

Universidad de las Ciencias Informáticas

Facultad 7



Título: Mercado de Datos para el Sistema de Gestión Estomatológica alasSiGEST.

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

Autores: Adrian Martínez Docal

Jesús Eiseel Peña Ceballo

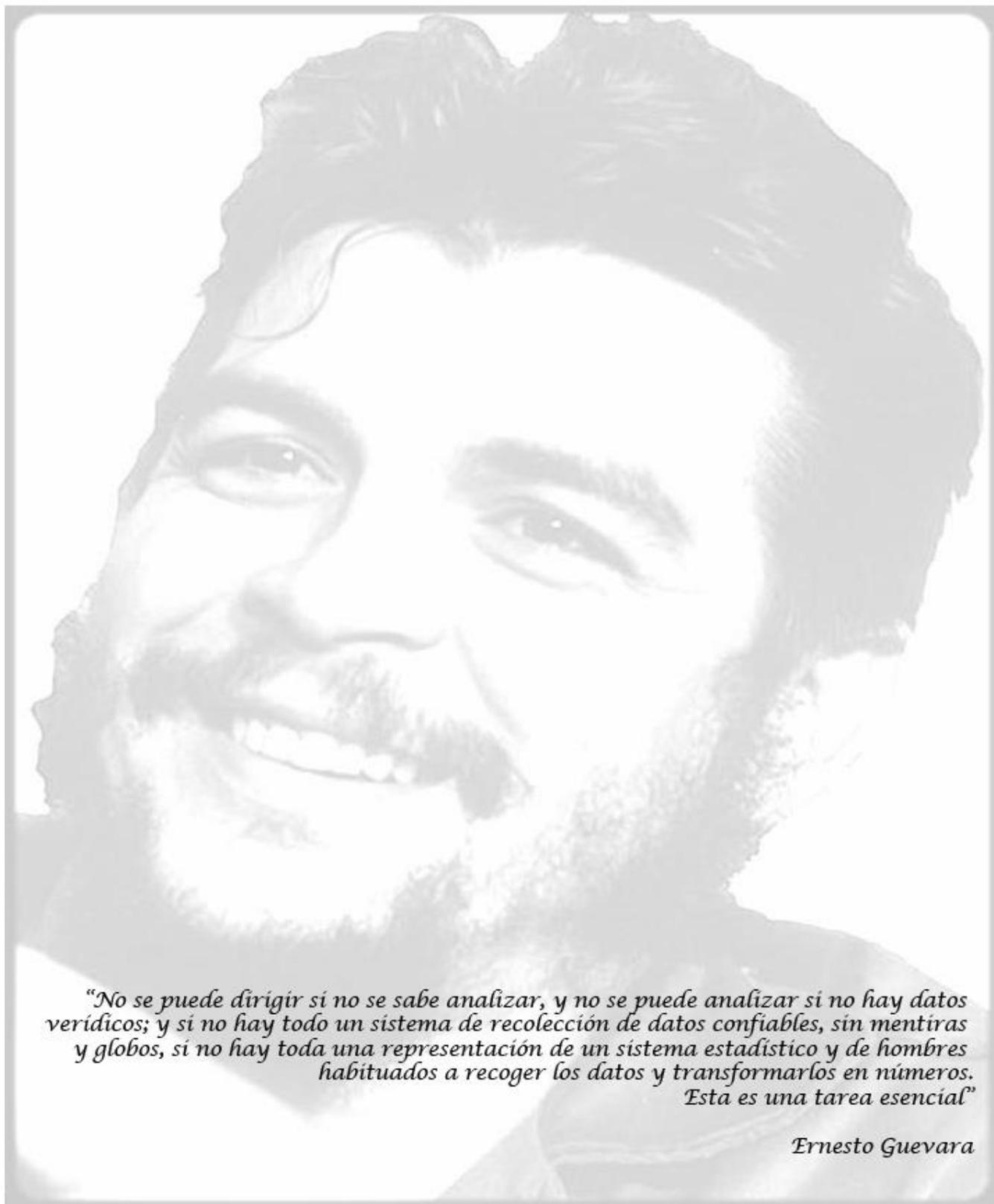
Tutores: Ing. Javier Villares Arias

Ing. Yelenis Díaz Larquin

Ing. Arlettys Fernández Ferrer

La Habana, 2013

“ Año 55 de la Revolución”



"No se puede dirigir si no se sabe analizar, y no se puede analizar si no hay datos verídicos; y si no hay todo un sistema de recolección de datos confiables, sin mentiras y globos, si no hay toda una representación de un sistema estadístico y de hombres habituados a recoger los datos y transformarlos en números. Esta es una tarea esencial"

Ernesto Guevara

DECLARACIÓN DE AUTORÍA

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

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

Adrian Martínez Docal

Jesús Eiseel Peña Ceballo

Firma del Autor

Firma del Autor

Arlettys Fernández Ferrer

Yelenis Díaz Larquin

Firma del Tutor

Firma del Tutor

Javier Villares Arias

Firma del Tutor

DATOS DE CONTACTOS

Ing. Javier Villares Arias. Graduado en el año 2009 en la Universidad de las Ciencias Informáticas de Ingeniero en Ciencias Informáticas, con categoría docente de Instructor. Durante su trayectoria laboral se ha desempeñado como profesor de las asignaturas Administración de Servidores Linux y Debates Históricos Contemporáneos, también se desempeña en el Departamento de Sistemas Especializados en Salud (SES) en los roles de Desarrollador en el proyecto de Estomatología y Arquitecto del departamento. Ha prestado servicios en el Departamento de Gestión Hospitalaria en el desarrollo del sistema alasHiS. Correo electrónico: jvillares@uci.cu

Ing. Yelenis Díaz Larquin. Graduada de Ingeniería en Ciencias Informáticas en la Universidad de las Ciencias Informáticas (UCI) en el 2012. Recién graduada en adiestramiento. Pertenece al Departamento de Sistemas Especializados en Salud (SES) del Centro de Informática Médica (CESIM) y desempeña el rol de Administradora de la Calidad del departamento.

Correo electrónico: ylarquin@uci.cu

Ing. Arlettys Fernández Ferrer. Graduada de Ingeniería en Ciencias Informáticas en la Universidad de las Ciencias Informáticas (UCI) en el 2012. Recién graduada en adiestramiento. Pertenece al Departamento de Sistemas Especializados en Salud (SES) del Centro de Informática Médica (CESIM) y desempeña el rol de Analista Principal del departamento.

Correo electrónico: afferrer@uci.cu

DEDICATORIA

De Chuchy:

Este trabajo está dedicado a la persona más especial en mi vida: mi mamá. Por su amor, sacrificio, apoyo y confianza que me han sido indispensables para llegar a convertirme en lo que soy.

A mi abuela por ser otra madre más y enseñarme tantas cosas de la vida. A mi tía y tío por quererme como un hijo y ayudarme todos estos años de mi vida. A mi hermana, mis primos Kennys y Omannys por el cariño y apoyo incondicional que me han brindado siempre.

De Adrián:

A mi mamá por ser la mejor madre del mundo y estar siempre ahí apoyándome en todo y querer lo mejor para mí. Por su confianza y sacrificio. A mi familia, por estar siempre conmigo en las buenas y las malas, por apoyarme y confiar siempre en mí.

AGRADECIMIENTOS

De Chuchy:

A mi mamá por apoyarme y darme siempre la fuerza y la confianza para cumplir la meta de culminar esta carrera universitaria.

A mi abuela, mi hermana, mi tía y tío, mis primos Omannys y Kennys, por los consejos y la ayuda que me han brindado en todos estos años.

A mis amigos de tantos años que me han demostrado que las buenas amistades perduran toda la vida: Brian, Jeordany, Albertico, Hector David, Yasiel.

A todos los buenos amigos con los que he compartido estos cinco años de carrera y viví tantos momentos inolvidables: Manu, Rey, Osmel, Diosbel, Ale, Oscar, Maikel, Alberto, Frances, Shogun. A Dayanelis por ser mi guía en los momentos difíciles y apoyarme en las buenas y en las malas.

A mi compañero de tesis Adrián por ayudarme a cumplir la difícil tarea de desarrollar satisfactoriamente este trabajo.

A mis tutores y a mi oponente por el tiempo dedicado y la preocupación que hicieron posible el desarrollo y la culminación de este trabajo.

A los integrantes del tribunal que por sus consejos y recomendaciones contribuyeron a mi superación y a mejorar el trabajo realizado.

De Adrián:

A mi mamá por ser la mejor madre del mundo y estar siempre ahí apoyándome en todo y querer lo mejor para mí. Por su confianza y sacrificio.

A mi hermano Alejandro por su apoyo. A mi familia, por estar siempre conmigo en las buenas y las malas, por apoyarme y confiar siempre en mí.

A mis compañeros de apartamento por haber compartido unos 5 años inolvidables.

A mis tutores y mi oponente por la preocupación y ayuda que me han brindado para hacer mi tesis.

A mi compañero de tesis Chuchy por haberme ayudado a la realización de mi sueño.

A mi tribunal, por darme consejos que hoy me hacen ver el camino más claro.

RESUMEN

La Estomatología es una de las especialidades de la Medicina que tiene gran importancia dentro del Sistema Nacional de Salud de Cuba. Su información es procesada por el Centro Nacional de Estomatología (CeNaEst) que recoge los datos de todas las clínicas estomatológicas del país. Por tal motivo surge la necesidad de almacenar grandes volúmenes de información para la toma de decisiones, así como la automatización del proceso de análisis. Para ello se traza como objetivo desarrollar un mercado de datos que contenga la información consolidada que necesitan a diario los directivos del CeNaEst.

Actualmente esta información se encuentra almacenada en el sistema alasSIGEST desarrollado por el departamento Sistemas Especializados para la Salud (SES) del Centro de Informática Médica (CESIM). Dicho sistema cuenta con una base de datos relacional sobre el sistema gestor de bases de datos PostgreSQL lo que impide hacer un análisis exhaustivo de la información consolidada. La metodología utilizada en la construcción del mercado de datos fue UCID, un híbrido entre las metodologías Hefesto y Kimball. Se utilizaron para el desarrollo del sistema herramientas de software libre.

Se obtuvo un producto de software que consiste en un mercado de datos, el cual brinda reportes y acceso eficiente a la información que necesitan los directivos del CeNaEst mejorando en gran medida el tiempo de respuesta. Los reportes generados se pueden analizar mediante gráficas y tablas con valores numéricos, garantizando una mejor comprensión y análisis a la hora de tomar decisiones.

PALABRAS CLAVES:

Mercado de Datos, toma de decisiones, información consolidada.

Índice

INTRODUCCIÓN.....	1
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA DE LOS ALMACENES DE DATOS	5
1 ALMACENES DE DATOS O DATA WAREHOUSE	5
1.1 MERCADOS DE DATOS EXISTENTES	8
1.1.1 NIVEL INTERNACIONAL	8
1.1.2 NIVEL NACIONAL.....	9
1.1.3 UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS.....	9
1.2 METODOLOGÍAS PARA EL DISEÑO E IMPLEMENTACIÓN DE UN MERCADO DE DATOS.....	10
1.2.1 KIMBALL	11
1.2.2 INMON	11
1.2.3 METODOLOGÍA HEFESTO	12
1.2.4 METODOLOGÍA UCID	12
1.3 PROCESAMIENTO DE TRANSACCIONES EN LÍNEA (OLTP).....	15
1.4 BASE DE DATOS MULTIDIMENSIONAL	16
1.4.1 ESQUEMA EN ESTRELLA	16
1.4.2 ESQUEMA COPO DE NIEVE.....	17
1.4.3 ESQUEMA CONSTELACIÓN O COPO DE ESTRELLAS.....	18
1.5 EXTRACCIÓN, TRANSFORMACIÓN Y CARGA (ETL).....	19
1.5.1 HERRAMIENTAS PARA ETL.....	19
1.5.1.1 <i>Clover ETL</i>	19
1.5.1.2 <i>ENHYDRAOCTOPUS</i>	20
1.5.1.3 <i>KETTLE</i>	20
1.6 OLAP.....	21
1.6.1 HERRAMIENTAS PARA OLAP	23
1.6.1.1 <i>JEDOX PALO</i>	23
1.6.1.2 <i>OLAP4J</i>	23
1.6.1.3 <i>MONDRIAN</i>	23
1.6.1.4 <i>SCHEMA WORKBENCH (KETTLE)</i>	24
1.6.1.5 <i>Lenguaje MDX</i>	24
1.7 SISTEMAS GESTORES DE BASES DE DATOS.....	25
1.7.1 ORACLE	25
1.7.2 MYSQL.....	26
1.7.3 POSTGRESQL.....	26
1.8 HERRAMIENTAS CASE	28
1.8.1 RATIONAL ROSE	28
1.8.2 VISUAL PARADIGM.....	28

1.8.3	ENTERPRISE ARCHITECT.....	28
1.9	PLATAFORMA PARA EL DESARROLLO DEL MERCADO DE DATOS	29
1.9.1	MÓDULOS DE LA PLATAFORMA PENTAHO BI.....	29
	CAPÍTULO 2: ANÁLISIS Y DISEÑO DE LA ESTRUCTURA DEL MERCADO DE DATOS.....	31
2.1	FASE 1: ANÁLISIS DE LOS REQUERIMIENTOS	31
2.1.1	REQUERIMIENTOS DE INFORMACIÓN.....	31
2.1.2	REQUERIMIENTOS FUNCIONALES.....	32
2.1.3	REQUERIMIENTOS NO FUNCIONALES.....	32
2.1.4	IDENTIFICAR INDICADORES Y PERSPECTIVAS.....	33
2.1.5	MODELO CONCEPTUAL.....	34
2.2	FASE 2: ANÁLISIS DE LAS FUENTES DE DATOS.....	36
2.2.1	CORRESPONDENCIA CON LOS REQUERIMIENTOS.....	36
2.2.2	CAMPOS QUE INTEGRAN LAS PERSPECTIVAS. NIVEL DE GRANULARIDAD.....	37
2.3	FASE 3: MODELADO DEL MERCADO DE DATOS.....	39
2.3.1	IDENTIFICACIÓN DE DIMENSIONES Y HECHOS.....	39
2.3.2	MODELO LÓGICO DEL MERCADO DE DATOS.....	40
2.4	ARQUITECTURA DEL MERCADO DE DATOS.....	41
2.5	POLÍTICAS DE RESPALDO Y RECUPERACIÓN	43
2.6	SEGURIDAD DEL MERCADO DE DATOS	43
	CAPÍTULO 3: IMPLEMENTACIÓN DEL MERCADO DE DATOS.....	45
3.1	REGLAS DE TRANSFORMACIÓN.....	45
3.2	ESTÁNDARES PARA EL DISEÑO DEL PROCESO ETL.....	46
3.3	PROCESOS DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA (ETL)	46
3.3.1	AUTOMATIZACIÓN DEL SISTEMA ETL.....	48
3.4	PROCESOS DE ANÁLISIS DE LA INFORMACIÓN CONTENIDA EN EL MERCADO DE DATOS.....	49
3.4.1	JERARQUÍAS DE LAS DIMENSIONES.....	49
3.4.2	MEDIDAS.....	50
3.4.3	CONSULTAS MDX.....	51
3.4.4	RESULTADOS DE LA HERRAMIENTA MONDRIAN. PENTAHO BI SERVER.....	52
3.5	DESPLIEGUE DEL MERCADO DE DATOS.....	54
3.6	PRUEBAS.....	55
3.6.1	PRUEBAS DE RENDIMIENTO Y ESTRÉS.....	56
3.6.2	LISTA DE CHEQUEO.....	58
	CONCLUSIONES.....	61
	REFERENCIAS BIBLIOGRÁFICAS.....	62

BIBLIOGRAFÍA.....	65
ANEXOS.....	67
1 CARGA DE LAS DIMENSIONES Y HECHOS DEL MERCADO DE DATOS.....	67
2. REQUERIMIENTOS DE INFORMACIÓN.....	72
3. REQUERIMIENTOS FUNCIONALES	74
4. REQUERIMIENTOS NO FUNCIONALES	76
5. INDICADORES Y PERSPECTIVAS.....	78

Índice de Figuras

Figura 1 Cubo Multidimensional (2)	16
Figura 2 Esquema en estrella (2).....	17
Figura 3 Esquema copo de nieve (2).....	17
Figura 4 Esquema Constelación (2)	18
Figura 5 Modelo Conceptual.....	35
Figura 6 Modelo Conceptual con atributos.	38
Figura 7 Matriz Bus.....	40
Figura 8 Estructura del Mercado de Datos.	41
Figura 9 Arquitectura del Mercado de Datos para el CeNaEst.	42
Figura 10 Carga de la dimensión d_especialidad	46
Figura 11 Carga de la dimensión d_paciente	47
Figura 12 Carga de la tabla de hechos “hechos_pacientes_protesis”	48
Figura 13 Ejecución del trabajo	48
Figura 14 Jerarquía de los atributos de la dimensión d_localidad.Figura 14	49
Figura 15 Jerarquía de los atributos de la dimensión d_tiempo.....	50
Figura 16 Diseño de los cubos multidimensionales.	51
Figura 17 Consultas MDX.....	52
Figura 18 Dimensión Localidad con las medidas e indicadores según el filtrado.	53
Figura 19 Información Gráfica	53
Figura 20 Otra vista de la información gráfica	54
Figura 21 Despliegue del Mercado de Datos.....	55
Figura 22 Prueba de rendimiento y estrés.....	57

Figura 23 Prueba de rendimiento a la transformación de la tabla de hechos “h_pacientes_egi”..	58
Figura 24 Comportamiento de los indicadores por etapas.....	59

INTRODUCCIÓN

El avance de las Tecnologías de la Información y las Comunicaciones (TIC), han causado un gran impacto en el mundo en las últimas décadas, provocando un aumento considerable de la información que se genera. Actualmente muchas empresas almacenan grandes volúmenes de datos para extraer la información útil que les facilite la toma de decisiones en dependencia de sus necesidades, pero muchas veces este proceso se ve afectado por uno de los mayores problemas de estos tiempos: “manejar grandes volúmenes de información”.

El Ministerio de Salud Pública de Cuba (MINSAP), organismo rector del Sistema Nacional de Salud, es el encargado de dirigir, ejecutar y controlar la aplicación de la política del Estado y del gobierno en cuanto a la salud pública, el desarrollo de las Ciencias Médicas y la industria médico-farmacéutica. El Centro Nacional de Estomatología (CeNaEst), tiene como objeto social, desarrollar la política y estrategia del MINSAP, de implementar el desarrollo de la docencia e investigación de los profesionales de la rama para alcanzar los niveles de excelencia exigidos por la población. Para apoyar el cumplimiento de dicho objetivo en Cuba se desarrollan diversos productos informáticos para facilitarles el trabajo a los encargados de mantener la salud pública. Algunos de estos productos van encaminados a las diferentes especialidades de la Estomatología, la cual es la rama de la odontología que se encarga del diagnóstico, estudio y tratamiento de las enfermedades que afectan la boca (principalmente los tejidos blandos) y en general de todo el sistema estomatognático. (1)

La Estomatología en el país tiene antecedentes que se remontan a los aborígenes, pasando por diversos momentos históricos concretos hasta el día de hoy, contribuyendo a elevar la salud bucal y mejorando la calidad de vida de la población. A partir del triunfo de la revolución, hasta la actualidad, ha crecido como ciencia y en estos momentos se encuentra en una etapa de florecimiento científico. Se ha continuado su desarrollo, introduciendo tecnologías de avanzadas, perfeccionando la calidad de la docencia e incrementando la solidaridad e internacionalismo con pueblos hermanos.

La Universidad de las Ciencias Informáticas (UCI) está conformada por varios centros productivos cuyo objetivo es apoyar y aumentar el desarrollo de software en el país, uno de ellos es el Centro de Informática Médica (CESIM) que pertenece a la Facultad 7. Este centro está conformado por diferentes

INTRODUCCIÓN

departamentos, los cuales desarrollan aplicaciones que van encaminadas a mejorar los procesos del Sistema Nacional de Salud (SNS). Muestra de ello es el departamento de Sistemas Especializados para la Salud (SES), que ha tenido como propósito apoyar el proceso de informatización en cada una de estas áreas. El Sistema de Gestión Estomatológica alasSiGEST es un producto creado en este ámbito, el cual tiene como objetivo la gestión de la información relacionada con las Consultas Especializadas del Servicio de Estomatología en CeNaEst.

