

Universidad de las Ciencias Informáticas

Facultad 7



Trabajo de Diploma para optar por el título de

Ingeniero en Ciencias Informáticas

**Implementación de un DataMart para la Unidad
Central de Cooperación Médica**

Autores: Leydis Hidalgo López

Alier Caballero Cartaya

Tutores: Ing. Alfredo Rodríguez Ruiz

Ing. Alejandro Catalá Aguirre

Ciudad de La Habana, Junio 2010

“Año 52 de la Revolución”

DECLARACIÓN DE AUTORÍA:

Declaramos que somos los únicos autores de este trabajo y autorizamos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmamos la presente a los _____ días del mes de Junio del año 2010.

Autora: Leydis Hidalgo López

Autor: Alier Caballero Cartaya

Tutor: Ing. Alfredo Rodríguez Ruiz

Tutor: Ing. Alejandro Catalá Aguirre

AGRADECIMIENTOS

Agradezco en primer lugar a mis padres por brindarme todo su amor durante toda mi vida, porque me han apoyado y confiado en mí siempre. Por ellos he llegado hasta aquí hoy y para ellos son mis logros.

A Beatriz por ser una segunda madre para mí en esta etapa de mi vida.

A mi familia en general por ser la mejor familia del mundo, darme su apoyo incondicional y confiar siempre en mí.

A mis amigas Diainys, Irina, Yunet, y Lisy que más que eso han sido mis hermanas, me siento muy orgullosa de haberlas conocido y haber compartido los momentos de alegrías y tristeza juntas. Espero que aunque estemos lejos unas de otras que nos separemos nunca. Las quiero.

A mi amigo Luis Orlando (Luichi) por estar a mi lado en las buenas y en las malas.

A Raúl por ser tan especial, apoyarme y hacerme sonreír en esta etapa tan importante. Gracias por existir.

A Cuqui, Annis, Iván, Edison, Abel, Jorgito, Marianela, Javier y todos los que de una forma u otra compartieron conmigo y me dieron ánimo para continuar.

A mi compañero de tesis por haber sido tan paciente y haberme soportado todo este tiempo.

A Roberto Acosta por ser mi ejemplo a seguir.

A mis tutores Alejandro Catalá y Alfredo Ruiz por apoyarnos, guiarnos y confiar en nosotros. Gracias.

A los miembros del tribunal y en especial a Héctor por ayudarnos a perfeccionar nuestro trabajo. Gracias por sus críticas y sugerencias.

A la Revolución y nuestro comandante Fidel Castro por haberme dado la oportunidad de estudiar en esta universidad ver hechos mis sueños realidad.

Leydis.

Agradecimientos

Quisiera comenzar agradeciendo a mis padres por ser las personas que me dieron vida y sin las cuales esta realización no sería posible. A mi madre, que ha sido mi impulso de siempre, mi fuerza en todo momento y mi eterno apoyo y cariño.

A mi padre, que es mi ejemplo y orgullo en la vida profesional, que nunca ha dejado de cuidarme y preocuparse. La realización de este trabajo es para que se sientan orgullosos de su hijo, por ellos han sido todos estos años de estudio.

A toda mi familia, a mi padrastro que tanto respeto, a mi hermano, mi abuelo, mis tíos y primo por apoyarme en todo.

A todos mis compañeros de clases y del apto que compartimos en tantos momentos buenos y malos. Como a todas mis amistades que me apoyaron en la realización de la Tesis.

Agradecer a mi querida compañera de tesis, a mis tutores que tanto esfuerzo empeñaron para llevar este resultado a cabo. A los miembros del tribunal y oponente por ser exigirme a emplearme al máximo en los estudios.

Quisiera agradecer de forma general a todas aquellas personas que de una forma u otra hicieron de mi estancia en la universidad la más agradable experiencia. Aquellos que me ayudaron, criticaron y depositaron en mí su confianza.

Gracias a todos,

Alier.

DEDICATORIA

Para mis padres, por su apoyo y todo lo que representan para mí.

Leydis.

