia entre el origen de los datos y los atributos del mercado?	0			
	4. ¿Se hace un uso adecuado del control del documento?	0			
	5. ¿En la sección de acrónimos se definen todos los acrónimos utilizados en el documento?	1		1	

Indicadores definidos en el desarrollo					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
	1. ¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?	0			
Semántica del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
crítico	1. ¿Se han identificado errores ortográficos en el documento?	0			
crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?	0			
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?	0			

Anexo 4: Lista de chequeo correspondiente al artefacto Diccionario de datos.

Estructura del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
crítico	6. ¿El entregable contiene las secciones obligatorias de la plantilla estándar definida para el expediente de proyecto?	0		0	
crítico	7. ¿El alcance del proyecto describe correctamente los datos de las dimensiones y hechos del mercado de datos?	0		0	

crítico	8. ¿El objetivo expresa correctamente el propósito del documento?	0		0	
	9. ¿Se hace un uso adecuado del control del documento?	0		0	
	10. ¿En la sección de acrónimos se definen todos los acrónimos utilizados en el documento?	0		0	
	11. ¿En el entregable, la definición de las variables se hace correctamente?	0		0	
	12. ¿Existe una adecuada correspondencia entre las variables definidas y las descripciones que tienen estas variables?	0		0	
	13. ¿En el entregable se crea una hoja por cada variable definida?	0		0	
	14. ¿Queda registrado en el entregable todos los posibles valores que van a tener las variables definidas?	0		0	

Indicadores definidos en el desarrollo

Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
	1. ¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?	1		1	

Semántica del documento

Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
crítico	4. ¿Se ha identificado errores ortográficos en el entregable?	0		0	
crítico	5. ¿Se entiende claramente lo que se ha especificado en el documento?	0		0	

	6. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?	0		0	
--	--	---	--	---	--

Anexo 5: Lista de chequeo correspondiente al artefacto Perfilado de datos.

Estructura del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
crítico	➤ ¿El entregable contiene las secciones obligatorias de la plantilla estándar definida para el expediente de proyecto? (ver expediente de proyecto)	0		0	
crítico	➤ ¿El alcance del proyecto describe correctamente los datos de las dimensiones y hechos del mercado de datos?	0		0	
crítico	➤ ¿El objetivo expresa correctamente el propósito del documento?	0		0	
	➤ ¿Se hace un uso adecuado del control del documento?	0		0	
	➤ ¿En la sección de acrónimos se definen todos los acrónimos utilizados en el documento?	0		0	
	➤ ¿Se identifican correctamente todos los objetos del negocio?	0		0	
	➤ ¿Queda reflejado correctamente en el entregable la definición de todas las tablas de la fuente de datos?	0		0	
	➤ ¿Existe calidad en los datos de la fuente?	0		0	
Indicadores definidos en el desarrollo					

Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
	1. ¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?	0		0	
Semántica del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
crítico	➤ ¿Se han identificado errores ortográficos en los entregables?	0		0	
crítico	➤ ¿Se entiende claramente lo que se ha especificado en el documento?	0		0	
	➤ ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?	0		0	Actualizar el índice.