El sistema alasSiGEST, permite gestionar la información de las especialidades: Estomatología General Integral, Ortodoncia, Periodoncia, Prótesis y Cirugía Bucomaxilofacial, pero debido a la gran cantidad de datos que se generan en el CeNaEst se dificulta el análisis de la información para dar soporte a la toma de decisiones. Por lo tanto se necesita contar con un sistema que permita extraer, transformar, integrar y almacenar los datos relevantes de las diferentes especialidades del Sistema de Estomatología con el objetivo de facilitar el acceso y exploración de la información requerida, reduciendo el tiempo de consultas complejas.

Un almacén de datos maneja grandes volúmenes de datos, debido a que consolida en su estructura la información recolectada durante años, proveniente de diversas fuentes, en un solo lugar centralizado. Es por esta razón que el depósito puede ser soportado y mantenido sobre diversos medios de almacenamiento. En síntesis, un almacén de datos no es solo mera información, sino un conjunto de herramientas para consultar, analizar y presentar información, que permiten obtener o realizar análisis, reportes, extracción y explotación de los datos de manera eficaz, para transformar dichos datos en información valiosa para la organización. (2)

Partiendo de lo anteriormente expuesto se identifica como **problema a resolver**: ¿Cómo facilitar la toma de decisiones en el CeNaEst?

En función del problema a resolver identificado se determina como **objeto de estudio**: Los almacenes de datos.

El objeto de estudio se enmarca en el **campo de acción**: Mercado de Datos para el Centro Nacional de Estomatología.

INTRODUCCIÓN

Para darle solución al problema se tiene como **objetivo general**: Desarrollar un mercado de datos que contribuya a facilitar el análisis de la información generada en el CeNaEst, de manera que apoye la toma de decisiones tácticas y estratégicas.

Para darle cumplimiento al objetivo planteado las **tareas** que se llevarán a cabo son:

1. Identificación las necesidades de información del cliente mediante la realización de preguntas claves del negocio.
2. Diseño el modelo conceptual del almacén de datos mediante el cual se adquiera una idea precisa del alcance del almacén.
3. Revisión las fuentes de información generadas en el CeNaEst en las cuales se obtenga la información relevante de las mismas.
4. Elaboración el modelo de datos del almacén obteniéndose su estructura para el CeNaEst.
5. Realización Extracción, Transformación y Carga para el almacén de datos.
6. Realización los procesos de análisis de la información contenida en un almacén de datos mediante el cual se agilicen las consultas de la información.

Una vez concluida esta investigación se espera como resultado contar con un sistema que contenga la información consolidada del Sistema de Gestión Estomatológica alasSiGEST, lo que facilitará la toma de decisiones al personal médico, ya que podrán obtener y consultar datos en diferentes niveles, además posibilitará un ahorro notable de recursos materiales.

Durante el desarrollo de esta investigación se emplean métodos teóricos, entre ellos el método Histórico-Lógico, ya que en la primera etapa de la investigación se desarrolla un estudio del estado del arte de la problemática que se plantea, haciendo un estudio de las fuentes bibliográficas necesarias para lograr un análisis más profundo de lo que se quiere desarrollar. Además se utilizan los métodos empíricos, porque estos emplean la observación, medición, experimentación y la entrevista, los cuales son indispensables para comenzar a desarrollar el presente trabajo. También se hace uso de la modelación para modelar diagramas, representar el proceso de desarrollo y propiciar un mejor entendimiento de la solución a implementar.

INTRODUCCIÓN

El contenido de este trabajo se encuentra estructurado de la siguiente manera:

Capítulo 1: Fundamentación Teórica. En este capítulo se describen los conceptos asociados al dominio del problema. Se aborda todo lo relacionado con Almacenes de Datos y se describen las tecnologías, herramientas y metodologías necesarias para dar solución al problema que se plantea.

Capítulo 2: En el capítulo se realiza el análisis y diseño según los pasos que describe la metodología UCID. Se elabora un modelo conceptual del mercado de datos a partir de los indicadores y perspectivas obtenidas del negocio. Se definen los hechos, dimensiones y medidas por los que quedará conformado el modelo de datos y se confecciona el modelo lógico del sistema, de forma que quede conformado el diseño del Mercado de datos. Además, se define la arquitectura del mercado de datos y las políticas de respaldo y seguridad.

Capítulo 3: En el presente capítulo se realizan los procesos de Extracción, Transformación y Carga, se definen las jerarquías de los atributos de las tablas de dimensiones y se diseñaran los cubos multidimensionales que permiten realizar los cruces necesarios para visualizar la información contenida en el Almacén de Datos. En general, se realiza el procesamiento analítico en línea para agilizar las consultas de la información en el almacén, mostrando los datos a los directivos desde diferentes perspectivas. Además se realizarán pruebas de rendimiento y estrés para probar la eficiencia del mercado de datos.

Capítulo 1: Fundamentación Teórica de los Almacenes de Datos

En este capítulo se describen los conceptos asociados al dominio del problema. Se aborda todo lo relacionado con Almacenes de Datos y se describen las tecnologías, herramientas y metodologías necesarias para dar solución al problema que se plantea.

1 Almacenes de Datos o Data Warehouse

Para llevar a cabo cualquier tipo de acción en una empresa, es necesario gestionar datos guardados en diversos formatos, fuentes y tipos, para luego depurarlos e integrarlos, además de almacenarlos en un solo destino, depósito o base de datos que permita su posterior análisis y exploración. Debido a esto es imperativo y de vital importancia contar con una herramienta que satisfaga todas estas necesidades. Como forma de solucionar lo anteriormente planteado se crean los Data Warehouse (DW), el cual posibilita la extracción de datos de sistemas operacionales y fuentes externas. Además permite la integración y homogenización de los datos de toda la empresa y provee información que haya sido transformada para que ayude en el proceso de toma de decisiones estratégicas y tácticas. El DW convertirá los datos operacionales de la empresa en una herramienta competitiva, debido a que pondrá a disposición de los usuarios indicados la información pertinente, correcta e integrada, en el momento que se necesita.

Según Ralph Kimball: "El Data Warehouse es una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis; es la unión de todos los mercados de datos de una entidad". (2)

Otra de las definiciones más famosas sobre DW, es la de W. H. Inmon (reconocido mundialmente como el padre del DW), quien define: "Un Data Warehouse es una colección de datos orientada al negocio, integrada, variante en el tiempo y no volátil para el soporte del proceso de toma de decisiones de la gerencia".

Características de un Data Warehouse según Inmon:

Orientado al Negocio: La información se clasifica en base a los aspectos que son de interés para la empresa. El DW excluye la información que no será utilizada exclusivamente en el proceso de toma de decisiones.

Integrado: La integración implica que todos los datos de diversas fuentes que son producidos por distintos departamentos, secciones y aplicaciones, tanto internas como externas, deben ser consolidados en una instancia antes de ser agregados al DW. A este proceso se le conoce como Extracción, Transformación y Carga de Datos (Extraction, Transformation and Load - ETL). La integración de datos, resuelve diferentes tipos de problemas relacionados con las convenciones de nombres, unidades de medidas, codificaciones y fuentes múltiples.

Variante en el tiempo: Debido al gran volumen de información que se manejará en el DW, cuando se realiza una consulta, los resultados deseados demorarán en originarse. Este espacio de tiempo que se produce desde la búsqueda de datos hasta su consecución es del todo normal en este ambiente y es, precisamente por ello, que la información que se encuentra dentro del depósito de datos se denomina de tiempo variable. Toda la información en el DW posee su propio sello de tiempo, lo cual constituye una de las principales ventajas del almacén de datos, ya que los datos son almacenados junto a sus respectivos históricos. Esto garantiza poder desarrollar análisis de la dinámica de la información, la cual es procesada como una serie de instantáneas, cada una representando un período de tiempo. Es decir, se podrá tener acceso a diferentes versiones de la misma información. El almacenamiento de datos históricos, es lo que permite al DW desarrollar pronósticos y análisis de tendencias y patrones, a partir de una base estadística de información.

No Volátil: La información es útil para el análisis y la toma de decisiones solo cuando es estable. Los datos operacionales varían momento a momento, en cambio, los datos una vez que entran en el DW no cambian. La actualización, o sea, insertar, eliminar y modificar, se hace de forma muy habitual en el ambiente operacional sobre una base, registro por registro, en cambio en el depósito de datos la manipulación básica de los datos es mucho más simple, debido a que solo existen dos tipos de operaciones, la carga de datos y el acceso a los mismos. Por esta razón es que en el DW no se requieren mecanismos de control de la concurrencia y recuperación. (2)

Las principales ventajas de los Almacenes de Datos son:

- Transforma datos orientados a las aplicaciones en información dirigida a la toma de decisiones.
- Elimina la producción y el procesamiento de datos que no son utilizados ni necesarios, producto de aplicaciones mal diseñadas o ya no utilizadas.

- Mejora la entrega de información, es decir, información completa, correcta, consistente, oportuna y accesible. Información que los usuarios necesitan, en el momento adecuado y en el formato apropiado.
- Integra y consolida diferentes fuentes de datos y departamentos empresariales, que anteriormente formaban islas, en una única plataforma sólida y centralizada.
- Permite la toma de decisiones estratégicas y tácticas.
- Permite reaccionar rápidamente a los cambios del mercado.
- Aumenta la competitividad del mercado.

La estructura de un almacén de datos permite una serie de análisis que mediante otros tipos de sistemas informáticos no se podrían realizar. Un almacén de datos ejecuta consultas muy dinámicas, solo con la combinación de dimensiones, lo que permite al usuario sin mucho conocimiento informático generar consultas complejas. Además, permite la extracción de datos provenientes de diversas fuentes y el almacenamiento y análisis de datos históricos, para la ayuda a la toma de decisiones, permitiendo obtener varias versiones de la misma información.

Mercado de Datos

Un mercado de datos es una versión especial de almacén de datos con el objetivo de responder a un determinado análisis, función o necesidad y con una población de usuarios específica. La diferencia entre un almacén y un mercado de datos es su alcance. Este último está pensado para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización, mientras el almacén es la organización en su conjunto.

La información que se necesita para la toma de decisiones en el CeNaEst se encuentra orientada a un único tema; el de los pacientes y sus respectivos datos en las especialidades Estomatología General Integral, Ortodoncia y Prótesis. Por tanto no es necesario la creación de un almacén de datos. Con un mercado de datos se tiene fácil acceso a los datos que se necesitan frecuentemente, mejora el tiempo de respuesta del usuario final, se simplifica el desarrollo de todo el mecanismo de la base de datos y con ello baja substancialmente todo el coste del proyecto, así como su duración.

1.1 Mercados de datos existentes

Las grandes corporaciones invierten importantes cantidades de sus presupuestos en la búsqueda de la forma más efectiva y rentable de tratar, almacenar y recuperar sus datos. Es por tal motivo que muchas de estas empresas ante un mercado tan competitivo implementan diferentes almacenes de datos para manejar su información de una manera más eficiente.

1.1.1 Nivel Internacional

La empresa Petroecuador cuenta con un mercado de datos que les proporciona a los directivos información de los volúmenes y precios de embarques de crudo y productos para la exportación e importación, almacenándose históricamente para permitir el análisis de tendencias y proyecciones brindando una mejor toma de decisiones al momento de diseñar las estrategias de producción. (3)

La Facultad Nacional de Ingeniería de Oruro en Bolivia cuenta con un mercado de datos que le permite obtener información acerca de las causas que influyen en cualquier fenómeno interno asociado con el proceso de enseñanza-aprendizaje de la Carrera de Ingeniería de Sistemas. También permite identificar las causas de la variación de tiempos de la permanencia estudiantil o cualquier otro fenómeno en la Carrera de Ing. de Sistemas. (4)

General Mills de Venezuela C.A. es una empresa transnacional dedicada a la manufactura de conservas alimenticias para el consumo humano. Todos los días se lleva a cabo un proceso automatizado de extracción de datos desde el sistema principal, los cuales después de analizados, procesados y transformados llenan las tablas de la base de datos del mercado de datos, esto les permite desarrollar reportes sobre precios, descuentos y promociones vigentes además del vencimiento de productos terminados. (3)

La Marina de Guerra del Perú (MGP) es el órgano encargado de la defensa marítima, fluvial y lacustre del país. Dicho órgano cuenta con un mercado de datos que le permite la planificación y el análisis estratégico, el cual brinda una adecuada información que permite la realización de planes referentes a cursos de capacitación o instrucción que sirven para una adecuada gestión y por ende una acertada toma de decisiones. Este mercado brinda a la MPG información consistente y disponible en tiempo real del personal que ha llevado cursos ya sea de calificación o carrera. Por lo tanto se tomarán decisiones más rápidas respecto a donde asignar determinado personal. (5)

1.1.2 Nivel Nacional

La Empresa de Proyectos de Arquitectura e Ingeniería (EMPAI) de Matanzas es una empresa que aplica el perfeccionamiento empresarial desde 1999. Ésta posee un mercado de datos que posibilita la gestión económica- financiera de la empresa. El mismo sirve de apoyo en la toma de decisiones de la dirección de gestión económica y forma parte del almacén de datos empresarial de la entidad. (6)

EMPAI también implementó un mercado de datos para el control de la información relevante del proceso de negocio de Gestión del Capital Humano para elevar la efectividad de la organización. Para la construcción del almacén usaron Microsoft SQL Server 2000 en la creación de la base de datos del mercado, para los cubos OLAP emplearon Microsoft Analysis Services y para el análisis de los datos contenidos en los cubos se auxiliaron de la herramienta de consulta OLAP, empleando ASP.NET en Microsoft Visual Studio.NET 2005. (6)

Una empresa como CIMEX, cuya labor incluye redistribuir productos a entidades para su venta a la población, con gran volumen de información a consultar de los productos comercializados en la corporación cuenta con un mercado de datos encargado de calcular las propuestas de compras permitiendo una gestión de compra-venta eficiente, con una finalidad fundamental: “Disminuir los costos, sin afectar al cliente, permitiendo prestaciones eficientes y con la calidad requerida, aumentando las ganancias o utilidades de las empresas”. (3)

El grupo empresarial de la industria químico farmacéutica de Cuba (QUIMEFA) es una empresa líder en su campo que incluye una amplia gama de productos químicos de uso industrial, productos avalados y destacados por su calidad. Dicha empresa hace uso de un mercado de datos el cual brinda reportes y acceso eficiente a la información que necesitan los directivos de QUIMEFA. Los reportes generados se pueden analizar mediante gráficas y tablas con valores numéricos, garantizando una mejor comprensión y análisis a la hora de tomar decisiones. (7)

1.1.3 Universidad de las Ciencias Informáticas

En la UCI se han realizado varios trabajos de diploma relacionados con mercados de datos. Uno de ellos es el trabajo de diploma realizado en el año 2010 titulado “Implementación de un DataMart para la Unidad Central de Cooperación Médica”. El mercado de datos desarrollado en este trabajo contiene la información

consolidada del proyecto Colaboración Médica, con el propósito de facilitar el control estadístico de la información en un período de tiempo y mejorar el proceso de toma de decisiones para los directivos de la UCCM.

El Centro de Informatización para la Seguridad Ciudadana de la Universidad diseñó un mercado de datos para el sistema Salas Situacionales del Sistema de Gestión Penitenciaria. El mercado de datos está basado en la metodología *Data Warehouse Engineering Process* (Procesos de Ingeniería de Almacenes de Datos), utilizando el enfoque propuesto por Ralph Kimball así como la arquitectura de dos capas. Como gestor de base de datos se hace uso de Oracle, proponiendo además, la suite de Inteligencia de Negocios Pentaho para el proceso de extracción, transformación y carga de los datos. (8)

En el año 2011 se diseñó un mercado de datos para el área de Comercio Exterior de la Oficina Nacional de Estadísticas, el cual consolida la información del área Comercio exterior, con el propósito de facilitar el control estadístico de la información y mejorar el proceso de toma de decisiones para los directivos del Comercio exterior en Cuba. (9)

Otro de los trabajos de diploma es: “Diseño del Datamart del subsistema de Conducidos”, con el objetivo de tener una manera de obtener reportes de forma eficiente relacionada con el subsistema de los conducidos del sistema integral de la estación Policía Nacional Revolucionaria (PNR), permitiendo trazar estrategias en el mejoramiento de la eficiencia y la eficacia del trabajo policial a diferentes niveles en el país.

Luego de la investigación realizada de las distintas empresas que han implementado mercados de datos se puede llegar a la conclusión que estos juegan un papel fundamental en dichas organizaciones a la hora de analizar los grandes volúmenes de información que se generan, reduciendo el tiempo de consultas analíticas complejas y facilitando la toma de decisiones tácticas y estratégicas. Por tal motivo se acentúa la importancia de desarrollar un mercado de datos para facilitar la toma de decisiones a los directivos del CeNaEst.

1.2 Metodologías para el diseño e implementación de un Mercado de datos

Existen varias metodologías de diseño y construcción de un mercado de datos. Cada fabricante de software de inteligencia de negocios busca imponer una metodología con sus productos.

1.2.1 Kimball

Esta metodología se basa en lo que Kimball denomina Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle). Este ciclo de vida está basado en cuatro principios básicos:

Centrarse en el negocio: Hay que concentrarse en la identificación de los requerimientos del negocio y su valor asociado, además de usar estos esfuerzos para desarrollar relaciones sólidas con el negocio, agudizando el análisis del mismo y la competencia consultiva de los implementadores.

Construir una infraestructura de información adecuada: Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa.

Realizar entregas en incrementos significativos: Crear el almacén de datos (DW) en incrementos entregables en plazos de 6 meses. Se debe usar el valor de negocio de cada elemento identificado para determinar el orden de aplicación de los incrementos. En este aspecto la metodología se realiza de forma semejante a las metodologías ágiles de construcción de software.

Ofrecer la solución completa: Proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Se debe tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible. Además se deberá entregar herramientas de consulta, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación.

La metodología de Kimball proporciona una base empírica y metodológica adecuada para las implementaciones de almacenes de datos pequeños y medianos, dada su gran versatilidad y su enfoque ascendente, que permite construir los almacenes en forma escalonada. Además presenta una serie de herramientas, tales como planillas, gráficos y documentos, que proporcionan una gran ayuda para iniciarse en el ámbito de la construcción de un almacén de datos. (10)

1.2.2 Inmon

Inmon ve la necesidad de transferir la información de los diferentes OLTP (Sistemas Transaccionales) de las organizaciones a un lugar centralizado donde los datos puedan ser utilizados para el análisis. La metodología Inmon se referencia normalmente como Top-down. Los datos son extraídos de los sistemas

operacionales por los procesos ETL y cargados, donde son validados y consolidados en el DW corporativo, donde además existen los llamados metadatos que documentan de una forma clara y precisa el contenido del DW. Una vez realizado este proceso, la actualización de los mercados de datos departamentales obtiene la información de este, y con las consiguientes transformaciones, organizan los datos en las estructuras particulares requeridas por cada uno de ellos, refrescando su contenido.

La Metodología de Inmon es más apropiada para sistemas complejos, donde se quiere asegurar la perdurabilidad y consistencia de la información aunque cambien los procesos de negocio de la organización. (11)

1.2.3 Metodología Hefesto

HEFESTO es una metodología propia, cuya propuesta está fundamentada en una muy amplia investigación, comparación de metodologías existentes y experiencias propias en procesos de confección de almacenes de datos.