RESUMEN

La colaboración médica es una premisa fundamental del sistema nacional de salud cubano. Su información es procesada por la Unidad Central de Cooperación Médica (UCCM) que con el proyecto Colaboración Médica mantiene un registro actualizado de los colaboradores y misiones médicas en el exterior. Por lo que surge la necesidad de almacenar grandes volúmenes de información para la toma de decisiones en la UCCM, así como la automatización del proceso de análisis. Para ello se traza como objetivo desarrollar un datamart que contenga la información consolidada del proyecto Colaboración Médica.

Actualmente la información se encuentra almacenada en una base de datos relacional sobre el sistema gestor MySQL por un período de tiempo indefinido lo cual provoca una carga intensa de la base de datos operacional y dificulta una respuesta rápida al cliente. Para la construcción del datamart para la UCCM se utiliza como guía la metodología Hefesto que consta de cuatro fases importantes destacándose; el proceso de Extracción, Transformación y Carga (ETL) para dar paso al procesamiento analítico en línea, utilizándose para ello herramientas de software libre.

El datamart desarrollado brinda fácil acceso a los datos que necesitan frecuentemente los directivos de la UCCM y permite que exista una fuente única de información relevante para la toma de decisiones mejorando el tiempo de respuesta. Los directivos obtendrán los valores numéricos que muestren lo que desean analizar concretamente y podrán obtener una representación gráfica para una mejor comprensión y estudio a la hora de tomar decisiones.

ÍNDICE

INTRODUCCIÓN..... 1

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA..... 5

1. ¿POR QUÉ UN ALMACÉN DE DATOS Y NO UNA BASE DE DATOS OPERACIONAL? 5

1.1. ¿POR QUÉ DATAMART Y NO DATAWAREHOUSE?..... 5

1.1.1. ¿QUÉ ES UN DATAWAREHOUSE? 5

1.1.2. ¿QUÉ ES UN DATAMART? 7

1.2. METODOLOGÍAS PARA EL DISEÑO E IMPLEMENTACIÓN DE UN DATAMART 7

1.2.1. LA METODOLOGÍA KIMBALL..... 7

1.2.2. LA METODOLOGÍA INMON..... 8

1.2.3. LA METODOLOGÍA HEFESTO 8

1.3. ¿QUÉ ES OLTP?..... 10

1.4. MODELADO MULTIDIMENSIONAL..... 10

1.4.1. ESQUEMA ESTRELLA 11

1.4.2. ESQUEMA COPO DE NIEVE 12

1.4.3. ESQUEMA DE CONSTELACIÓN..... 12

1.5. ETL 13

1.6. OLAP..... 14

1.6.1. ROLAP 14

1.6.2. MOLAP 15

1.6.3. HOLAP 15

1.7. HERRAMIENTAS PARA LA CONSTRUCCIÓN DEL DATAMART 15

1.7.1.	SISTEMAS GESTORES DE BASE DE DATOS	15
1.7.1.1.	ORACLE.....	16
1.7.1.2.	MYSQL.....	16
1.7.1.3.	POSTGRESQL.....	17
1.7.2.	HERRAMIENTAS DE INTEGRACIÓN DE DATOS.....	18
1.7.2.1.	CLOVER ETL.....	18
1.7.2.2.	ENHYDRAOCTOPUS.....	19
1.7.2.3.	KETTLE.....	19
1.7.3.	HERRAMIENTAS PARA OLAP	20
1.7.3.1.	JEDOX PALO	21
1.7.3.2.	OLAP4J.....	21
1.7.3.3.	MONDRIAN.....	21
1.7.3.4.	SCHEMA WORKBENCH (KETTLE)	22
1.7.4.	HERRAMIENTAS CASE.....	22
1.7.4.1.	RATIONAL ROSE.....	22
1.7.4.2.	VISUAL PARADIGM.....	22
1.7.4.3.	ENTERPRISE ARCHITECT.....	23
1.7.5.	SERVIDOR DE APLICACIONES.....	23
1.7.5.1.	JBOSS.....	24
1.7.5.2.	TOMCAT.....	24
1.8.	PROPUESTA DE LA PLATAFORMA A UTILIZAR	24
1.8.1.	MÓDULOS DE LA PLATAFORMA PENTAHO BI	25

1.9. DATAMARTS EXISTENTES.....	25
1.9.1. NIVEL INTERNACIONAL	25
1.9.2. NIVEL NACIONAL	26
1.9.3. UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS	27
CAPÍTULO 2. ANÁLISIS Y DISEÑO	28
2. PASOS Y APLICACIÓN DE LA METODOLOGÍA HEFESTO	28
2.1.1. ANÁLISIS DEL PROCESO	28
2.2. PASO1. ANÁLISIS DE REQUERIMIENTOS	29
2.2.1. IDENTIFICAR PREGUNTAS	29
2.2.2. INDICADORES Y PERSPECTIVAS DE ANÁLISIS	30
2.2.3. MODELO CONCEPTUAL.....	32
2.3. PASO2. ANÁLISIS DE LOS OLTP	33
2.3.1. CORRESPONDENCIA CON LOS REQUERIMIENTOS.....	33
2.3.2. CAMPOS QUE INTEGRAN LAS PERSPECTIVAS. NIVEL DE GRANULARIDAD	34
2.4. PASO3. MODELO LÓGICO DE LA ESTRUCTURA DEL DATAMART.....	36
2.5. PASO4. PROCESO DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA (ETL).....	36
CAPÍTULO 3. PROCESO ANALÍTICO EN LÍNEA	42
3. PROCESOS DE ANÁLISIS DE LA INFORMACIÓN CONTENIDA EN EL DATAMART	42
3.1. JERARQUÍA DE LAS TABLAS DE DIMENSIONES	42
3.2. MEDIDAS	43
3.3. CONSULTAS.....	44
3.4. RESULTADOS DE LOS CLIENTES DE MONDRIAN.....	44

3.4.1. JPIVOT	44
3.5. DESPLIEGUE DEL DATAMART	47
CONCLUSIONES	49
BENEFICIOS	50
RECOMENDACIONES	51
REFERENCIAS BIBLIOGRÁFICAS	52
BIBLIOGRAFÍA	55
ANEXOS.	57
1. BASE DE DATOS DE COLABORACIÓN MÉDICA.	57
2. CARGA DE LAS DIMENSIONES DEL DATAMART.	58
2.1. DIMENSIÓN MISIÓN	58
2.2. DIMENSIÓN LOCALIZACIÓN	58
2.3. DIMENSIÓN NÓMINA.	58
2.4. DIMENSIÓN PASAPORTE.	58
2.5. DIMENSIÓN MOVIMIENTO.	58

INTRODUCCIÓN

El desarrollo de la informática y las telecomunicaciones con sus grandes avances tecnológicos han provocado a nivel mundial un aumento considerable de la información, llegando a nombrarse la época actual como Era o Sociedad de la Información. Todo ese volumen de datos es almacenado principalmente por las empresas para utilizarlo en dependencia de sus necesidades y extraer la información útil para la toma de decisiones, pero muchas veces este proceso se ve afectado por el mayor problema de principios del siglo XXI “manejar de forma óptima grandes volúmenes de información”.

Durante los últimos años un grupo de instituciones cubanas y el propio Ministerio de Salud Pública (MINSAP) han desarrollado sistemas encaminados a lograr la informatización de la salud. En todos los casos el objetivo ha sido proveer al Sistema Nacional de Salud de información confiable, consistente y oportuna para la toma de decisiones y el mejoramiento de los procesos médicos asistenciales, garantizando de esta manera el incremento en la calidad y seguridad de la atención médica a la población.[1]

Cada resultado obtenido por el gobierno cubano en la esfera de la salud desde el triunfo de la Revolución ha sido compartido en gesto solidario con diversas naciones del mundo. Gracias al legado del pensamiento latinoamericanista de José Martí, Ernesto Che Guevara y la continuidad de estas ideas por el Comandante en Jefe Fidel Castro.