Hefesto comienza recolectando las necesidades de información de los usuarios y se obtienen las preguntas claves del negocio. Luego se deben identificar los indicadores resultantes de los interrogativos y sus respectivas perspectivas de análisis, mediante las cuales se construirá el modelo conceptual de datos del DW. Después se analizarán los OLTP para señalar las correspondencias con los datos fuentes y seleccionar los campos de estudio de cada perspectiva. Una vez realizado este proceso, se pasará a la construcción del modelo lógico del depósito, precisándose las jerarquías que intervendrán. Por último se definirán los procesos de carga, transformación, extracción y limpieza de los datos fuente. (2)

1.2.4 Metodología UCID

Esta metodología es un híbrido entre las metodologías Hefesto y Kimball, la cual es adoptada por UCID (Unidad de Compatibilización, Integración y Desarrollo de Soluciones Informáticas para la Defensa). La misma permite aprovechar la facilidad que brinda Hefesto al ser una metodología sencilla y entendible con las facilidades que proporciona Kimball al generar artefactos para la documentación a la hora de realizar el DW, además de un conjunto de actividades que describe la metodología ascendente de Kimball. Incluye también algunas actividades y artefactos del Centro de Tecnología de Gestión de Datos DATEC (Data Technology Center) (12)

Sus principales características son:

- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- Se basa en los requerimientos del usuario, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.
- Reduce la resistencia al cambio, ya que involucra al usuario final en cada etapa para que tome decisiones respecto al comportamiento y funciones del DW.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- Es independiente de las herramientas que se utilicen para su implementación.
- Es independiente de las estructuras físicas que contengan el DW y de su respectiva distribución.
- Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.
- Se aplica tanto para almacén de datos como para mercado de datos.

A continuación se exponen las fases de desarrollo de la metodología UCID:

1. Análisis de requerimientos.

- Identificar necesidades de información.
- Identificar perspectivas e indicadores.

Roles: Jefe de línea, Analista, Arquitecto de datos.

Artefactos generados: Modelo conceptual, Acta de validación.

2. Análisis de las fuentes de datos.

- Determinar estado general de los sistemas fuentes.
- Conformar indicadores.
- Establecer correspondencias.

- Definir nivel de granularidad.

Roles: Arquitecto de datos.

Artefactos generados: Registro de sistemas fuentes, Especificación de indicadores, Modelo conceptual ampliado.

3. Modelado del almacén de datos.

- Definir modelo lógico para el almacén de datos.
- Identificar y diseñar dimensiones y hechos.
- Diseñar modelo físico.

Roles: Arquitecto de datos.

Artefactos generados: Modelo de datos, Diccionario de datos.

4. Integración de datos.

- Mapeo de datos.
- Cargas incrementales de datos (Diseño del sistema de extracción, transformación y carga).
- Automatización del sistema ETL.

Roles: Arquitecto de datos.

Artefactos generados: Mapa lógico de datos, Transformaciones, Trabajos.

5. Representación de información.

- Diseñar dimensiones y cubos de información.
- Configurar herramientas de análisis.

Roles: Arquitecto de datos, Analista.

Artefactos generados: Esquemas, Manual de usuario.

6. Pruebas.

Propuesta de pruebas a realizar:

- Pruebas de rendimiento.
- Pruebas de carga.
- Pruebas de estrés.

Roles: Arquitecto de datos, Analista, Asegurador de calidad, Revisor técnico, Probador.

Artefactos generados: Lista de no conformidades, Plan de revisiones, Plan de pruebas.

Después de un estudio de las metodologías más usadas para la construcción del almacén de datos se escoge la metodología creada en UCID.

La metodología de Inmon plantea ir de lo más general a lo más específico, o sea construir un almacén de datos del cual se nutrirán posteriormente los mercados de datos que se crean en los departamentos, razón por la cual no es factible para desarrollar este proyecto. Sin embargo Kimball es ideal para proyectos de poco tiempo, debido a que plantea que se debe crear por cada departamento un conjunto de mercados de datos independientes orientados a los temas que estén relacionados con él. En este caso la metodología de Kimball se adapta mejor a las necesidades del proyecto.

Por otro lado la metodología HEFESTO a pesar de ser sencilla y entendible no define de forma explícita los artefactos que se generan en sus fases. Es por eso que se selecciona la metodología UCID para la implementación del mercado de datos, ya que esta integra los factores positivos de la metodología HEFESTO y Kimball de manera que se satisfagan los requerimientos del proyecto.

1.3 Procesamiento de Transacciones en Línea (OLTP)

OLTP (On Line Transaction Processing), representa toda aquella información transaccional que genera la empresa en su accionar diario, además de las fuentes externas con las que puede llegar a disponer. Estas fuentes de información son de características muy disímiles entre sí, en formato, procedencia, función. Entre los OLTP más habituales que pueden existir en cualquier organización se encuentran: archivos de textos, hipertextos, hojas de cálculos, bases de datos transaccionales, informes semanales, mensuales, anuales, entre otros. (13)

1.4 Base de Datos Multidimensional

El análisis de los datos en un tiempo finito ha traído consigo estudios sobre la mejor forma de almacenar y representar estos datos para que puedan ser consultados de una forma más eficiente. El uso del Modelo Multidimensional es una de las aproximaciones más acertadas y seguidas por los especialistas. Las bases de datos multidimensionales, proveen una estructura que permite tener acceso flexible a los datos, para explorar y analizar sus relaciones, y resultados consiguientes. Estas se pueden visualizar como un cubo multidimensional, donde las variables asociadas existen a lo largo de varios ejes o dimensiones, y la intersección de las mismas representa la medida, indicador o el hecho que se está evaluando. (2)
(Ver figura 1.)

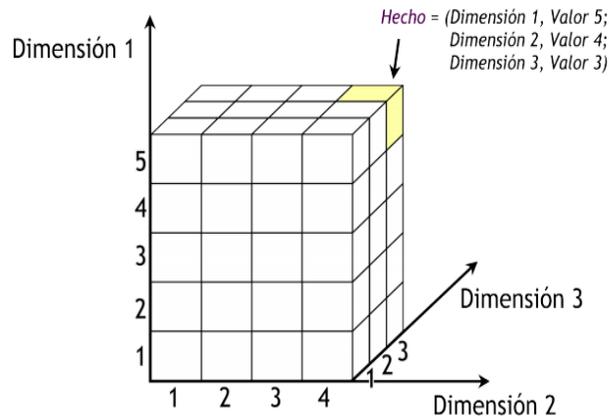


Figura 1 Cubo Multidimensional (2)

Las bases de datos multidimensionales implican tres variantes posibles de modelamiento, que permiten realizar consultas de soporte de decisión: esquema en estrella, esquema copo de nieve y esquema constelación o copo de estrellas.

1.4.1 Esquema en estrella

El esquema en estrella, consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves. El esquema en estrella es el más simple de interpretar y optimiza los tiempos de respuesta ante las consultas de los usuarios. Este modelo es soportado por casi todas las herramientas de consulta y análisis, y los metadatos son fáciles de documentar y mantener, sin embargo es el menos robusto para la carga y es el más lento de construir.

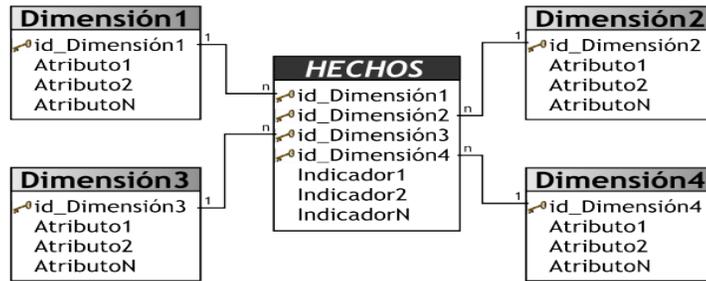


Figura 2 Esquema en estrella (2)

1.4.2 Esquema copo de nieve

Este esquema representa una extensión del modelo en estrella cuando las dimensiones se organizan en jerarquías de dimensiones.

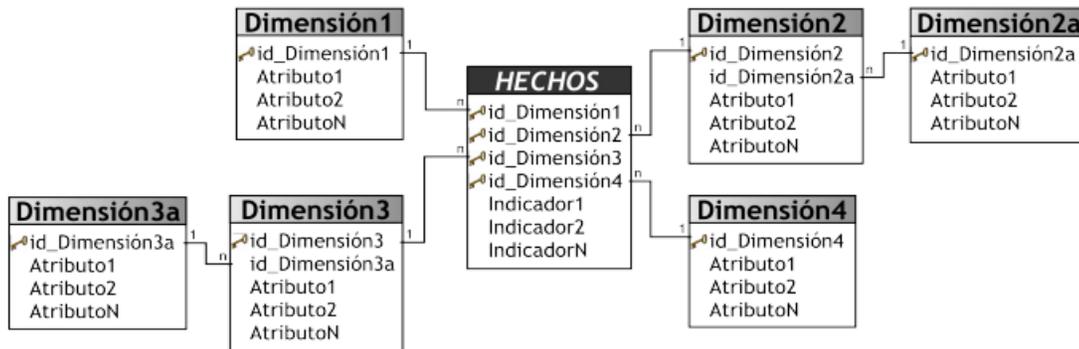


Figura 3 Esquema copo de nieve (2)

Existe una tabla de hechos central que está relacionada con una o más tablas de dimensiones, quienes a su vez pueden estar relacionadas con una o más tablas de dimensiones. Este modelo es más cercano a un modelo de entidad relación, que al modelo en estrella, debido a que sus tablas de dimensiones están normalizadas. Uno de los motivos principales de utilizar este tipo de modelo, es la posibilidad de segregar los datos de las dimensiones y proveer un esquema que sustente los requerimientos de diseño. Además su flexibilidad permite que pueda implementarse después de que se haya desarrollado un esquema en estrella.

1.4.3 Esquema constelación o copo de estrellas

Este modelo está compuesto por una serie de esquemas en estrella, y como se puede apreciar en la figura 5, está formado por una tabla de hechos principal y por una o más tablas de hechos auxiliares. Dichas tablas yacen en el centro del modelo y están relacionadas con sus respectivas tablas de dimensiones. No es necesario que las diferentes tablas de hechos compartan las mismas tablas de dimensiones, ya que las tablas de hechos auxiliares pueden vincularse con solo algunas de las tablas de dimensiones asignadas a la tabla de hechos principal y también pueden hacerlo con nuevas tablas de dimensiones.

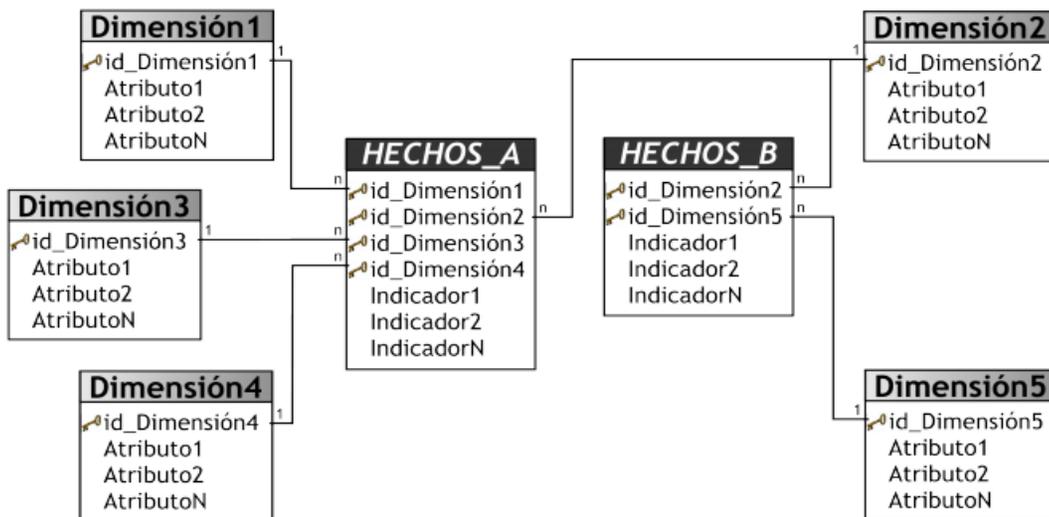


Figura 4 Esquema Constelación (2)

Mediante la investigación realizada se escoge el esquema Constelación para la modelación de la base de datos multidimensional pues su diseño y cualidades son muy similares a las del esquema en estrella. Este último es el más simple de interpretar, pero con la diferencia de que el esquema constelación permite tener más de una tabla de hechos, ya que como se gestionan tres especialidades y los datos de las mismas no se relacionan, se define una tabla de hecho por cada especialidad, por lo cual se podrán analizar más aspectos claves del negocio con un mínimo esfuerzo adicional de diseño. Además, una

misma dimensión puede utilizarse para varias tablas de hechos, en este caso existen dimensiones que comparten los mismos hechos como es el caso de las dimensiones localidad, tiempo, entidad y paciente.

1.5 Extracción, Transformación y Carga (ETL)

Para poder extraer los datos desde los OLTP, para luego manipularlos, integrarlos y transformarlos, y posteriormente cargar los resultados obtenidos en el DW, es necesario contar con algún sistema que se encargue de ello. Precisamente los ETL (Extracción, Transformación y Carga) son los que realizarán esta función. Los ETL extraen datos de las diversas fuentes que se requieran, los transforman para resolver posibles problemas de inconsistencias entre los mismos y finalmente, después de haberlos depurado se procede a su carga en el depósito de datos. (2)

Las funciones específicas de los ETL son las siguientes:

- Extracción de los datos desde las aplicaciones y bases de datos
- Transformación de estos datos para reconciliarlos en todos los sistemas, realizar cálculos o análisis sintáctico de cadenas, enriquecerlos con información de búsqueda externa y, además, adaptarlos al formato preciso por el sistema objetivo.
- Carga de los datos resultantes en las diversas aplicaciones de Inteligencia de Negocios: almacenes de datos históricos generales (almacenes de datos) o almacenes de datos empresariales, almacenes de datos históricos individuales (Mercados de datos).

1.5.1 Herramientas para ETL

Entre las principales herramientas utilizadas para realizar los procesos de Extracción, Transformación y Carga se encuentran:

1.5.1.1 Clover ETL

Es un ambiente de transformación de datos de código abierto basado en Java, para datos estructurados, capaz de funcionar como aplicación independiente (*standalone*) o estar incluida en otra aplicación.

Características principales:

- Realiza transformaciones gráficas basadas en XML para la descripción de los metadatos de los registros.

- Se distribuye bajo la licencia LGPL.
- Soporta 4 diferentes tipos de datos: *string*, *numeric*, *date*, *bytes*.
- Arquitectónicamente, está conceptualizado en unidades lógicas separadas llamadas unidades de transformación que engloban funcionalidades de la transformación e inteligencia, cada una de las cuales puede ser utilizada como componente de una aplicación independiente en otras aplicaciones y servicios. Cada componente corre como un hilo de ejecución separado, creando un ambiente más tolerante a fallas. (3)

1.5.1.2 ENHYDRAOCTOPUS

Es una herramienta de ETL para transformaciones de datos por conexión JDBC [JDBC, 2007]. Octopus solo soporta fuentes de datos que vengan con el manejador JDBC, incluye también drivers especiales que permiten la conectividad con archivos CSV, XML, MS-SQL y archivos de propietarios. Octopus utiliza archivos XML para cargar los trabajos, así como para definir los parámetros de las transformaciones dadas. (14)

Características principales:

- La característica principal de Octopus es el requerimiento de que para acceder a las fuentes de datos que lo utilicen debe existir un manejador JDBC disponible.
- Herramienta poderosa capaz de: normalizar datos, crear llaves artificiales, tablas y llaves primarias.
- Todos los trabajos que corren en Octopus son independientes de los proveedores de bases de datos.

1.5.1.3 KETTLE

Conocido actualmente como *Pentaho Data Integration*, incluye un conjunto de herramientas para realizar ETL. Uno de sus objetivos es que el proceso de ETL sea fácil de generar, mantener y desplegar. Se compone de 4 herramientas: (15)

- SPOON: permite diseñar de forma gráfica la transformación ETL.
- PAN: ejecuta un conjunto de transformaciones diseñadas con SPOON, conocidas como trabajos (*jobs*), creando dependencias entre dichas transformaciones.
- CHEF: permite mediante una interfaz gráfica, diseñar la carga de datos incluyendo un control de estado de los trabajos.
- KITCHEN: permite ejecutar los trabajos batch diseñados con Chef.

Las características principales son:

- Tiene una interfaz visual con indicadores de las transformaciones.
- Es una aplicación escrita en Java con algunas características avanzadas escritas en Java Script.
- Basado en metadatos.
- Como soporte se encuentran los foros de Pentaho y la comunidad Pentaho.
- Con respecto a la escalabilidad, soporta la arquitectura de procesamiento en paralelo para distribuir las tareas de ETL a través de múltiples servidores.

Para el proceso de integración de datos del Mercado de datos se adopta la herramienta Kettle, pues la misma funciona en distintos sistemas operativos como Windows, Unix y Linux. Ofrece una licencia pública GPL y soporta Oracle, DB2, SQL Server y Sybase así como MySQL, PostgreSQL, Hypersonic, FireBird SQL e Ingres. Además existe una experiencia previa de trabajo con la herramienta en la universidad.

1.6 OLAP

OLAP (procesamiento analítico en línea). Es una solución cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Para ello utiliza estructuras multidimensionales (o Cubos OLAP) que contienen datos resumidos de grandes bases de datos o Sistemas Transaccionales (OLTP). Los cubos, las dimensiones y las jerarquías son la esencia de la navegación multidimensional del OLAP. Al describir y representar la información en esta forma, los usuarios pueden navegar intuitivamente en un conjunto complejo de datos.

Sin embargo, el solo describir el modelo de datos en una forma más intuitiva, hace muy poco para ayudar a entregar la información al usuario más rápidamente. (16)

La razón de usar OLAP para las consultas es la velocidad de respuesta. En este modelo los datos son vistos como cubos los cuales consisten en categorías descriptivas (dimensiones) y valores cuantitativos (medidas). El modelo multidimensional de datos simplifica a los usuarios formular consultas complejas, arreglar datos de un reporte y cambiar de datos resumidos a detallados.

Se pueden definir varios tipos de OLAP, dependiendo de las técnicas que se utilicen a la hora de obtener los datos, la forma en la que están estructurados, etc. Dentro de estos tipos se encuentran herramientas que permiten realizar análisis de los datos:

- **ROLAP**

Herramientas ROLAP (*Relational On-line Analytical Process*). Son herramientas OLAP que crean vistas multidimensionales extrayendo los datos de bases de datos relacionales. Estas herramientas simulan los datos multidimensionales usando sofisticadas técnicas de indexación, cachés, metadata. (17)

- **MOLAP**

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

- **HOLAP**

Herramientas HOLAP (*Hybrid On-line Analytical Process*). Permiten un análisis híbrido de la información, unificando lo mejor de los dos tipos anteriores. El análisis HOLAP ayuda a reducir costes de hardware ya que se necesita menos disco que en las bases de datos relacionales. Además, la respuesta de las consultas sobre las bases de datos multidimensionales son más rápidas que sobre las relacionales. Como aspecto negativo, los datos multidimensionales deben ser cargados antes de ser consultados y refrescados cuando se actualizan los datos de la organización.

Para la construcción del mercado de datos se utilizará la herramienta ROLAP teniendo en cuenta que los datos se extraen de una base de datos relacional soportada por un sistema gestor de base de datos PostgreSQL, permitiendo el uso total de la seguridad e integridad de los datos resultando escalable para grandes volúmenes. Este soporta técnicas de optimización de accesos para acelerar las consultas y se accede directamente a los datos del mercado de datos y además que se utilizara Mondrian como motor de consultas y este está diseñado para ROLAP.

1.6.1 Herramientas para OLAP

A continuación se describen las características principales de algunos de los proveedores de OLAP más populares.

1.6.1.1 JEDOX PALO

Jedox Palo es un servidor de bases de datos multidimensional capaz de centralizar y administrar casi un número infinito de hojas de cálculo. El sistema opera en tiempo real, soporta la consolidación de jerarquías así como numerosas funciones de inteligencia empresarial y es un servidor de código abierto. Palo es un servidor de datos multidimensional (MOLAP) orientado a celdas, específicamente desarrollado para almacenamiento y análisis de datos en hojas de cálculo. (15)

1.6.1.2 OLAP4J

OLAP4J (Online Analytical Processing for Java) es una interfaz de aplicación para el ambiente Java 2 Platform, Enterprise Edition, que soporta la creación, almacenamiento y administración de datos para una aplicación OLAP. (3)

1.6.1.3 MONDRIAN

Mondrian es un motor de ROLAP desarrollado en Java que permite analizar grandes conjuntos de datos almacenados en un almacén de datos. Posee un alto desempeño y análisis interactivo de volúmenes de información de cualquier tamaño, además explora dimensionalmente los datos y realiza cálculos avanzados utilizando el lenguaje MDX (multidimensional expressions o expresiones multidimensionales). (18)

Para acceder a las funcionalidades que brinda Mondrian se requiere de un cliente, es el caso de JPivot o JRubik.