Cuba realiza la primera ayuda médica en el año 1960 al enviar a Chile una brigada médica emergente, para atender los damnificados de un terremoto que afectó a ese país y dejó miles de fallecidos. Durante estas cuatro décadas un total de 79 487 colaboradores distribuidos en 97 países han brindado su ayuda donde se han destacado entre otras funciones en la asistencia médica clínico-quirúrgica a la población en los lugares más apartados e intrincados con la mayor expresión de solidaridad e internacionalismo.

La colaboración médica cubana es una premisa fundamental del sistema nacional de salud y ocupa un lugar priorizado dentro de la política de sus convicciones humanistas. Actualmente su información es procesada por la Unidad Central de Cooperación Médica (UCCM), que cuenta con un proyecto realizado en la Universidad de las Ciencias Informáticas para mantener un registro actualizado de los colaboradores y misiones médicas en el exterior y enviar información detallada y consolidada a los

niveles superiores del MINSAP al Ministerio de Relaciones Exteriores (MINREX) y al consejo de Estado.

La información se encuentra almacenada en una base de datos relacional sobre el sistema gestor MySQL, lo cual le permite almacenar todos los datos relacionados con los colaboradores médicos cubanos y sus respectivas misiones en los distintos países. El requisito fundamental de este almacenamiento es conservar toda la información relevante sobre los colaboradores y misiones médicas por un período de tiempo indefinido lo cual provoca una carga intensa de la base de datos operacional y dificulta una respuesta rápida al cliente.

En Junio del 2009 se presentó una tesis elaborada en la Universidad de las Ciencias informáticas cuyo tema es el de “Implementación de un Datawarehouse para el Control de Recurso Humano de la salud”. La misma apoyó a los directivos del Observatorio Nacional de los Recursos Humanos (RRHH) de la salud en la toma de decisiones con relación a las actividades de colaboración médica y la docencia en la formación posgraduada, donde se combinaron algunos datos del proyecto Colaboración y Docencia Médica para obtener un control de los colaboradores y su formación docente. Este almacén de datos no abarcó toda la información relevante para la UCCM, solamente lo relacionado con los RRHH de la salud.

De ahí la necesidad de contar con un sistema que permita extraer y filtrar una mayor cantidad de datos para transformarlos, integrarlos y almacenarlos con el fin de poder acceder a la información para dar soporte en el proceso de toma de decisiones de los directivos de la UCCM y reducir el tiempo de consultas largas y complejas. Además, poder convertir los datos en información homogénea, de mayor calidad, y que se mantenga en el tiempo.

Luego de analizar el proceso de almacenamiento de la información de la UCCM se identifica como **problema científico**: ¿Cómo almacenar grandes volúmenes de información relevante para la toma de decisiones de los directivos del proyecto Colaboración Médica en un período de tiempo?

El problema planteado delimita el **objeto de estudio**: El proceso de consultoría de la información de la UCCM. El objeto de estudio se enmarca en el **campo de acción**: El proceso de almacenamiento y análisis de la información de la UCCM para la toma de decisiones.

Por lo antes planteado se definió como **objetivo de la investigación:** desarrollar un datamart que contenga la información consolidada del proyecto Colaboración Médica.

Para lograr el objetivo de la investigación se trazaron las siguientes **tareas de investigación:**

1. Identificar las necesidades de información del cliente obteniéndose las preguntas claves del negocio.
2. Diseñar el modelo conceptual del almacén de datos mediante el cual se adquiere una idea precisa del alcance del almacén.
3. Revisar las fuentes de información generadas en la UCCM en las cuales se consiga la información relevante de las mismas.
4. Elaborar el modelo de datos del almacén obteniéndose su estructura para la UCCM.
5. Realizar Extracción, Transformación y Carga para el almacén de datos.
6. Realizar los procesos de análisis de la información contenida en un almacén de datos mediante el cual se agilicen las consultas de la información.

Como resultados se espera contar con un almacén de datos que contenga 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 presente trabajo consta de 3 capítulos, estructurados de la siguiente manera:

Capítulo I: Fundamentación teórica. Se aborda todo lo relacionado con almacenes de datos, así como una descripción de todas las herramientas, tecnologías y metodologías a utilizar para dar solución al problema planteado.

Capítulo II: Análisis y Diseño. Se describen todos los pasos de la metodología utilizada, con el objetivo de facilitar el trabajo que significa la construcción de un datamart. En principio se identifican las necesidades de información del cliente para el diseño de un modelo conceptual, así como los pasos para realizar Extracción, Transformación y Carga (ETL).

Capítulo III: Proceso Analítico en Línea. Se expone la construcción del cubo de datos, donde se definen las jerarquías de los atributos de las dimensiones a las que pertenecen, y la implementación de los indicadores o medidas. Se muestran las herramientas clientes disponibles para que los usuarios puedan consultar la información almacenada, y es representado mediante el diagrama de despliegue

los requerimientos tecnológicos y la integración entre el almacén y las base de datos del sistema de Colaboración Médica.

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

En el capítulo se aborda todo lo relacionado con almacenes de datos, así como una descripción de todas las herramientas, tecnologías y metodologías a utilizar para dar solución al problema planteado.

1. ¿POR QUÉ UN ALMACÉN DE DATOS Y NO UNA BASE DE DATOS OPERACIONAL?

Los datos de un almacén de datos y los datos operacionales son accedidos por usuarios que los utilizan de formas diferentes. Los usuarios que acceden a los datos operacionales, comúnmente efectúan tareas predefinidas y actualizables que requieren acceso a una sola base de datos de una aplicación y almacenan datos actuales. Por el contrario, los usuarios que accedan al almacén, efectúan tareas complejas que requieren acceso a un conjunto de datos desde fuentes múltiples, frecuentemente no son predecibles y almacenan datos históricos.

Características del almacén de datos:

- Maneja un gran volumen de datos, debido a que consolida en su estructura la información recolectada durante largos períodos de tiempo, proveniente de diversas fuentes, en un solo lugar centralizado.
- Organiza y almacena los datos que se necesitan para el procesamiento analítico e informático, con el propósito de responder a las interrogantes de los directivos de una empresa y a la vez brindarles una interfaz amigable, comprensible y fácil de utilizar, para que puedan tomar decisiones sobre los datos sin tener que poseer demasiados conocimientos informáticos.
- Constituye un conjunto de herramientas para consultar, analizar y presentar información, que permiten obtener o realizar análisis, reportes, extracción y explotación de los datos, con alto rendimiento, para transformar dichos datos en información valiosa para la empresa.

Todas las características expuestas anteriormente, son imposibles de realizar en un típico ambiente operacional, esta es una de las razones de ser de un almacén de datos y se justifica su construcción por la necesidad de los directivos de la UCCM de tener una fuente única de información relevante de los colaboradores y misiones médicas para la toma de decisiones, reducir el tiempo de consultas largas y complejas, así como conservar la información por un largo período de tiempo.

1.1. ¿POR QUÉ DATAMART Y NO DATAWAREHOUSE?

1.1.1. ¿QUÉ ES UN DATAWAREHOUSE?

El Datawarehouse posibilita la extracción de datos de sistemas operacionales y fuentes externas, permite la integración y homogenización de los datos de toda la empresa, provee información que ha sido transformada y sumariada, para que ayude en el proceso de toma de decisiones estratégicas y tácticas. [2]

Según **W. H. Inmon** (considerado por muchos el padre del Data Warehouse), 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”. [3]

Según **Ralph Kimball** (considerado el principal promotor del enfoque dimensional para el diseño de almacenes de datos), un *Data Warehouse* es una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis. [4]

Entre las principales características de un almacén de datos se encuentran:

Orientado a temas:

En el almacén de datos la información se clasifica en base a los aspectos que son de interés para la UCCM, se excluye la información que no será usada por el proceso de sistemas de soporte de decisiones. Estará orientado a todos los datos relevantes sobre colaboradores, su localización, misión, movimientos, fechas y otras informaciones significativas.