- **JPivot:** Es una librería de Java Server Pages (JSP) personalizados que presentan tablas OLAP y permite realizar operaciones tales como Drill Down y Drill Up además de consultas OLAP por medio del lenguaje MDX. Está diseñado para trabajar con motores OLAP como el de Mondrian por lo cual no utiliza las interfaces de programación de aplicaciones o APIs (*Application Programming Interface*) de Mondrian directamente sino que implementa su propio modelo OLAP.
- **JRubik:** Está basado en los componentes de JPivot y también se puede conectar a fuentes OLAP basadas en Mondrian. Las consultas son realizadas mediante el lenguaje MDX y cuenta con componentes similares a los de JPivot, solo que su interfaz no está basada en JSP sino en Swing, por lo que es una aplicación Desktop.

1.6.1.4 SCHEMA WORKBENCH (KETTLE)

Es un entorno visual para el desarrollo y prueba de cubos OLAP Mondrian. Esta herramienta se utiliza para la creación de los archivos XML que se usan para la construcción de los cubos. Además permite la ejecución de consultas MDX contra el esquema y la base de datos. (3)

Como parte del estudio realizado se escoge la herramienta para OLAP Mondrian, por su alto desempeño, análisis interactivo de grandes o pequeños volúmenes de información y exploración dimensional de los datos. Para la construcción de los cubos se utilizará la herramienta SCHEMA WORKBENCH.

1.6.1.5 Lenguaje MDX

MDX es un acrónimo de Multidimensional Query Expression. El MDX es en los sistemas OLAP el equivalente al SQL en los sistemas gestores de bases de datos relacionales. Eso significa que es el lenguaje a través del cual se puede explotar la información que reside en los motores OLAP y satisfacer las consultas analíticas. Además, permite realizar consultas con una sintaxis parecida al lenguaje natural y tiene funciones y fórmulas que lo hacen muy potente para el análisis de datos. (19)

Las expresiones multidimensionales son la vía para acceder a la información contenida en los cubos multidimensionales de los almacenes de datos. Permiten crear conjuntos de datos, miembros calculados, fórmulas definidas por el usuario, e incluso trabajar con fórmulas y rangos dinámicos.

1.7 Sistemas Gestores de Bases de Datos

Entre los principales gestores de bases de datos se encuentran:

1.7.1 Oracle

Oracle es un Sistema Gestor de Bases de Datos (SGBD) con características objeto-relacionales, que pertenece al modelo evolutivo de SGBD. Sus características principales son las siguientes:

- Entorno cliente/servidor.
- Gestión de grandes bases de datos.
- Usuarios concurrentes.
- Alto rendimiento en transacciones.
- Sistemas de alta disponibilidad.
- Disponibilidad controlada de los datos de las aplicaciones.
- Adaptación a estándares de la industria, como SQL-92.
- Gestión de la seguridad.
- Autogestión de la integridad de los datos.
- Opción distribuida.
- Portabilidad.
- Compatibilidad.
- Conectabilidad.
- Replicación de entornos. (20)

1.7.2 MySQL

MySQL es un sistema de gestión de bases de datos relacional, licenciado bajo la GNU (*General Public License*). El software MySQL tiene una doble licencia. Los usuarios pueden elegir entre usar el software MySQL como un producto Open Source bajo los términos de la licencia GNU o pueden adquirir una licencia comercial estándar de MySQL AB. Su diseño multihilo le permite soportar una gran carga de forma muy eficiente. Las principales características de este gestor de bases de datos son las siguientes:

- Aprovecha la potencia de sistemas multiprocesador, gracias a su implementación multihilo.
- Soporta gran cantidad de tipos de datos para las columnas.
- Dispone de API's en gran cantidad de lenguajes (C, C++, Java, PHP, etc.).
- Gran portabilidad entre sistemas.
- Soporta hasta 32 índices por tabla.
- Gestión de usuarios y contraseñas, manteniendo un buen nivel de seguridad en los datos. (21)

1.7.3 PostgreSQL

PostgreSQL es un sistema de gestión de bases de datos objeto-relacional ya que incluye características de la orientación a objetos, como puede ser la herencia, tipos de datos, funciones, restricciones, disparadores, reglas e integridad transaccional. Es el sistema de gestión de bases de datos de código abierto más potente del mercado y en sus últimas versiones no tiene nada que envidiarle a otras bases de datos comerciales. (21)

La última serie de producción es la 9.1, sus características técnicas la hacen uno de los gestores de bases de datos más potentes y robustos del mercado. Su desarrollo comenzó hace más de 16 años, y durante este tiempo, estabilidad, potencia, robustez, facilidad de administración e implementación de estándares han sido las características que más se han tenido en cuenta durante su desarrollo. PostgreSQL funciona muy bien con grandes cantidades de datos y una alta concurrencia de usuarios accediendo a la vez al sistema. (22)

Principales características de este gestor de bases de datos:

- Implementación del estándar (lenguaje) SQL92/SQL99.
- Soporta distintos tipos de datos: además del soporte para los tipos base, también soporta datos de tipo fecha, monetarios, elementos gráficos, datos sobre redes (MAC, IP), cadenas de bits, etc. También permite la creación de tipos propios.
- Incorpora una estructura de datos array.
- Incorpora funciones de diversa índole: manejo de fechas, geométricas, orientadas a operaciones con redes, etc.
- Permite la declaración de funciones propias, así como la definición de disparadores.
- Soporta el uso de índices, reglas y vistas.
- Incluye herencia entre tablas (aunque no entre objetos, ya que no existen), por lo que a este gestor de bases de datos se le incluye entre los gestores objeto-relacionales.
- Permite la gestión de diferentes usuarios, como también los permisos asignados a cada uno de ellos.
- Soporta casi toda la sintaxis SQL (incluyendo sub consultas, transacciones, tipos y funciones definidas por el usuario).
- Cuenta con un amplio conjunto de enlaces con lenguajes de programación como por ejemplo C, C++, Java, Python, PHP y otros.

Después del estudio realizado sobre los distintos sistemas gestores de base de datos se utilizará el PostgreSQL teniendo en cuenta que posee licencia GPL. Está disponible en múltiples plataformas, cuenta con varias herramientas gráficas de alta calidad para administrar las bases de datos (pgAdmin, pgAccess) y para hacer diseño de bases de datos (Tora, Data Architect), y contiene un amplio conjunto de enlaces con lenguajes de programación como por ejemplo C, C++, Java, Python, PHP y otros. Estos aspectos lo convierten en la herramienta ideal para el desarrollo del mercado de datos para el Sistema de Gestión Estomatológica alasSiGEST.

1.8 Herramientas CASE

Para la modelación de los distintos diagramas las herramientas más utilizadas son:

1.8.1 Rational Rose

Rational Rose es la herramienta CASE que comercializan los desarrolladores de UML (Unified Modeling Lenguaje) y que soporta de forma completa la especificación del UML. Propone la utilización de cuatro tipos de modelo para realizar un diseño del sistema, utilizando una vista estática y otra dinámica de los modelos del sistema, uno lógico y otro físico. Permite crear y refinar estas vistas creando de esta forma un modelo completo que representa el dominio del problema y el sistema de software. (3)

1.8.2 Visual Paradigm

Visual Paradigm para UML 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. (23)

1.8.3 Enterprise Architect

Enterprise Architect es una herramienta de construcción y modelado de software de alto rendimiento con una trazabilidad completa desde los requisitos iniciales hasta las decisiones de diseño de software, Enterprise Architect provee el tipo de visualización, colaboración eficiente y robusta requerida en los entornos de desarrollo de software que actualmente son altamente demandantes. Como una solución de modelado verdaderamente ágil, provee una sobrecarga de instalación baja, un rendimiento brillante y una interfaz intuitiva. Es una herramienta de análisis de negocio y UML orientada a objetos, provee el límite competitivo para el desarrollo de software, administración de proyecto, administración de requerimientos y análisis de negocio. (24)

A través de la investigación realizada de algunas de las herramientas CASE se selecciona Visual Paradigm para el modelado y construcción del mercado de datos teniendo en cuenta que la Suite Visual

Paradigm contiene entre sus productos la herramienta Database Visual ARCHITECT, la cual es una herramienta para crear diagramas entidad-relación, permite el modelado visual de objetos y modelos de datos, hacer diagramas E/R y diagramas de clase a partir de la ingeniería inversa a bases de datos. Soporta varios motores de bases de datos entre ellas: Oracle, DB2, Cloudscape/Derby, Sybase Adaptive Server Enterprise, Sybase SQL Anywhere, Microsoft SQL Server, PostgreSQL y MySQL. Además la Suite Visual Paradigm ofrece licencia pública.

1.9 Plataforma para el desarrollo del Mercado de Datos

La plataforma **Open Source Pentaho Business Intelligence** cubre las amplias necesidades de Análisis de los Datos y de los Informes empresariales. Las soluciones de Pentaho están escritas en Java y tienen un ambiente de implementación también basado en Java. Eso hace que Pentaho sea una solución muy flexible para cubrir una amplia gama de necesidades empresariales – tanto las típicas como las sofisticadas y específicas al negocio. (25)

1.9.1 Módulos de la Plataforma Pentaho BI

- **Reporting:** es el módulo de los informes y ofrece la solución adecuada a las necesidades de los usuarios. Pentaho Reporting permite generar informes de gran capacidad, la distribución de los resultados del análisis en múltiples formatos - todos los informes incluyen la opción de imprimir o exportar a formato PDF, XLS, HTML y texto. Los reportes Pentaho permiten también programación de tareas y ejecución automática de informes con una determinada periodicidad.
- **Análisis:** suministra a los usuarios un sistema avanzado de análisis de información. Con uso de las tablas dinámicas (*pivot tables, crosstabs*), generadas por Mondrian, el usuario puede navegar por los datos, ajustando la visión de los mismos, los filtros de visualización, añadiendo o quitando los campos de agregación. Además, con el *Microsoft Excel Analysis Services*, se puede analizar los datos dinámicos en Microsoft Excel (usando la conexión a OLAP server Mondrian).
- **Integración de Datos:** se realiza con una herramienta Kettle ETL (*Pentaho Data Integration*) que permite implementar los procesos ETL.

La suite Open Source Pentaho Business Intelligence ha sido seleccionada por contar con la herramienta Mondrian dentro del módulo de Análisis y Kettle dentro del módulo Integración de Datos, ambas herramientas fueron seleccionadas durante el estudio de las herramientas de integración de datos y las herramientas para OLAP.

Conclusiones del Capítulo

En el capítulo se realizó un análisis sobre los almacenes de datos, sus ventajas y beneficios para el personal médico a la hora de tomar decisiones. Se hizo un estudio del arte sobre los mismos, a nivel internacional y nacional, haciendo referencia en este último a los realizados en la UCI, los cuales han sido usados para solucionar diversos problemas de la sociedad y aumentar el nivel de investigación y desarrollo en esta área. Se escogió como herramienta para el ETL Kettle, pues la misma funciona en distintos sistemas operativos y soporta robustos gestores de bases de datos, para los procesos OLAP se eligió el Mondrian, por su alto desempeño, análisis interactivo de grandes o pequeños volúmenes de información y exploración dimensional de los datos. Para la construcción de los cubos se utilizó la herramienta SCHEMA WORKBENCH. Como gestor de base de datos se escogió PostgreSQL porque está disponible en múltiples plataformas, cuenta con varias herramientas gráficas de alta calidad y cuenta con un amplio conjunto de enlaces con lenguajes de programación. Como herramienta Case el Visual Paradigm y como plataforma para el desarrollo del mercado de datos Pentaho.

Capítulo 2: Análisis y Diseño de la estructura del Mercado de Datos.

En el capítulo se realiza el análisis y diseño según los pasos que describe la metodología UCID. Se elabora un modelo conceptual del mercado de datos a partir de los indicadores y perspectivas obtenidas del negocio. Se definen los hechos, dimensiones y medidas por los que quedará conformado el modelo de datos y se confecciona el modelo lógico del sistema, de forma que quede conformado el diseño del Mercado de datos. Además, se define la arquitectura del mercado de datos y las políticas de respaldo y seguridad.

2.1 Fase 1: Análisis de los requerimientos

2.1.1 Requerimientos de Información.

El objetivo principal es obtener e identificar las necesidades de información clave, que es esencial para llevar a cabo las metas y estrategias del negocio, lo que facilitará una eficaz y eficiente toma de decisiones. Se comienza recolectando las necesidades de información de los directivos del CeNaEst y se obtienen las preguntas claves del negocio. Para ello se les realizó una entrevista con el propósito de conocer cuáles eran sus necesidades reales, los resultados que esperaban y los reportes que considerasen más importantes para su trabajo en el centro. Ellos expresaban la necesidad de que se les facilitara el trabajo con algunos reportes e información de importancia histórica que les es difícil manejar.

Algunos de los resultados obtenidos en la entrevista: (Ver Anexos 2. Requisitos de Información)

- R11 - Cantidad de pacientes atendidos en Estomatología general integral (EGI) por cada una de las entidades en un tiempo determinado y en una localidad determinada.
- R12 - Cantidad de pacientes atendidos en Ortodoncia por cada una de las entidades en un tiempo determinado y en una localidad determinada.
- R13 - Cantidad de pacientes atendidos en Prótesis por cada una de las entidades en un tiempo determinado y en una localidad determinada.
- R14 - Porcentaje de pacientes atendidos en la especialidad de ortodoncia por entidad.
- R15 - Porcentaje de pacientes atendidos en la especialidad de ortodoncia por localidad.

RI6 - Porcentaje de pacientes atendidos en la especialidad de estomatología general integral (EGI) por localidad.

2.1.2 Requerimientos funcionales

Los requerimientos funcionales representan las capacidades y condiciones que el mercado de datos debe cumplir para dar respuesta a los requerimientos de información de los directivos del CenaEst. A continuación se muestran algunos de los requerimientos funcionales que fueron identificados (Ver Anexos 3. Requerimientos Funcionales):

RF1 - Autenticar usuario.

RF2 - Adicionar roles.

RF3 - Eliminar roles.

RF4 - Adicionar usuarios.

RF5 - Eliminar usuarios.

RF6 - Insertar reportes.

RF7 - Mostrar datos de origen.

2.1.3 Requerimientos no funcionales

Por las peculiaridades de los sistemas de almacenamiento y análisis de datos el usuario final debe tener conocimiento en el trabajo con las herramientas de integración de datos. Los mismos especifican criterios que pueden usarse para juzgar la operación de un sistema. En fin los requisitos no funcionales son características o capacidades que el sistema debe tener o cumplir. A continuación se mostrarán algunos de los requisitos no funcionales (Ver Anexos 4. Requerimientos no funcionales):

RNF1 - Usabilidad: mostrar los mensajes, títulos y demás textos que aparezcan en la interfaz del sistema en idioma español.

RNF2 - Usabilidad: navegar en los reportes del Almacén de Datos de manera ágil.

RNF3 - Confiabilidad: garantizar el cumplimiento de actualización de los datos en el almacén.

RNF4 - Eficiencia: el tiempo de respuesta debe ser en tiempo real.

RNF5 - Interfaz: los reportes estadísticos deben contar con una interfaz simple que facilite la interacción usuario-aplicación.

RNF6 - Interfaces Hardware: proporcionar características mínimas de hardware a los servidores.

2.1.4 Identificar Indicadores y Perspectivas

Los indicadores y perspectivas identificadas responden a los requerimientos informacionales descritos en la sección 2.1.1. A continuación se describen los conceptos de cada uno de ellos.

Indicadores: Son en general valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: saldos, promedios, cantidades, sumatorias, fórmulas, etc.

Perspectivas: Se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de responder a las preguntas planteadas. (Ver Anexos 5. Indicadores y Perspectivas)

Cantidad de pacientes atendidos en EGI por cada una de las entidades en un tiempo determinado y en

Indicador

Perspectivas

una localidad determinada.

Perspectivas

Cantidad de pacientes atendidos en Ortodoncia por cada una de las entidades en un tiempo determinado

Indicador

Perspectivas

y en una localidad determinada.

Perspectivas

Cantidad de pacientes atendidos en **Prótesis** por cada una de las **entidades** en un **tiempo** determinado y

Indicador

en una **localidad** determinada.

Perspectivas

Perspectivas

Porcentaje de pacientes atendidos en la **especialidad de ortodoncia** por **entidad**.

Indicador

Perspectivas

Porcentaje de pacientes atendidos en la **especialidad de prótesis** por **localidad**.

Indicador

Perspectivas

Porcentaje de pacientes atendidos en la **especialidad de estomatología general integral (EGI)** por

Indicador

Perspectivas

localidad.

Perspectivas

2.1.5 Modelo Conceptual

Una vez identificados los indicadores y perspectivas se procede a la confección del modelo conceptual de datos del mercado de datos, en el que se reflejan las perspectivas a la izquierda y los indicadores a la derecha (Ver Figura 5), proporcionando una idea precisa y clara del alcance del almacén de datos. Este modelo posee un alto nivel de definición de los datos, lo que permite que pueda ser presentado ante los usuarios y explicado con facilidad.

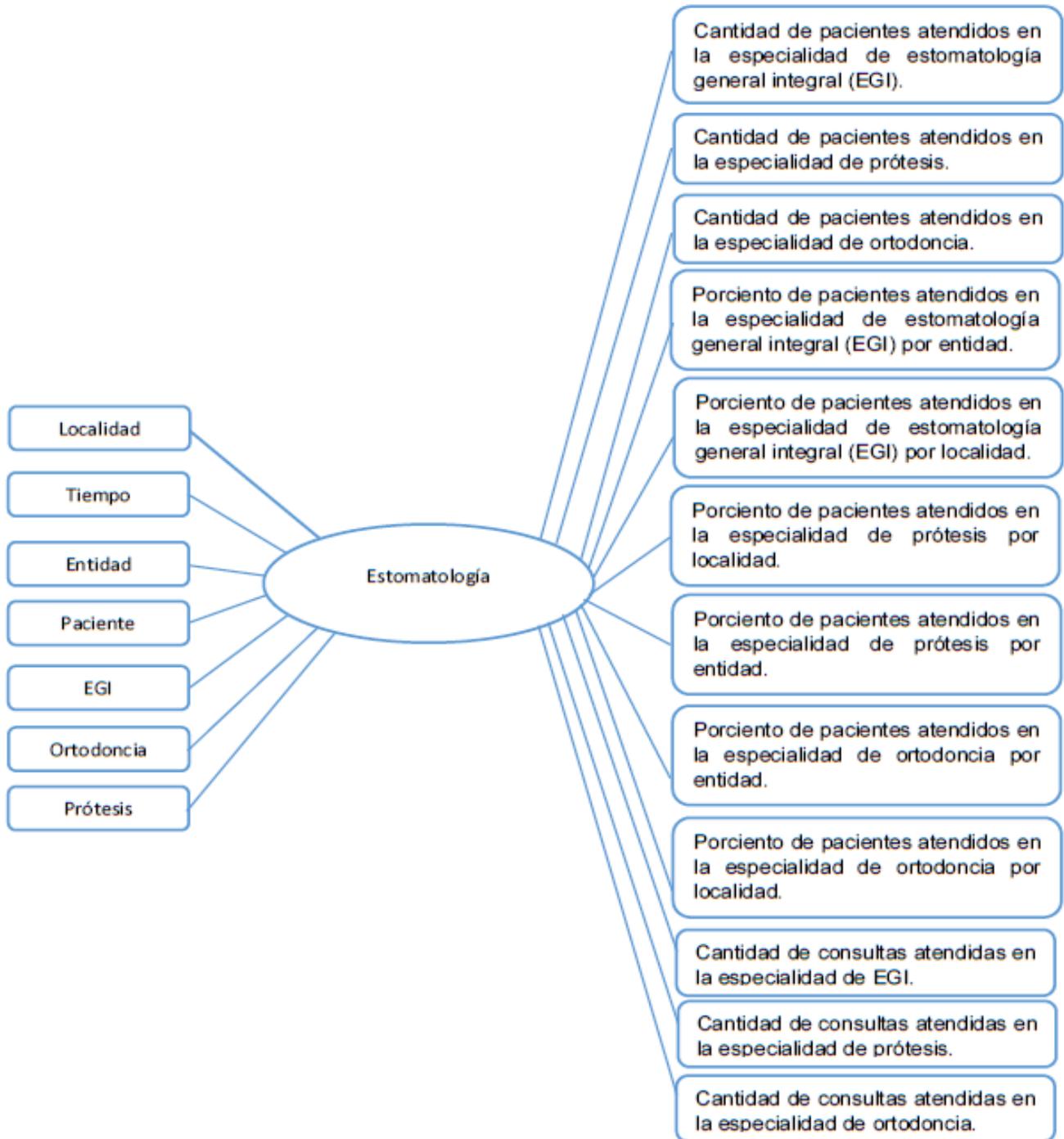


Figura 5 Modelo Conceptual

2.2 Fase 2: Análisis de las fuentes de datos

Para gestionar los datos correspondientes al Sistema de Gestión Estomatológica alasSIGEST se cuenta con una base de datos relacional y como gestor de base de datos se utiliza PostgreSQL en la versión 8.3.8. Al sistema se conectan diario aproximadamente 60 usuarios en el CeNaEst, y los usuarios ejecutan entre 10 y 100 transacciones diarias aproximadamente en condiciones normales, o sea que no hay cargas masivas de datos.