Integrado:

A través de los años los diseñadores y programadores que han dado su aporte en el desarrollo de la informática en la UCCM no se han basado en ningún estándar para definir nombres de variables, tipos de datos, etc., ya sea por carecer de ellos o por no creer que sean necesarios. Por lo cual, cada uno por su parte ha dejado en cada aplicación, módulo y tabla, su propio estilo personalizado, confluyendo de esta manera en la creación de modelos muy inconsistentes e incompatibles entre sí. La integración implica que los datos de diversas fuentes que son producidos por distintos departamentos, secciones y aplicaciones, tanto internas como externas se consoliden e integren para la toma de decisiones de los directivos de la UCCM.

Variante en el tiempo:

En el datawarehouse los datos de los colaboradores y misiones médicas son almacenados junto a sus respectivos históricos. Gracias al sello de tiempo se puede tener acceso a diferentes versiones de la misma información. Esto permite a los directivos de la UCCM 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 de los colaboradores y misiones médicas una vez que entran en el almacén no cambian. Esto permite elaborar reportes a partir de los datos guardados de forma confiable y sin haber sido actualizada o cambiada como ocurriría en una base de datos operacional.

1.1.2. ¿QUÉ ES UN DATAMART?

Un datamart 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 datawarehouse y un datamart es su alcance. El datamart está pensado para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización. Es el almacén natural para los datos departamentales. En cambio, el ámbito del datawarehouse es la organización en su conjunto. [5] En síntesis los datamarts son pequeños datawarehouses centrados en un tema o un área de negocio específico dentro de una organización.

La información que se necesita para la toma de decisiones en la UCCM se encuentra orientada a un único tema; el de los colaboradores y sus respectivas misiones médicas, por tanto no es necesario la creación de un datawarehouse pues con un datamart se tiene fácil acceso a los datos que se necesitan frecuentemente, crea una vista colectiva para grupos de usuarios, mejora el tiempo de respuesta del usuario final, se destaca por una definición de requerimientos más fácil y rápida, se simplifica el desarrollo de todo el mecanismo de su base de datos y con ello baja substancialmente todo el coste del proyecto, así como su duración.

1.2. METODOLOGÍAS PARA EL DISEÑO E IMPLEMENTACIÓN DE UN DATAMART

1.2.1. LA METODOLOGÍA KIMBALL

Apunta a una solución completa que se puede implementar en poco tiempo. Esto cumple con lo que se necesita en una metodología de Business Intelligence (BI), dar resultados rápidos para demostrar el valor de la solución al negocio y no perder la confianza de los usuarios.

Se enfoca principalmente en el diseño de la base de datos que almacenará la información para la toma de decisiones.

El ciclo de vida de Kimball: [6]

- Ilustra el flujo general de implementación de un datawarehouse (DW).
- Identifica secuencia de tareas ordenadas y actividades principales que deben suceder concurrentemente.
- Muchas necesidades deben ser acomodadas para lograr única necesidad de la organización.
- No todos los detalles de las tareas del ciclo de vida deben ser ejecutados en todos los proyectos.

1.2.2. LA METODOLOGÍA INMON

Por el contrario puede tener una implementación mucho más tardada, y es recomendada cuando se hace demasiado difícil representar el modelo a través de dimensiones y la complejidad de la solución se hace demasiado grande.

Ambas metodologías pueden implementarse en un mismo datawarehouse entonces la recomendación es que si es necesario utilizar esta metodología hacerlo en iteraciones ya avanzadas y siempre empezar con Kimball.

1.2.3. LA METODOLOGÍA HEFESTO

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

La metodología HEFESTO puede resumirse a través del siguiente gráfico:

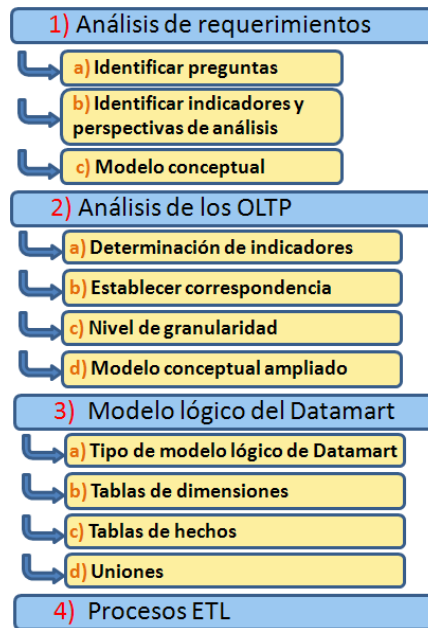


Fig.1 Metodología Hefesto.

Para trabajar con la metodología Hefesto se 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 hecho esto, 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. [7]

Esta metodología cuenta con las siguientes características: [8]

- 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 datawarehouse.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.

- Es independiente de las estructuras físicas que contengan el datawarehouse y de su respectiva distribución.

Mediante los estudios realizados se escoge la metodología Hefesto como guía para la construcción del datamart pues cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente. Resulta independiente del tipo de ciclo de vida que se emplee para su realización y de las herramientas que se utilicen para su implementación, aplicándose tanto para datamart como para datawarehouse. Existe una experiencia previa de su utilización en proyectos y trabajos de diplomas de la facultad; ejemplo la tesis presentada en junio del 2009 en la Facultad 7 con el tema “Implementación de un Datawarehouse para el Control de Recurso Humano de la salud”.

1.3. ¿QUÉ ES OLTP?

Procesamiento Analítico Transaccional (*On Line Transaction Processing*), representa toda aquella información transaccional que genera la empresa en su accionar diario, además, de las fuentes externas que puede llegar a disponer, estas fuentes de información, son de características muy disímiles entre sí, en formato, procedencia, función, etc. [9]

Entre los OLTP más habituales que pueden existir en cualquier organización se encuentran: [10]

- Archivos de textos.
- Hipertextos.
- Hojas de cálculos.
- Informes semanales, mensuales, anuales, etc.
- Bases de datos transaccionales.

1.4. MODELADO 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 rápida. El uso del Modelo Multidimensional es una de las aproximaciones más acertadas y seguidas por los especialistas.

Los datamarts gestionan el depósito de datos y lo organizan en forma de una base de datos multidimensional que almacena los datos en diversas dimensiones, conformando un cubo multidimensional, en donde el cruce de los valores de los atributos de cada dimensión a lo largo de las abscisas, determinan un hecho específico que se define como datos instantáneos en el tiempo, que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de dimensiones. (Ver Fig.2)

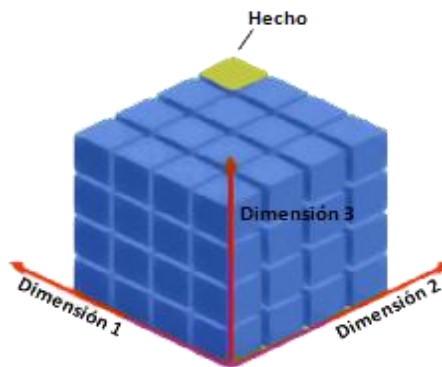


Fig.2 Cubo multidimensional.

1.4.1. ESQUEMA ESTRELLA

Un esquema en estrella es un modelo de datos que tiene una tabla de hechos que contiene los datos para el análisis, rodeada de las tablas de dimensiones. Este aspecto, de tabla de hechos (o central) más grande rodeada de radios o tablas más pequeñas es lo que asemeja a una estrella, dándole nombre a este tipo de construcciones.

Las tablas de dimensiones tendrán siempre una clave primaria simple, mientras que en la tabla de hechos, la clave principal estará compuesta por las claves principales de las tablas dimensionales.