Durante la ejecución de este paso del análisis de los sistemas fuentes, se logró detectar algunos errores que pueden tener cierta repercusión en las decisiones que tomen los directivos del CeNaEst. Uno de los problemas que se detectaron fue la existencia de campos que contienen valores nulos, además de algunas tuplas duplicadas. Para resolver estos problemas se definieron algunas reglas de transformación en los procesos de Extracción, Transformación y Carga del Mercado de Datos.

2.2.1 Correspondencia con los requerimientos

El objetivo de este análisis es examinar los OLTP disponibles que contengan la información requerida, así como también sus características, para poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos.

A continuación se describe cómo serán calculados alguno de los indicadores, con el objetivo de que todos los elementos del modelo conceptual estén correspondidos en los OLTP.

- “Cantidad de pacientes atendidos en la especialidad de estomatología general integral (EGI)”: representa el total de todos los de los pacientes que hayan sido atendidos según una entidad, localidad y un tiempo determinado además de la especialidad EGI.
- “Cantidad de pacientes atendidos en la especialidad de prótesis”: representa el total de todos los pacientes que hayan sido atendidos según una entidad, localidad y un tiempo determinado además de especialidad de prótesis.
- “Cantidad de pacientes atendidos en la especialidad de ortodoncia”: representa el total de todos los de los pacientes que hayan sido atendidos según una entidad, localidad y un tiempo determinado además de especialidad de ortodoncia.

- “Porcentaje de pacientes atendidos en la especialidad de ortodoncia por localidad”: representa el porcentaje de pacientes atendidos en ortodoncias por una localidad determinada.
- “Porcentaje de pacientes atendidos en la especialidad de ortodoncia por entidad”: representa el porcentaje de pacientes atendidos en ortodoncias por una entidad determinada.
- “Porcentaje de pacientes atendidos en la especialidad de estomatología general integral (EGI) por localidad.”: representa el porcentaje de pacientes atendidos en EGI por una localidad determinada.

2.2.2 Campos que integran las perspectivas. Nivel de Granularidad.

Después de haber recogido toda la información necesaria, se presentan los campos principales y de interés de cada perspectiva seleccionada de la base de datos, que ayudarán a consultar los indicadores, en la figura 6 se representa el modelo conceptual con cada uno de los campos o atributos elegidos para cada perspectiva:

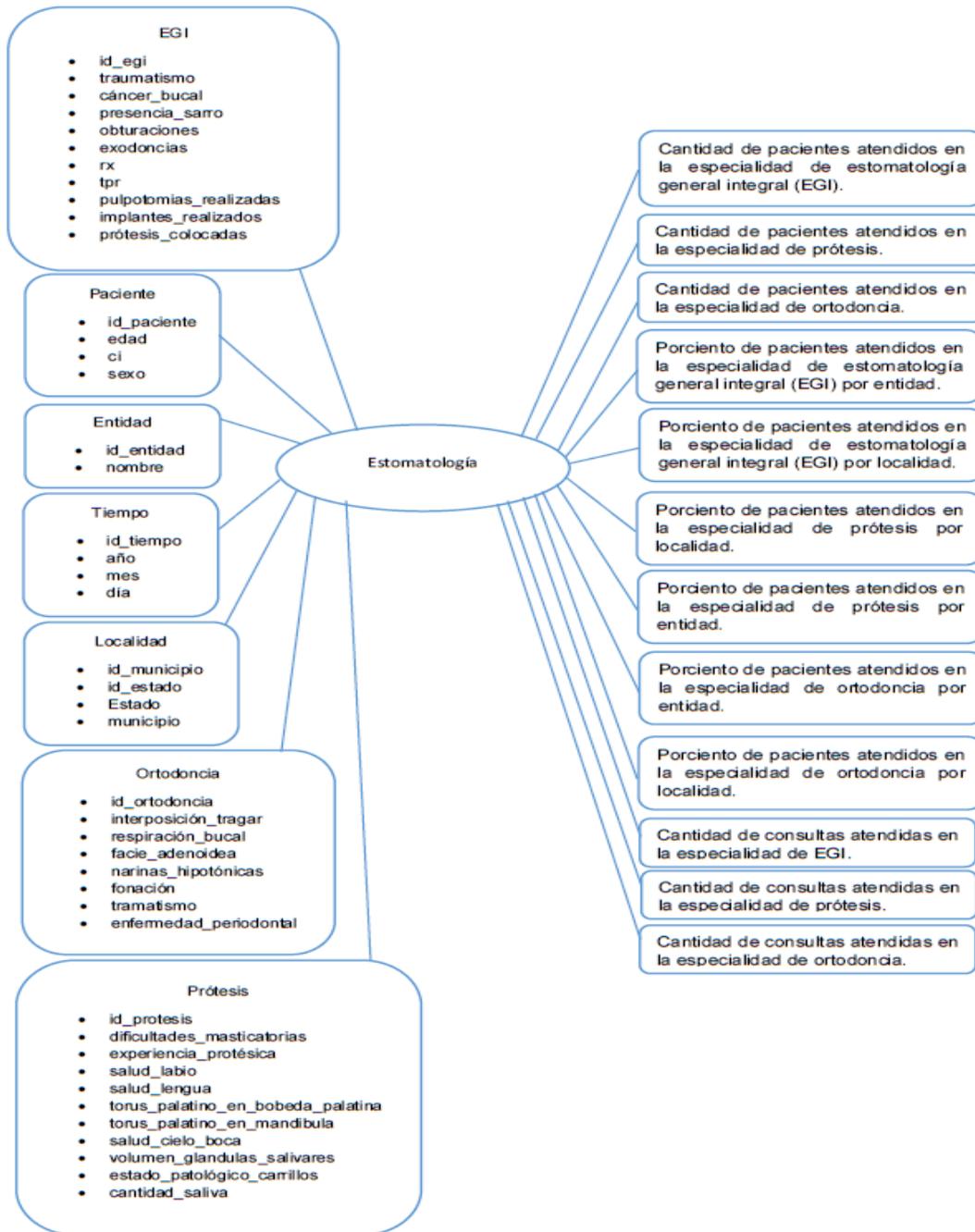


Figura 6 Modelo Conceptual con atributos.

2.3 Fase 3: Modelado del Mercado de datos

2.3.1 Identificación de dimensiones y hechos

Generalmente cada perspectiva definida en el modelo conceptual constituirá una dimensión. Para llevar a cabo este proceso primeramente es necesario determinar el nombre que identificará la tabla de dimensión. Seguidamente se añadirá un campo que constituya la clave principal de dicha tabla y finalmente se redefinirán los nombres de los campos que conformarán la dimensión en caso que no sean lo necesariamente intuitivos. Se definieron las dimensiones:

- d_egi
- d_localidad
- d_entidad
- d_paciente
- d_ortodoncia
- d_protesis
- d_tiempo

Las tablas de hechos contienen los hechos que serán analizados para apoyar el proceso de toma de decisiones. Los hechos son datos instantáneos en el tiempo, que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de dimensiones. Los registros del hecho poseen una clave primaria que está compuesta por las claves primarias de las tablas de dimensiones relacionadas a este. Se definieron las tablas de hechos:

- h_pacientes_egi
- h_pacientes_ortodoncia
- h_pacientes_protesis

Matriz bus o matriz dimensional

La Matriz bus es una herramienta potente para la planificación y la comunicación, pues permite definir la arquitectura general de los datos para el mercado de datos, y determinar el impacto que provocaría durante el desarrollo del sistema un cambio a la solución. En ella se describen las relaciones que existen entre las tablas de hechos y las dimensiones. Las columnas de la matriz representan las dimensiones utilizadas en el mercado de datos y las filas, los hechos identificados. Las celdas sombreadas con una X indican que la columna de dimensión está relacionada con la fila del proceso de negocio. Esto posibilita ver de inmediato cuáles son las dimensiones que merecen una atención especial debido a su participación o relación con múltiples hechos. La figura 7 muestra la matriz bus para el mercado de datos del CeNaEst.

		Dimensiones					
Hechos	dim_localidad	dim_entidad	dim_tiempo	dim_paciente	dim_egi	dim_prótesis	dim_ortodoncia
h_paciente_egi	X	X	X	X	X		
h_paciente_prótesis	X	X	X	X		X	
h_paciente_ortodoncia	X	X	X	X			X

Figura 7 Matriz Bus.

2.3.2 Modelo lógico del Mercado de datos

En este paso se confecciona el modelo lógico de la estructura del mercado de datos. Para obtener la estructura del mercado se selecciona el esquema constelación por todas las ventajas mencionadas en el capítulo anterior. En la figura 8 se representa según el esquema seleccionado el diseño de las dimensiones y las tablas de hechos con sus atributos asociados.



Figura 8 Estructura del Mercado de Datos.

2.4 Arquitectura del Mercado de datos

Los mercados de datos como solución informática que apoyan la toma de decisiones en las entidades que los implementan también necesitan una descripción detallada de la arquitectura. La arquitectura de los mercados de datos es una forma de representar la estructura global de los datos, la comunicación, los procesos y la presentación del usuario final. (26)

La arquitectura general del mercado de datos para el CeNaEst está compuesta por las fuentes de datos (sistema alasSIGEST) y por tres subsistemas fundamentales: el subsistema de integración, el subsistema de almacenamiento y el subsistema de visualización.

En la figura 9 se muestra la arquitectura para el mercado de datos del CeNaEst:

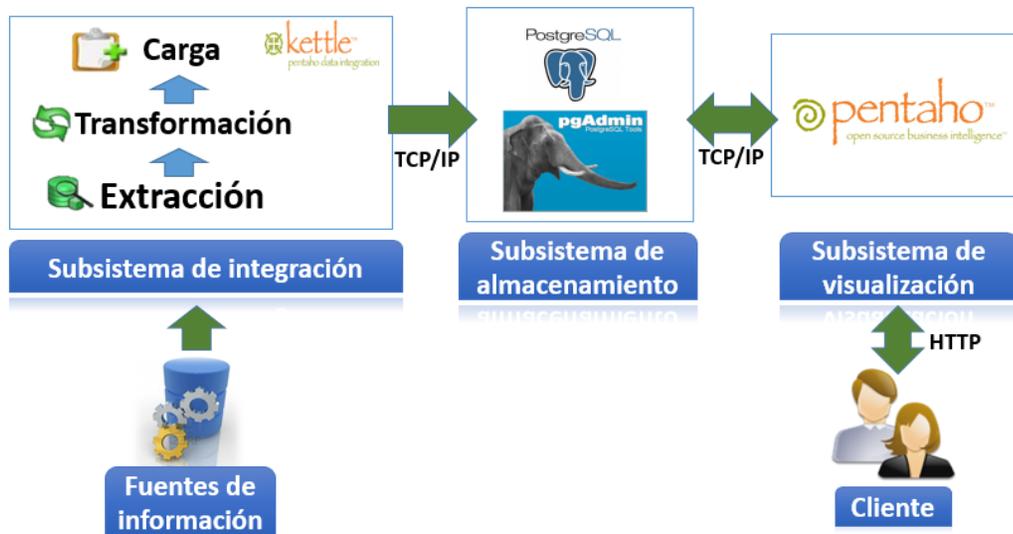


Figura 9 Arquitectura del Mercado de Datos para el CeNaEst.

A continuación se explicarán cada uno de los subsistemas que definen la arquitectura del mercado de datos para el CeNaEst

- Fuente de datos: se refiere al sistema alasSIGEST que constituye el origen de los datos.
- Subsistema de integración: se encarga de integrar, estandarizar y limpiar los datos necesarios del sistema alasSIGEST, con el fin de cargarla hacia el mercado de datos, a través de los procesos de extracción, transformación y carga de los datos.
- Subsistema de almacenamiento: es el encargado de almacenar toda la información correspondiente al área específica del mercado de datos. El almacén estará compuesto por las dimensiones y las tablas de hechos identificadas anteriormente, las cuales contendrán los datos que permitirán a los directivos del CeNaEst tomar decisiones tácticas y estratégicas.
- Subsistema de visualización: tiene como objetivo principal ofrecer la interfaz a los directivos del CeNaEst para consultar la información almacenada en el mercado de datos permitiendo mostrar los reportes que necesitan.

El sistema alasSIGEST proporciona los datos necesarios al subsistema de integración el cual lleva a cabo los procesos de transformar e integrar la información para su almacenamiento en el mercado de datos.

El subsistema de almacenamiento recoge la información transformada (a través de los procesos ETL) por el subsistema de integración para almacenarla en el mercado de datos para el CeNaEst soportado por el SGBD PostgreSQL.

Finalmente el subsistema de visualización muestra a los directivos del CeNaEst la información que necesitan mediante gráficos y tablas a través de la herramienta Pentaho BI Server.

2.5 Políticas de Respaldo y Recuperación

En el Mercado de Datos para el CeNaEst las políticas de respaldo y recuperación se miden en tres aspectos fundamentales:

- Periodicidad de las salvadas: las salvadas de toda la información contenida en la BD se realizarán trimestralmente verificando en todo momento que exista una copia escrita de la información almacenada en el servidor.
- Tablas involucradas: las tablas que se involucran en este proceso son las 7 tablas de dimensiones y las 3 tablas de hechos identificadas en el proceso de análisis.
- Salvadas existentes: a pesar de que actualmente no existen salvadas en esta área se prevé la realización de reemplazos de estas cada tres meses, así como también el chequeo de su estado mensualmente, mediante pruebas de rendimiento y flexibilidad.

2.6 Seguridad del Mercado de Datos

La seguridad en el mercado de datos está dada por los niveles de acceso al sistema, regida fundamentalmente por los permisos y roles que los usuarios tienen a la hora de interactuar con la base de datos y la aplicación. Para la seguridad en la base de datos se decidió definir un rol de Administrador el cual posee total acceso a la base de datos del sistema; y para la seguridad en la aplicación se definieron los roles Administrador y Analista, el primero tiene acceso total para administrar los usuarios, roles y reportes; el segundo tiene acceso de solo lectura a los distintos reportes.

Conclusiones del capítulo

En el capítulo se desarrolló cada uno de los pasos de la metodología UCID. Se obtuvo las preguntas claves del negocio, identificándose los indicadores y perspectivas. Se elaboró el modelo conceptual de datos proporcionando una idea clara del alcance del almacén. También se confeccionó el modelo lógico de la estructura del mercado de datos partir de la identificación de los hechos y las dimensiones. Además, se definió la arquitectura del mercado de datos y las políticas de respaldo y seguridad.

Capítulo 3: Implementación del Mercado de Datos.

En el siguiente capítulo se realizan los procesos de Extracción, Transformación y Carga, se definen las jerarquías de los atributos de las tablas de dimensiones y se diseñarán los cubos multidimensionales que permiten realizar los cruces necesarios para visualizar la información contenida en el Almacén de Datos. En general, se realiza el procesamiento analítico en línea para agilizar las consultas de la información en el almacén, mostrando los datos a los directivos desde diferentes perspectivas. Además se realizarán pruebas de rendimiento y estrés para probar la eficiencia del mercado de datos.

3.1 Reglas de Transformación

Una vez construido el modelo lógico del almacén de datos es necesario realizar el proceso de Extracción, Transformación y Carga (ETL), pero antes de realizar la carga de datos, es conveniente efectuar una limpieza de los mismos, para evitar valores faltantes y anómalos. Al generar los ETL, se debe tener en cuenta cuál es la información que se desea almacenar en el depósito de datos, para ello se pueden establecer restricciones y condiciones adicionales, este proceso se realizará en la herramienta Kettle perteneciente a la suite Pentaho.

Con el fin de lograr la carga de la información deseada en el mercado de datos, se establecerá un conjunto de condiciones adicionales o reglas de transformación teniendo en cuenta el análisis realizado en el capítulo anterior. Estas reglas deben analizarse cuidadosamente para evitar pérdidas significativas de datos.

Teniendo en cuenta los problemas detectados en el análisis realizado a la fuente de datos se establece las siguientes reglas de transformación:

- A los campos que contengan valores nulos se les asignará el valor: “No Existe”, siempre que el tipo de dato en el destino lo permita.
- No pueden existir tuplas duplicadas.

3.2 Estándares para el diseño del proceso ETL

Antes de comenzar a diseñar los procesos de extracción, transformación y carga que se realizarán a los datos, fue necesario establecer un estándar para organizar el trabajo. Para dar cumplimiento a esto, se determinó nombrar a todas las transformaciones con el nombre que tiene la tabla que constituye el destino de los datos transformados, además se agregará al nombre el prefijo “Transf_”.

A modo de seguridad se creó un repositorio o catálogo donde se almacenarán todas las transformaciones y trabajos diseñados, además serán almacenados en el formato que provee la herramienta Kettle Pentaho Data Integration (ktr y kjb).

3.3 Procesos de Extracción, Transformación y Carga (ETL)

El proceso de la dimensión d_especialidad (Ver Anexos 1. Carga de las dimensiones y hechos del Mercado de Datos) comienza insertando en la dimensión mediante una consulta SQL el id y el nombre de la especialidad, para así obtener el nombre de la especialidad que se puede atender a un paciente. Para obtener dichos atributos se procede de la siguiente manera: (Ver Figura 10)



Figura 10 Carga de la dimensión d_especialidad

Para este proceso se utilizaron varios componentes como el de Entrada Tabla el cual permite mediante una consulta SQL obtener los datos que se desean agregar al almacén de datos. Seguidamente se utiliza el componente Selecciona/Renombra valores el cual se encarga del mapeo de datos y para finalizar mediante el componente Insertar/Actualizar se realiza una búsqueda de la fila en la tabla usando una o más llaves de búsqueda, si no la encuentra inserta la fila, si la encuentra y contiene la misma información no hace nada, y si no son iguales la actualiza.

Otro ejemplo de transformación es la carga de la dimensión d_paciente (Figura 11), la cual tiene la entrada de datos mediante una consulta SQL en el componente Entrada Tabla. Se utiliza el componente

Filas Únicas para evitar que haya tuplas repetidas, además se inserta el componente Información del Sistema el cual va a dar la fecha del sistema para después con el componente Valor Java Script Modificado transformar los datos y que lleguen al almacén en el formato deseado, además de hacer cálculos como el utilizado para conocer la edad del paciente. Seguidamente se hace el mapeo de los datos, se insertan y actualizan los datos.

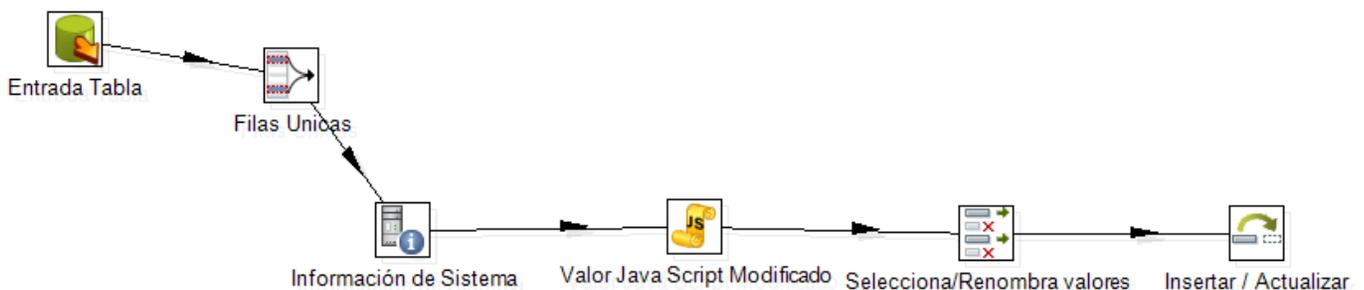


Figura 11 Carga de la dimensión d_paciente

Después de haber cargado todos los datos en las tablas de dimensiones, se procede a insertar los datos en las tablas de hechos, que tienen el objetivo de relacionar todas las dimensiones del almacén de datos. Para esto se realiza una búsqueda en la base de datos origen mediante una consulta SQL con el componente Entrada Tabla para extraer los datos a necesarios, después se utiliza el componente Búsqueda de valor en base de datos, el cual permite comparar el id de la base de datos origen con el del mercado de datos y de esta manera devolver un valor en la tabla de búsqueda, se hace esto sucesivamente hasta que se verifiquen todas las dimensiones. A continuación se utiliza el componente Ordenar campos, mediante el cual los campos por los que se va a agrupar estén ordenados. Seguidamente se agrupan los datos teniendo en cuenta ciertas reglas como; suma, promedio, primer valor nulo, donde el objetivo es organizar los valores o calcular los valores en un grupo definido de campos. Para terminar la transformación se emplea el componente Insertar/Actualizar que posibilita insertar los valores en la tabla de hechos. (Ver Figura 12)

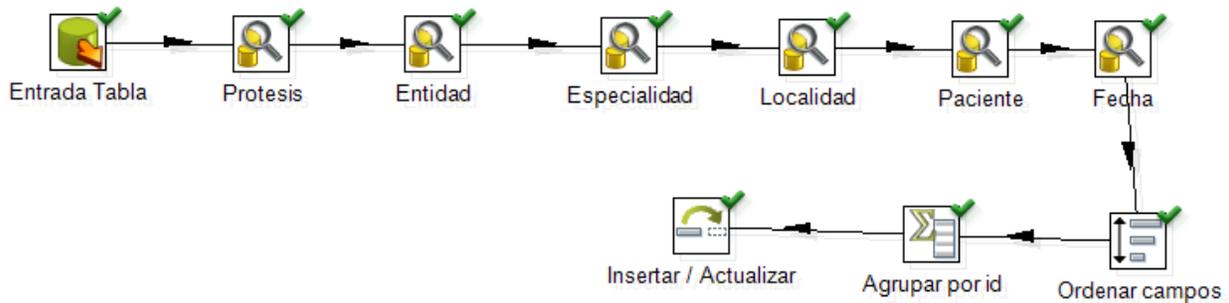


Figura 12 Carga de la tabla de hechos “hechos_pacientes_protesis”

Luego de haber culminado la carga de las tablas de hechos se crea un trabajo (*job*), en el que se ejecutan todas las cargas de las dimensiones y de las tablas de hechos del almacén de datos. Este proceso utiliza el componente (*Start*) el cual da comienzo a la extracción, transformación y carga. Seguidamente se ejecuta el componente trabajo (Ver Figura 13). El mismo realiza la carga de todas las dimensiones del almacén y a continuación se realiza la carga para las tablas de hechos.

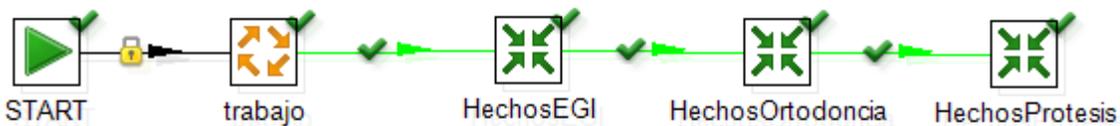


Figura 13 Ejecución del trabajo

3.3.1 Automatización del sistema ETL.

Teniendo en cuenta la frecuencia de actualización de los datos en el CeNaEst, se propone la automatización de los procesos de extracción, transformación y carga semanalmente. El objetivo de esta automatización es añadir al depósito aquellos datos nuevos que fueron generados después de la última actualización. La misma se llevó a cabo mediante la programación del trabajo diseñado.

3.4 Procesos de Análisis de la Información contenida en el Mercado de Datos

3.4.1 Jerarquías de las dimensiones.

Para el esquema constelación, se deberán especificar las jerarquías que existirán dentro de cada tabla de dimensión, teniendo siempre presente cuál es el objetivo de las mismas. Una jerarquía representa una relación lógica entre dos o más atributos dentro de una misma dimensión. Sus principales características son:

- Pueden existir varias en una misma dimensión.
- Están compuestas por dos o más niveles.
- Se tiene una relación “1-n” entre atributos consecutivos de un nivel superior y uno inferior.

La principal ventaja de manejar jerarquías, reside en poder analizar los datos desde su nivel más general al más detallado y viceversa.

En la figura 14 y figura 15 se muestran las jerarquías de los atributos de algunas de las dimensiones que componen el mercado de datos:



Figura 14 Jerarquía de los atributos de la dimensión `d_localidad`. Figura 14



Figura 15 Jerarquía de los atributos de la dimensión d_tiempo.

3.4.2 Medidas.

Las medidas son los valores numéricos calculables que se muestran en la consulta al Almacén de Datos. Luego de establecer las jerarquías de cada una de las dimensiones se elaboran las medidas o indicadores en busca de las respuestas de lo que desean conocer los directivos del CeNaEst. Para ello se utilizan las consultas MDX (*multidimensional expressions* o expresiones multidimensionales) que se ejecutan sobre el mercado de datos para obtener los indicadores que contribuirán a la toma de decisiones.

Desde la base de datos multidimensional se cargan las dimensiones y las tablas de hechos para calcular las medidas. Utilizando la herramienta Schema Workbench se diseñan los cubos de datos (uno por cada tabla de hechos). Se agregan las medidas y se formulan las consultas necesarias para obtener las cantidades de pacientes por cada especialidad y los porcentos en dependencia de las necesidades de los directivos del CeNaEst.

La figura 16 muestra el diseño de los cubos multidimensionales en la herramienta Schema Workbench:

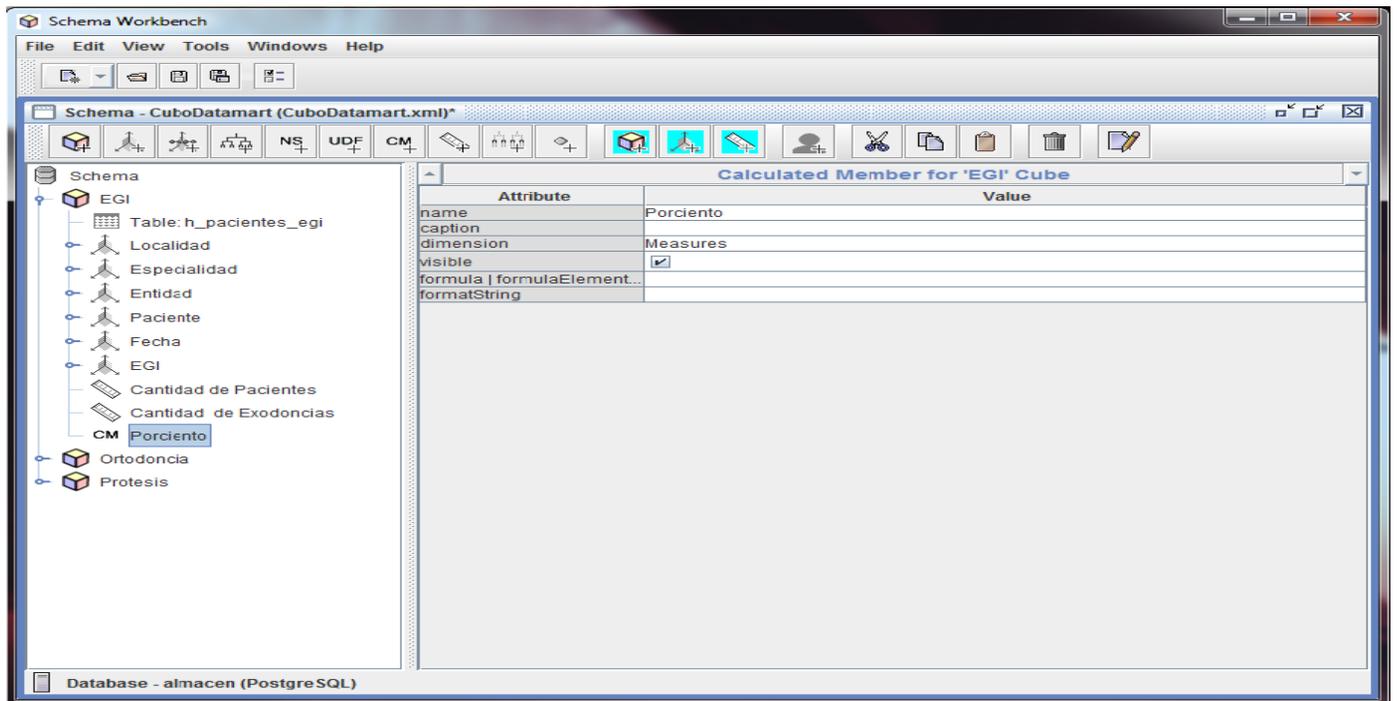


Figura 16 Diseño de los cubos multidimensionales.

3.4.3 Consultas MDX

Para calcular los datos numéricos que serán visualizados en los resultados se utilizarán las consultas MDX (Ver Figura 17). Una de las ventajas de las expresiones multidimensionales es que permiten consultar objetos multidimensionales y devolver un conjunto de celdas que contenga los datos del cubo.

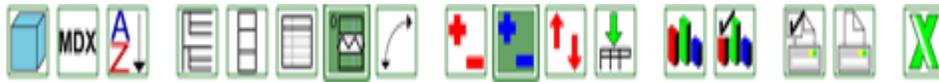
MDX Query Editor



Figura 17 Consultas MDX.

3.4.4 Resultados de la herramienta mondrian. Pentaho BI Server.

El cubo es almacenado en un archivo XML para luego ser cargado en la herramienta Mondrian, escogida por su alto desempeño para realizar el procesamiento analítico en línea. El cliente JPivot estará disponible a través de la web mediante el servidor Pentaho BI Server, donde se muestra un menú con las opciones necesarias para obtener la información desde diferentes perspectivas, como por ejemplo filtrar la información según los campos que se desean, ajustar la visualización de los datos y añadir o quitar campos que no sean necesarios visualizar. Los resultados serán mostrados mediante gráficos o tablas, los cuales pueden ser exportados en formato PDF o EXCEL.



				Medidas
Entidad	Fecha	Localidad	Obturaciones	Cantidad de Consultas
+ Entidad	+ Anno	+ Estado	No	45

Slicer:

Figura 18 Dimensión Localidad con las medidas e indicadores según el filtrado.



Figura 19 Información Gráfica

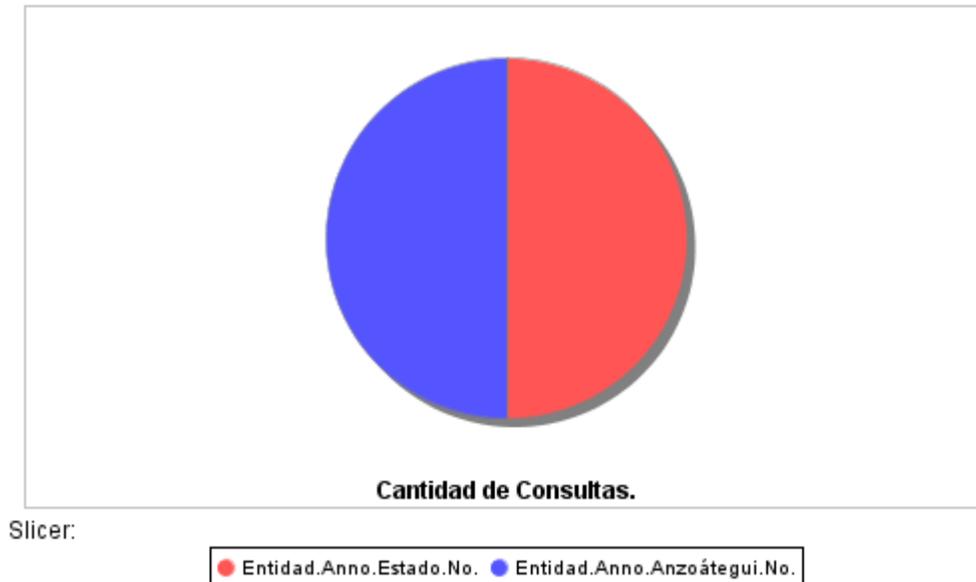


Figura 20 Otra vista de la información gráfica

3.5 Despliegue del Mercado de Datos.

Para realizar el despliegue del mercado de datos para el CeNaEst se necesitarán un total de 3 servidores, uno para la base de datos del sistema alasSIGEST implementado sobre el gestor de base de datos PostgreSQL, otro para el mercado de datos sobre el gestor de base de datos PostgreSQL 8.4, y un servidor de aplicación Web donde se ejecuta el Pentaho Mondrian sobre el Pentaho BI Server. Se realiza de esta forma ya que se recomienda que los servidores de origen y el almacén de datos estén separados para una mayor seguridad de los mismos y mayor protección con respecto a fallas.

Los servidores se comunican mediante el protocolo TCP/IP y la PC cliente se conecta mediante el protocolo HTTP para realizar las consultas al mercado de datos a través de la herramienta Mondrian. Los servidores de base de datos deben de tener como mínimo 150 Giga Bytes de capacidad de almacenamiento en disco y 1 Giga Bytes de memoria RAM. Para el servidor de aplicación se recomienda que posea 2 GB de memoria RAM y para las PC cliente 80 Giga Bytes de capacidad de almacenamiento en disco y 512 Mega Bytes de memoria RAM.

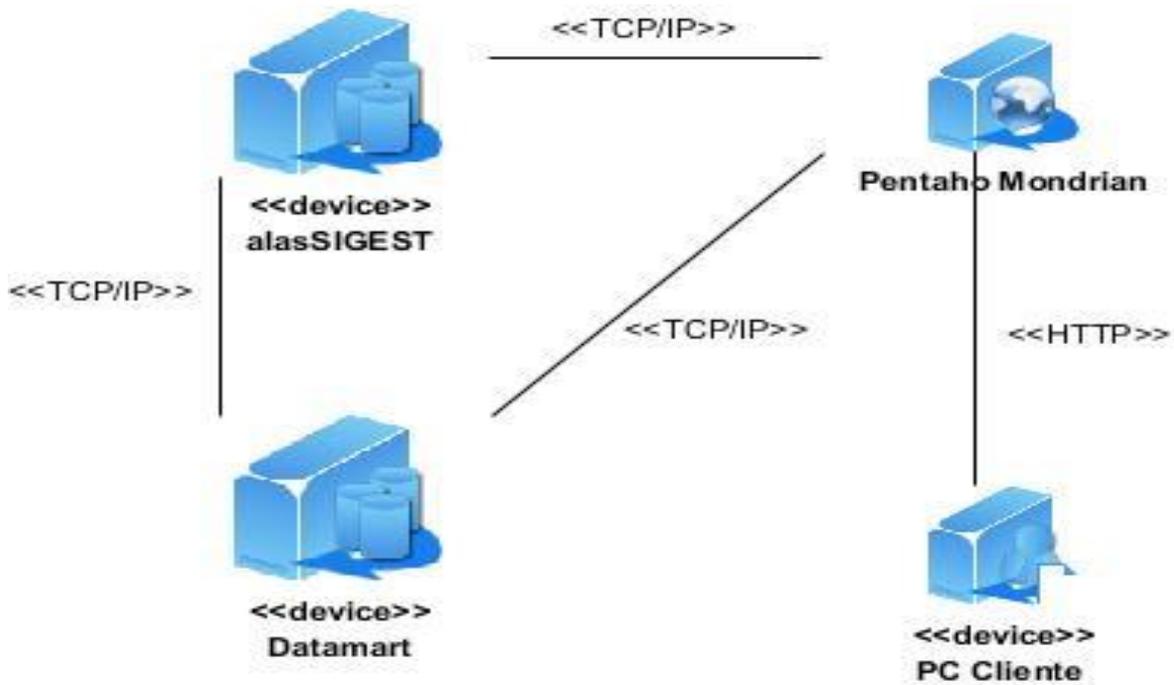


Figura 21 Despliegue del Mercado de Datos

3.6 Pruebas

Las pruebas de software son los procesos que permiten verificar y revelar la calidad de un producto. Son utilizadas para identificar posibles fallos de implementación, calidad y usabilidad en la solución. Para determinar el nivel de calidad se deben efectuar pruebas que permitan comprobar el grado de cumplimiento de las especificaciones iniciales del sistema.

Para la validación del mercado de datos se aplicaron diferentes tipos de pruebas, las cuales se especifican a continuación:

- **Pruebas de rendimiento:** Son las pruebas que se realizan, desde una perspectiva, para determinar lo rápido que realiza una tarea un sistema en condiciones particulares de trabajo. También puede servir para validar y verificar otros atributos de la calidad del sistema, tales como la escalabilidad, fiabilidad y uso de los recursos.
- **Prueba de estrés:** Se utiliza normalmente para romper la aplicación. Se va doblando el número de usuarios que se agregan a la aplicación y se ejecuta una prueba de carga hasta que se

rompe. Este tipo de prueba se realiza para determinar la solidez de la aplicación en los momentos de carga extrema y ayuda a los administradores para determinar si la aplicación rendirá lo suficiente en caso de que la carga real supere a la carga esperada.

3.6.1 Pruebas de rendimiento y estrés.

En las aplicaciones web, además de validar los requisitos funcionales, existen varios aspectos adicionales a analizar, como por ejemplo, el rendimiento del servidor, la facilidad de navegación y el tiempo de respuesta a eventos, con el objetivo de determinar no solo la velocidad, sino la eficiencia de los mismos. Para corroborar el tiempo de visualización de un reporte, se utilizó la herramienta Jakarta-jmeter en su versión 2.3.3, en una máquina con un procesador Celeron de 1.60 GHz, 2 Gb de memoria RAM y 150 Gb de capacidad de almacenamiento en disco.

El jmeter es un programa de código abierto, implementado en Java que permite realizar pruebas de comportamiento funcional y medir el rendimiento. Originalmente fue diseñado para probar las aplicaciones Web, pero desde entonces se ha expandido a otras funciones de prueba. También se puede utilizar para realizar pruebas de carga y estrés a los servidores.

Plan de Prueba

Se configuró un plan de prueba con la visualización del reporte “Porcentaje de Pacientes por Entidad” implementado en el mercado de datos. Para ello primero se creó un servidor proxy HTTP, se configuró un acceso de 15, 30 y 60 usuarios conectados concurrentemente al mercado de datos.

Resultados de la prueba

Luego de concluidas las pruebas de rendimiento se obtuvo como tiempos de respuesta del reporte 2.6/sec, 7.0/sec y 19.4/sec para 15, 30 y 60 accesos de usuarios respectivamente. En la figura 22 se puede observar los resultados de la prueba para 30 usuarios conectados.

CAPÍTULO 3: IMPLEMENTACIÓN DEL MERCADO DE DATOS PARA EL CENAEST

Label	# Muestras	Media	Mediana	Linea de 90%	Mín	Máx	% Error	Rendimiento
/pentaho/gwtr...	45	38	25	80	4	218	100,00%	5,6/min
/pentaho/cont...	45	130	25	77	4	3502	0,00%	5,6/min
/pentaho/cont...	45	61	22	78	5	722	0,00%	5,6/min
/pentaho/cont...	45	64	20	75	5	1276	0,00%	5,6/min
/pentaho/cont...	45	38	24	88	4	195	0,00%	5,6/min
/pentaho/mant...	45	45	18	110	4	364	100,00%	5,6/min
/pentaho/mant...	405	99	26	282	4	3044	0,00%	47,0/min
/pentaho/cont...	45	55	23	95	4	667	0,00%	5,6/min
/pentaho/mant...	45	34	12	69	2	449	100,00%	5,6/min
/pentaho/mant...	45	53	22	103	4	469	0,00%	5,6/min
/pentaho/mant...	45	84	42	214	5	646	0,00%	5,7/min
/pentaho/Solut...	45	63	26	140	4	282	100,00%	5,7/min
/pentaho/mant...	90	70	35	211	5	674	0,00%	11,2/min
/pentaho/cont...	45	38	21	86	5	179	0,00%	5,6/min
/ztb/update/fire...	45	6202	7973	11009	0	12040	100,00%	5,5/min
/media/plugin...	45	13695	14384	14546	9832	14683	100,00%	5,4/min
/safebrowsing...	45	12320	12336	12480	12057	12532	100,00%	5,4/min
/pentaho/cont...	45	94	77	167	23	180	0,00%	5,5/min
/pentaho/View...	45	179	199	249	56	338	0,00%	5,5/min
/pentaho/Pivot...	45	219	263	330	56	342	0,00%	5,5/min
/pentaho/jpivot...	90	284	143	805	6	1850	0,00%	11,0/min
/pentaho/jpivot...	90	193	109	446	5	1117	0,00%	11,0/min
/pentaho/jpivot...	90	157	108	307	6	994	0,00%	11,0/min
/pentaho/adho...	45	218	188	413	26	871	0,00%	5,5/min
/pentaho/adho...	45	198	174	377	34	488	0,00%	5,5/min
/pentaho/adho...	45	230	199	483	22	573	0,00%	5,5/min
/pentaho/adho...	45	221	211	406	26	709	0,00%	5,5/min
TOTAL	3780	805	64	499	0	16693	14,29%	7,0/sec

Figura 22 Prueba de rendimiento y estrés.

Con la realización de esta prueba se pudo observar que el tiempo de respuesta es aceptable debido a la cantidad de usuarios conectados concurrentemente, así como las características que posee la computadora donde se realizó la prueba.

Además se realizaron pruebas de rendimiento a las transformaciones utilizando los componentes de monitoreo que brinda la herramienta Kettle Pentaho Data Integration integrada en la Suite de Pentaho Open Source Business Intelligence. A continuación se muestra la prueba realizada a la transformación de la tabla de hechos “h_pacientes_egi”.

#	Nombre paso	Numero Copia	Leído	Escrito	Entrada	Salida	Actualizado	Rejected	Errores	Activo	Tiempo
1	Agrupar por id	0	70	45	0	0	0	0	0	Finalizado	2.3s
2	Insertar / Actualizar	0	45	45	45	43	0	0	0	Finalizado	2.6s
3	EGI	0	70	70	70	0	0	0	0	Finalizado	1.9s
4	Entidad	0	70	70	70	0	0	0	0	Finalizado	1.9s
5	Especialidad	0	70	70	70	0	0	0	0	Finalizado	2.0s
6	Entrada Tabla	0	0	70	70	0	0	0	0	Finalizado	1.6s
7	Localidad	0	70	70	70	0	0	0	0	Finalizado	2.2s
8	Fecha	0	70	70	70	0	0	0	0	Finalizado	2.3s
9	Ordenar campos	0	70	70	0	0	0	0	0	Finalizado	2.3s
10	Paciente	0	70	70	70	0	0	0	0	Finalizado	2.3s

Figura 23 Prueba de rendimiento a la transformación de la tabla de hechos “h_pacientes_egi”.

Una vez realizada la prueba a la tabla de hechos “h_pacientes_egi” durante el proceso de extracción, transformación y carga de los datos se comprobó que la misma se ejecutó satisfactoriamente en un período de tiempo razonable.

3.6.2 Lista de chequeo

La lista de chequeo es un documento que consta de una serie de preguntas para medir el cumplimiento de varias reglas establecidas en el desarrollo de un trabajo. Al mercado de datos para el CeNaEst se le aplicó una lista de chequeo para medir tres aspectos fundamentales:

- **Estructura del documento:** abarca todos los aspectos definidos por el expediente de proyecto o el formato establecido por el proyecto.
- **Indicadores definidos por etapas:** 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 estructura de la lista de chequeo está formada por los siguientes elementos:

- **Peso:** define si el indicador a evaluar es crítico o no.
- **Indicadores a evaluar:** son los indicadores a evaluar en las secciones **Estructura del documento**, **Semántica del documento** e **Indicadores definidos por la etapa**

- **Evaluación (Eval):** es la forma de evaluar el indicador en cuestión. El mismo se evalúa de 1 en caso de que exista alguna dificultad sobre el indicador y 0 en caso de que el indicador revisado no presente problemas.
- **No Procede (N.P.):** se usa para especificar que el indicador no es necesario evaluarlo en ese caso.
- **Cantidad de elementos afectados:** especifica la cantidad de errores encontrados sobre el mismo indicador.
- **Comentario:** especifica los señalamientos o sugerencias que quiera incluir la persona que aplica la lista de chequeo. Pueden o no existir señalamientos o sugerencias.

Luego de definir la estructura de la lista de chequeo esta es aplicada al mercado de datos para el CeNaEst. La lista se encuentra en el artefacto lista de chequeo. A continuación se muestra en la Figura 24 el comportamiento de los 24 indicadores, organizados por secciones donde se encontró una no conformidad en la sección “Indicadores definidos por etapas”:

NC1. No se tuvo en cuenta los formatos fuentes y tipos de datos de las perspectivas de análisis.

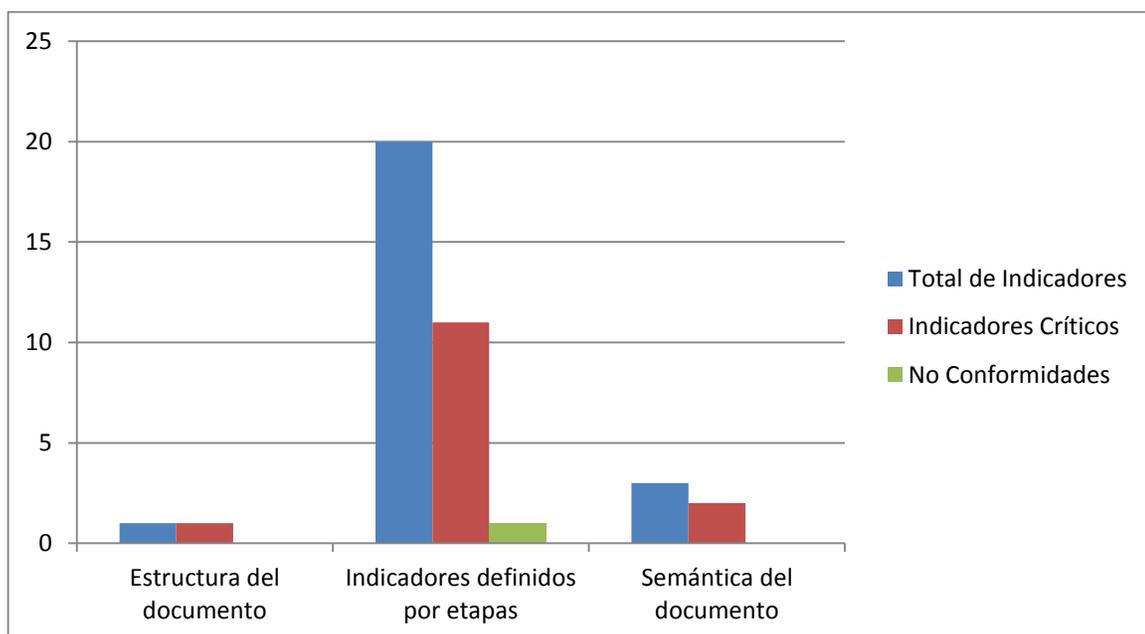


Figura 24 Comportamiento de los indicadores por etapas

Conclusiones del capítulo

En el capítulo se realizaron los procesos de Extracción, Transformación y Carga, se definieron las jerarquías de los atributos de las tablas de dimensiones y se diseñaron los cubos multidimensionales que permiten realizar los cruces necesarios para visualizar la información contenida en el Almacén de Datos. En general, se realizó el procesamiento analítico en línea para agilizar las consultas de la información en el almacén, mostrando los datos a los directivos desde diferentes perspectivas. Además se realizaron pruebas de rendimiento y estrés para probar la eficiencia del mercado de datos.

CONCLUSIONES

En el presente trabajo desarrolló un producto de software que consiste en un Mercado de Datos para el CeNaEst. El mismo permite el almacenamiento de grandes volúmenes de información relevante, transformando los datos operativos en información analítica enfocada a la toma de decisiones. Las principales contribuciones de este trabajo de diploma son:

- Se realizó un estudio del estado del arte de los principales sistemas de soporte a la toma de decisiones haciendo énfasis en los almacenes de datos y sistemas de software que realizan procesos similares, lo que permitió conocer las tendencias actuales en cuanto a la construcción y diseño de este tipo de herramientas .
- Se obtuvo un análisis de los principales requerimientos de información en el CeNaEst posibilitando una visión global del negocio.
- Se realizó el diseño y la implementación de la solución informática utilizando como guía la metodología Híbrida que cubrió todas las etapas del desarrollo, evitando pasar por alto detalles que causen una mala implementación de la solución mercado de datos.

REFERENCIAS BIBLIOGRÁFICAS

Referencias Bibliográficas

1. Perú, Universidad <http://www.universidadperu.com>. [En línea] 2005. [Citado el: 5 de 2 de 2013.] <http://www.universidadperu.com/estomatologia-peru.php>.
2. Bernabeu, Ing. Ricardo Darío. Data Warehousing: Investigación y sistematización de conceptos- Hefesto: Metodología propia para la construcción de un Data Warehouse. Argentina : s.n., 2007.
3. López, Leydis Hidalgo y Cartaya, Alier Caballero. Implementación de un DataMart para la Unidad Central de Colaboración Médica. La Habana, Cuba : s.n., 2010.
4. Barrientos, Miguel Angel Montalvo. IMPLEMENTACIÓN DE UN DATA MART PARA LA FACULTAD NACIONAL DE INGENIERÍA CASO: CARRERA ING. DE SISTEMAS. Oruro, Bolivia : s.n., 6 de abril de 2010.
5. Ing. Mellado Méndez, Augusto. DATAMART CAPACITACIÓN. San Martín de Porres, Perú : s.n., 2009.
6. Hidalgo, Ing. Gisel González y Rizo, MSc Emma Rizo. Data Mart de Gestión Económico - Financiera de la Empresa de. Matanzas, Cuba : s.n.
7. Nueva Aguilar, Yuniesky. Construcción de un Data Mart para el Sistema Integral de Gestión de Medicamentos, La Habana, Cuba, junio 2011.
8. Izquierdo, Malena García y Martínez., Manuel Macías. Diseño del repositorio de datos para la sala situacional de un sistema de gestión penitenciaria. La Habana, Cuba : s.n.
9. Mascareño Hodge, Leidys Susel. Peña Consuegra, Patricia. Sistema de Información de Gobierno. Mercado de datos Comercio exterior. La Habana, Cuba, Junio 2011.
10. Rivadera, Gustavo R. La metodología de Kimball para el diseño de almacenes de datos (Data Warehouses) .
11. Lopez, Jorge Luis Tufiño. Desarrollo del datamart para el Sistema Nacional de Vigilancia Tecnológica de Software Libre. Quito, Ecuador : s.n., Agosto, 2011.

REFERENCIAS BIBLIOGRÁFICAS

12. Lorenzo, Tte. Ing. Orestes Rodríguez. Guía para el desarrollo de soluciones de almacenes de datos. La Habana, Cuba, abril 2012.
13. Bernabeu, Ing. Ricardo Darío. ARQUITECTURA DEL DATA WAREHOUSING.
14. Morales, González. Inteligencia empresarial para la toma de decisiones en la PYME enfocada en la administración de la relación con el cliente (CRM) utilizando análisis de la canasta de compra (MBA). Escuela de Ingeniería y Ciencias, Universidad de las Américas. [En línea] 2007. http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/gonzalez_m_ac/capitulo_3.html.
15. Guanche Cañizares, Mauricio. Tesis Desarrollo de un DataMart para la obtención de las razones financieras de los subsistemas de Cobros y Pagos, Caja y Banco del proyecto ERP-Cuba. La Habana, Cuba, Junio 2011.
16. La revista de tecnología y estrategia de negocio en Internet. La empresa multidimensional: OLAP. 11/1/2002. Número: 54 [En línea] <http://www.idg.es/iWorld/articulo.asp?id=143456>.
17. Persistencia Molap, Rolap, Holap. [En línea]. 09-2009 http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx.
18. Analisis del estado de Mondrian (*OLAP Open Source*). Giménez, Javier. 2007.
19. <http://www.enrichiosca.es/2009/05/tutorial-introduccion-al-mdx-capitulo-1.html>
20. El SGBDR Oracle. Velasco., Roberto Hernando.
21. PostgreSQL vs. MySQL. Pecos, Daniel.
22. Martinez, Rafael. <http://www.postgresql.org.es>. [En línea] [Citado el: 8 de 12 de 2012.] http://www.postgresql.org.es/sobre_postgresql.
23. Paradigm, Visual. <http://www.visual-paradigm.com>. [En línea] [Citado el: 21 de 1 de 2013.] <http://www.visual-paradigm.com./product/vpuml/>.

REFERENCIAS BIBLIOGRÁFICAS

24. Enterprise Architect - Herramienta de diseño UML. <http://sparxsystems.com.ar>. [En línea] [Citado el: 21 de 1 de 2013.] <http://sparxsystems.com.ar/products/ea.html>.
25. Kafati., Elizabeth Gutiérrez. La plataforma Pentaho Open Source Business Intelligence.
26. SEN, A. and SINHA, A. P. A Comparison of Datawarehousing Methodologies, Communication of the ACM, pp. 79-84, 2005.

Bibliografía

1. Barrientos, Miguel Angel Montalvo. IMPLEMENTACIÓN DE UN DATA MART PARA LA FACULTAD NACIONAL DE INGENIERÍA CASO: CARRERA ING. DE SISTEMAS. Oruro, Bolivia : s.n., 6 de abril de 2010. 7.
2. Bernabeu, Ing. Ricardo Darío. Data Warehousing: Investigación y sistematización de conceptos- Hefesto: Metodología propia para la construcción de un Data Warehouse. Argentina : s.n., 2007. 2.
3. Bernabeu, Ing. Ricardo Darío. ARQUITECTURA DEL DATA WAREHOUSING.
4. Casares, Claudio. DataWarehousing. [En línea] 10-2009.
<http://www.programacion.com/bbdd/tutorial/warehouse/>
5. Curto, Josep. Data Warehousing, Data Warehouse y Datamart.[En línea] 10-2009.
<http://informationmanagement.wordpress.com/2007/10/07/data-warehousing-data-warehouse-y-datamart/>
6. Enterprise Architect - Herramienta de diseño UML. <http://sparxsystems.com.ar>.
<http://sparxsystems.com.ar>. [En línea] [Citado el: 21 de 1 de 2013.]
<http://sparxsystems.com.ar/products/ea.html>. 42.
7. Hidalgo, Ing. Gisel González y Rizo, MSc Emma Rizo. Data Mart de Gestión Económico - Financiera de la Empresa de. Matanzas, Cuba : s.n. 11.
8. Izquierdo, Malena García y Martínez., Manuel Macías. Diseño del repositorio de datos para la sala situacional de un sistema de gestión penitenciaria. La Habana, Cuba : s.n. 15.
9. Kafati., Elizabeth Gutiérrez. La plataforma Pentaho Open Source Business Intelligence.
10. López, Leydis Hidalgo y Cartaya, Alier Caballero. Implementación de un DataMart para la Unidad. La Habana, Cuba : s.n., 2010. 6.
11. La revista de tecnología y estrategia de negocio en Internet. 54, 11/1/2002.

12. Lopez, Jorge Luis Tufiño. Desarrollo del datamart para el Sistema Nacional de Vigilancia Tecnológica de Software Libre. Quito, Ecuador : s.n., Agosto, 2011. 18.
13. Martinez, Rafael. <http://www.postgresql.org.es>. <http://www.postgresql.org.es>. [En línea] [Citado el: 8 de 12 de 2012.] http://www.postgresql.org.es/sobre_postgresql. 39.
14. Mellado Méndez, Ing. Augusto. DATAMART CAPACITACIÓN. San Martín de Porres, Perú : s.n., 2009. 10.
15. Nueva Aguilar, Yuniesky. Tesis Construcción de un Data Mart para el Sistema Integral de Gestión de Medicamentos, La Habana, Cuba, junio 2011.
16. Oporto Díaz, Mg. Samuel. Metodologías para el Data Warehousing.
17. Paradigm, Visual. <http://www.visual-paradigm.com>. <http://www.visual-paradigm.com>. [En línea] [Citado el: 21 de 1 de 2013.] <http://www.visual-paradigm.com/product/vpuml/>. 41.
18. Persistencia Molap, Rolap, Holap.
19. Perú, Universidad de. <http://www.universidadperu.com>. <http://www.universidadperu.com>. [En línea] 2005. [Citado el: 5 de 2 de 2013.] <http://www.universidadperu.com/estomatologia-peru.php>. 1.
20. PostgreSQL vs. MySQL. Pecos, Daniel.
21. Rivadera, Gustavo R. La metodología de Kimball para el diseño de almacenes de datos (Data warehouses) . 17.
22. Roberto Hernando. El SGBDR Oracle. Velasco.

ANEXOS

1 Carga de las dimensiones y hechos del Mercado de Datos.

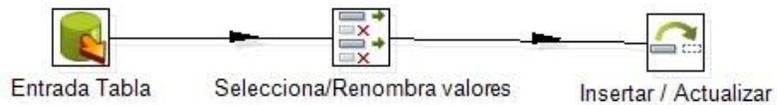
1.1 Dimensión “d_egi”



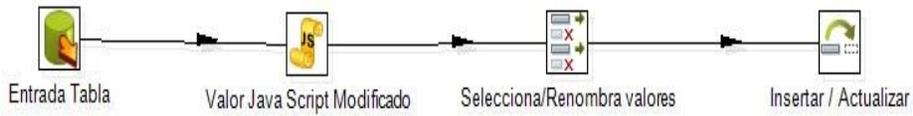
1.2 Dimensión “d_entidad”



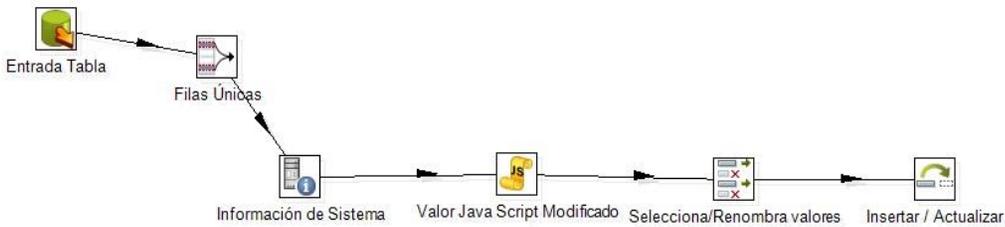
1.3 Dimensión “d_especialidad”



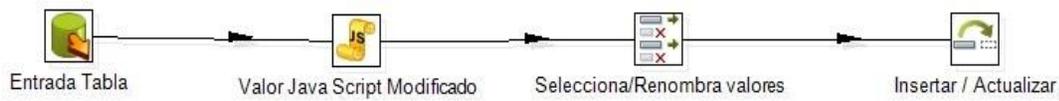
1.4 Dimensión “d_ortodoncia”



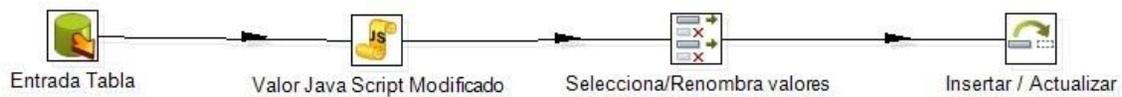
1.5 Dimensión “d_paciente”



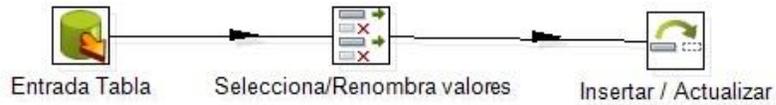
1.6 Dimensión “d_prótesis”



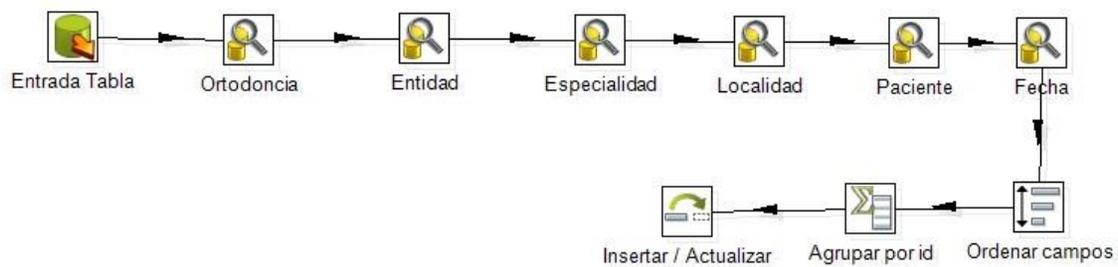
1.7 Dimensión “d_tiempo”



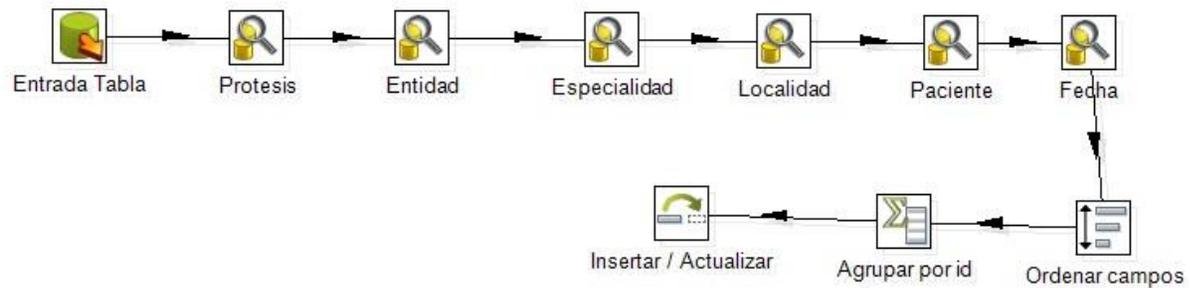
1.8 Dimensión “d_localidad”



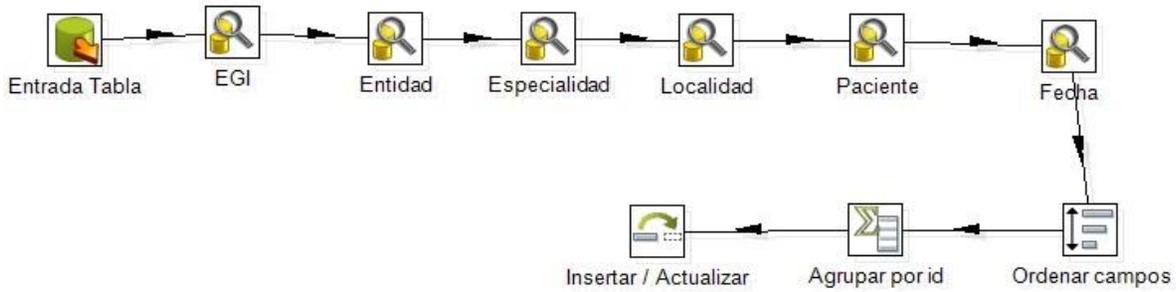
1.9 Tabla de hechos “hechos_pacientes_ortodoncia”



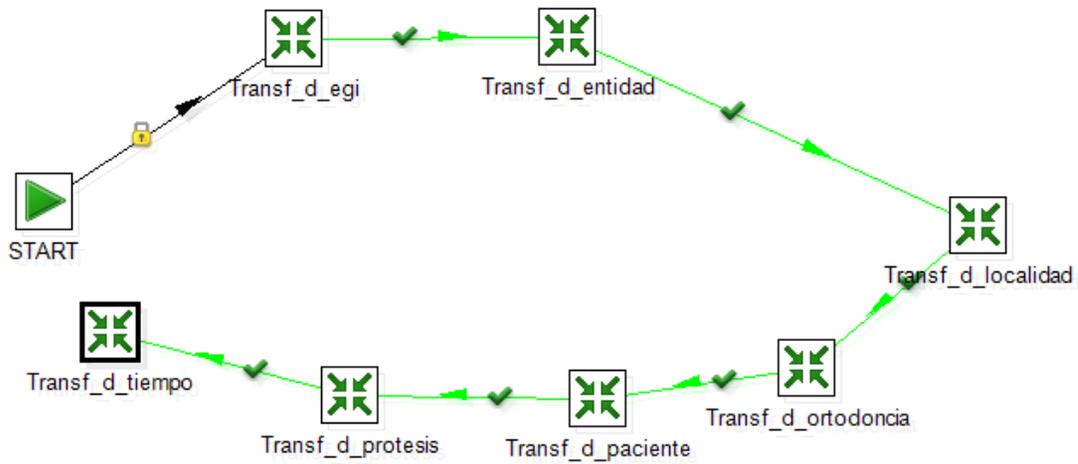
1.10 Tabla de hechos “hechos_pacientes_prótesis”



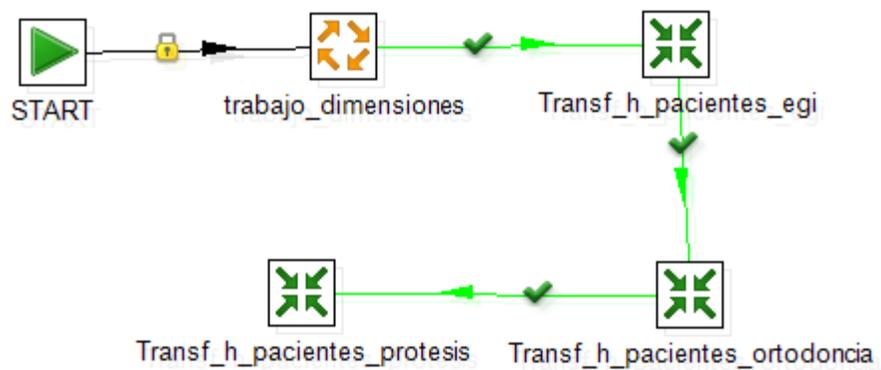
1.11 Tabla de hechos “hechos_pacientes_egi”



1.12 Trabajo que carga todas las dimensiones.



1.13 Trabajo que carga todo.



2. Requerimientos de Información

- RI1 - Cantidad de pacientes atendidos en EGI por cada una de las entidades en un tiempo determinado y en una localidad determinada.
- RI2 - Cantidad de pacientes atendidos en Ortodoncia por cada una de las entidades en un tiempo determinado y en una localidad determinada.
- RI3 - Cantidad de pacientes atendidos en Prótesis por cada una de las entidades en un tiempo determinado y en una localidad determinada.
- RI4 - Cantidad de consultas atendidas con traumatismo por entidad en un tiempo determinado y en una localidad determinada.
- RI5 - Cantidad de consultas atendidas que se haya detectado cáncer bucal por entidad en un tiempo determinado y en una localidad determinada.
- RI6 - Cantidad de consultas atendidas por obturaciones por entidad en un tiempo determinado y en una localidad determinada.
- RI7 - Cantidad de consultas atendidas por presencia de sarro por entidad en un tiempo determinado y en una localidad determinada.
- RI8 - Cantidad de consultas atendidas con colocación de implantes realizados por entidad en un tiempo determinado y en una localidad determinada.
- RI9 - Cantidad de consultas atendidas con pulpotomía realizada por entidad en un periodo determinado y en una localidad determinada.
- RI10 - Cantidad de consultas atendidas con rayos x realizados por entidad en un tiempo determinado y en una localidad determinada.
- RI11 - Cantidad de consultas atendidas con tpr realizado por entidad en un tiempo determinado y en una localidad determinada.

- RI12 - Cantidad de consultas atendidas con interposición al tragar por entidad en un tiempo determinado y en una localidad determinada.
- RI13 - Cantidad de consultas atendidas con respiración bucal por entidad en un tiempo determinado y en una localidad determinada.
- RI14 - Cantidad de consultas atendidas con facie adenoidea por entidad en un tiempo determinado y en una localidad determinada.
- RI15 - Cantidad de consultas atendidas con Narinas hipotónicas por entidad en un tiempo determinado y en una localidad determinada.
- RI16 - Cantidad de consultas atendidas con Fonación/Función afectada por entidad en un tiempo determinado y en una localidad determinada.
- RI17 - Cantidad de consultas atendidas con traumatismo por ortodoncia por entidad en un tiempo determinado y en una localidad determinada.
- RI18 - Cantidad de consultas atendidas con enfermedad periodontal por entidad en un tiempo determinado y en una localidad determinada.
- RI19 - Cantidad de consultas atendidas con dificultades masticatorias por entidad en un tiempo determinado y en una localidad determinada.
- RI20 - Cantidad de consultas atendidas con experiencia protésica por entidad en un tiempo determinado y en una localidad determinada.
- RI21 - Cantidad de consultas atendidas con disposición para recibir tratamiento de ambos sexos por entidad en un tiempo determinado y en una localidad determinada.
- RI22 - Cantidad de consultas atendidas según estado de salud de los labios por entidad en un tiempo determinado y en una localidad determinada.

- RI23 - Cantidad de consultas atendidas según estado de salud de la lengua por entidad en un tiempo determinado y en una localidad determinada.
- RI24 - Cantidad de consultas atendidas que presentan torus palatino en la bóveda palatina por entidad en un tiempo determinado y en una localidad determinada.
- RI25 - Cantidad de consultas atendidas con presencia de torus palatino en la mandíbula por entidad en un tiempo determinado y en una localidad determinada.
- RI26 - Cantidad de consultas atendidas con estado de salud del cielo de la boca por entidad en un tiempo determinado y en una localidad determinada.
- RI27 - Cantidad de consultas atendidas con volumen de las glándulas salivares (normal o patológico) por entidad en un tiempo determinado y en una localidad determinada.
- RI28 - Cantidad de consultas atendidas con estado patológico de los carrillos por entidad en un tiempo determinado y en una localidad determinada.
- RI29 - Cantidad de consultas atendidas con cantidad de saliva (escasa o normal) por entidad en un tiempo.
- RI30 - Porcentaje de pacientes atendidos en la especialidad de estomatología general integral (EGI) por entidad.
- RI31 - Porcentaje de pacientes atendidos en la especialidad de estomatología general integral (EGI) por localidad.
- RI32 - Porcentaje de pacientes atendidos en la especialidad de prótesis por localidad.
- RI33 - Porcentaje de pacientes atendidos en la especialidad de prótesis por entidad.
- RI34 - Porcentaje de pacientes atendidos en la especialidad de ortodoncia por entidad.
- RI35 - Porcentaje de pacientes atendidos en la especialidad de ortodoncia por localidad.

3. Requerimientos funcionales

- RF1 - Autenticar usuario.
- RF2 - Adicionar roles.

- RF3 - Eliminar roles.
- RF4 - Adicionar usuarios.
- RF5 - Eliminar usuarios.
- RF6 - Insertar reportes.
- RF7 - Modificar reportes.
- RF8 - Eliminar reportes.
- RF9 - Extraer información.
- RF10 - Realizar transformación y carga.
- RF11 - Abrir navegador OLAP.
- RF12 - Mostrar editor MDX.
- RF13 - Mostrar padres.
- RF14 - Ocultar repeticiones.
- RF15 - Intercambiar ejes.
- RF16 - Mostrar gráfico.
- RF17 - Configurar gráfico.
- RF18 - Configurar impresión.
- RF19 - Exportar a PDF.
- RF20 - Exportar a excel.
- RF21 - Mostrar propiedades.
- RF22 - Suprimir filas.

- RF23 - Detallar miembros.
- RF24 - Entrar en detalles.
- RF25 - Mostrar datos de origen.

4. Requerimientos no funcionales

- RNF1 - Usabilidad: mostrar los mensajes, títulos y demás textos que aparezcan en la interfaz del sistema en idioma español.
- RNF2 - Usabilidad: navegar en los reportes del Almacén de Datos de manera ágil.
- RNF3 - Confiabilidad: garantizar el cumplimiento de actualización de los datos en el almacén.
- RNF4 - Eficiencia: el tiempo de respuesta debe ser en tiempo real.
- RNF5 - Interfaz: los reportes estadísticos deben contar con una interfaz simple que facilite la interacción usuario-aplicación.
- RNF6 - Interfaces Hardware: proporcionar características mínimas de hardware a los servidores.
- RNF7 - Usabilidad: establecer tiempo de entrenamiento requerido para que usuarios normales sean productivos operando el sistema.
- RNF8 - Usabilidad: diseñar un reporte del Almacén de Datos de manera sencilla y ágil.
- RNF9 - Confiabilidad: garantizar la persistencia de la información.
- RNF10 - Eficiencia: el sistema debe permitir la abundancia de usuarios conectados simultáneamente sin que se afecte el tiempo de respuesta, al menos 700.
- RNF11 - Eficiencia: en el proceso de integración solo tendrá conectado un usuario que tendrá la tarea de vigilar el proceso de integración de datos.

- RNF12 - Soporte: lograr que los elementos definidos en el almacén tengan una nomenclatura homogénea.
- RNF13 - Restricciones del Diseño: utilizar el lenguaje de programación definido durante la investigación.
- RNF14 - Requisitos para a documentación de usuarios en línea y ayuda del sistema: confeccionar manual de usuario.
- RNF15 - Interfaz: acceder al sistema.
- RNF16 - Interfaces de usuario: los requerimientos de interfaz de usuario se centran en la presentación de la información de cara al cliente.
- RNF17 - Interfaz: las interfaces de salida no serán cargadas con información innecesaria.
- RNF18 - Interfaces Hardware: proporcionar características mínimas de hardware a las estaciones de trabajo.
- RNF19 - Interfaces Software: instalar en las estaciones de trabajo el software necesario para el correcto funcionamiento del sistema.
- RNF20 - Interfaces Software: Las configuraciones de software de las máquinas clientes deben contar al menos con los siguientes elementos: Firefox 2.0 o superior, Java Virtual Machine, Schema Workbench en caso de que un usuario capacitado requiera la construcción de esquemas multidimensionales para el diseño de nuevos reportes y Pentaho Data Integration.
- RNF21 - Interfaces de Comunicación: la comunicación entre la base de datos de integración y el almacén de datos es a través del protocolo TCP/IP.
- RNF22 - Interfaces de Comunicación: el sistema necesita estar conectado directamente a un dispositivo de red.
- RNF23 - Requisitos de Licencia: las herramientas a utilizar se encuentran bajo licencia GPL.

5. Indicadores y Perspectivas

Cantidad de pacientes atendidos en EGI por cada una de las entidades en un tiempo determinado y en

Indicador **Perspectivas**

una localidad determinada.

Perspectivas

Cantidad de pacientes atendidos en Ortodoncia por cada una de las entidades en un tiempo

Indicador **Perspectivas**

determinado y en una localidad determinada.

Perspectivas

Cantidad de pacientes atendidos en Prótesis por cada una de las entidades en un tiempo determinado

Indicador **Perspectivas**

y en una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con traumatismo por entidad en un tiempo determinado y en una

Indicador **Perspectivas**

localidad determinada.

Perspectivas

Cantidad de consultas atendidas que se haya detectado cáncer bucal por entidad en un tiempo

Indicador

Perspectivas

determinado y en una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con obturaciones por entidad en un tiempo determinado y en una

Indicador

Perspectivas

localidad determinada.

Perspectivas

Cantidad de consultas atendidas con presencia de sarro por entidad en un tiempo determinado y en

Indicador

Perspectivas

una localidad determinada.

Perspectivas

Cantidad de consultas atendidas de implantes realizados por entidad en un tiempo determinado y

Indicador

Perspectivas

en una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con pulpotomía realizada por entidad en un periodo determinado y en

Indicador

Perspectivas

una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con rayos x realizados por entidad en un tiempo determinado y en

Indicador

Perspectivas

una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con tpr realizado por entidad en un tiempo determinado y en una

Indicador

Perspectivas

localidad determinada.

Perspectivas

Cantidad de consultas atendidas con Interposición al tragar por entidad en un tiempo determinado y en

Indicador

Perspectivas

una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con Respiración bucal por entidad en un tiempo determinado y en

Indicador

Perspectivas

una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con Facie adenoidea por entidad en un tiempo determinado y en una

Indicador

Perspectivas

localidad determinada.

Perspectivas

Cantidad de consultas atendidas con Narinas hipotónicas por entidad en un tiempo determinado y en

Indicador

Perspectivas

una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con Fonación/Función afectada por entidad en un tiempo

Indicador

Perspectivas

determinado y en una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con traumatismo por ortodoncia por entidad en un tiempo

Indicador

Perspectivas

determinado y en una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con enfermedad periodontal por entidad en un tiempo determinado y

Indicador

Perspectivas

en una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con dificultades masticatorias de ambos sexos por entidad en un

Indicador

Perspectivas

tiempo determinado y en una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con experiencia protésica por entidad en un tiempo determinado y en

Indicador

Perspectivas

una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con disposición para recibir tratamiento de ambos sexos por entidad

Indicador

Perspectivas

en un tiempo determinado y en una localidad determinada.

Perspectivas

Cantidad de consultas atendidas según estado de salud de los labios por entidad en un tiempo

Indicador

Perspectivas

determinado y en una localidad determinada.

Perspectivas

Cantidad de consultas atendidas según estado de salud de la lengua por entidad en un tiempo

Indicador

Perspectivas

determinado y en una localidad determinada.

Perspectivas

Cantidad de consultas atendidas que presentan torus palatino en la bóveda palatina por entidad en un

Indicador

Perspectivas

tiempo determinado y en una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con presencia de torus palatino en la mandíbula por entidad en un

Indicador

Perspectivas

tiempo determinado y en una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con estado de salud del cielo de la boca por entidad en un tiempo

Indicador

Perspectivas

determinado y en una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con volumen de las glándulas salivares (normal o patológico) por

Indicador

Perspectivas

entidad en un tiempo determinado y en una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con estado patológico de los carrillos por entidad en un tiempo

Indicador

Perspectivas

determinado y en una localidad determinada.

Perspectivas

Cantidad de consultas atendidas con cantidad de saliva (escasa o normal) por entidad en un tiempo.

Indicador

Perspectivas

Por ciento de pacientes atendidos en la especialidad de estomatología general integral (EGI) por entidad.

Indicador

Perspectivas

Por ciento de pacientes atendidos en la especialidad de EGI por localidad.

Indicador

Perspectivas

Por ciento de pacientes atendidos en la especialidad de prótesis por localidad.

Indicador

Perspectivas

Por ciento de pacientes atendidos en la especialidad de prótesis por entidad.

Indicador

Perspectivas

Por ciento de pacientes atendidos en la especialidad de ortodoncia por entidad.

Indicador

Perspectivas

Por ciento de pacientes atendidos en la especialidad de ortodoncia por localidad.

Indicador

Perspectivas