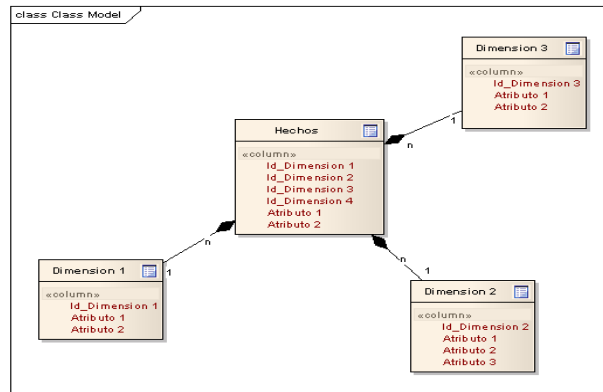


Fig.3 Esquema estrella.

1.4.2. ESQUEMA COPO DE NIEVE

En un esquema en copo de nieve se tiene una tabla central de hechos en la que se guardan las medidas del negocio que se quiere analizar, y en las tablas adyacentes se tendrán las dimensiones (parámetros) de que dependen los datos del negocio. Si por alguna dimensión se requiere más de una tabla se dice que el esquema resultante es un esquema en copo de nieve.

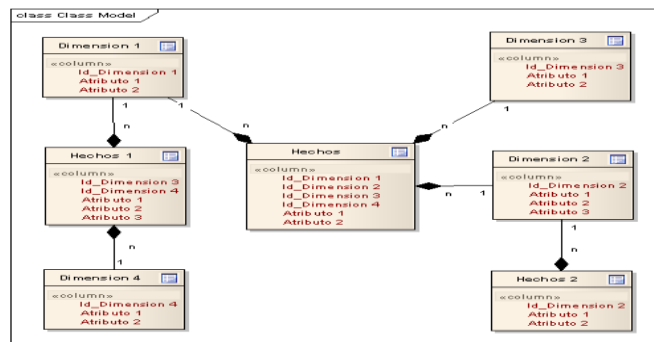


Fig.4 Esquema copo de nieve.

1.4.3. ESQUEMA DE CONSTELACIÓN

El esquema de constelación está compuesto por una serie de esquemas en estrella, está formado por una tabla de hechos principal y por una o más tablas de hechos auxiliares. Las tablas de hechos se ubican en el centro del modelo y se relacionan con sus respectivas tablas de dimensiones. Este esquema es similar al esquema de estrella, permite tener más de una tabla de hechos, contribuye a la reutilización de dimensiones y no es soportado por todas las herramientas de análisis y consulta.

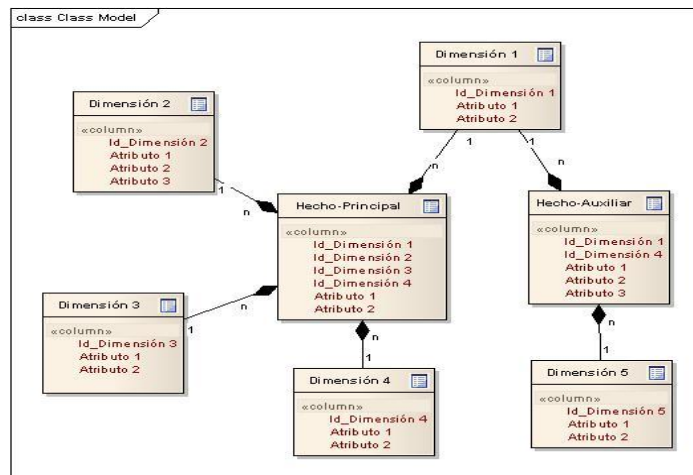


Fig.5 Esquema de constelación.

Dado un estudio previo de los esquemas para el modelado multidimensional se utilizará el esquema de estrella pues es ideal por su simplicidad y velocidad para ser usado en análisis multidimensionales (OLAP, Datamarts, *Executive Information System* (EIS), entre otros). Además, el diseño de esquemas en estrella permite implementar la funcionalidad de una base de datos multidimensional utilizando una base de datos relacional. Otra razón para utilizar este tipo de esquema es su sencillez desde el punto de vista del usuario final. Las consultas no son complicadas. Por último, es el esquema con mejor rendimiento y velocidad pues permite indexar las dimensiones de forma individualizada sin que repercuta en el rendimiento de la base de datos en su conjunto.

1.5. ETL

ETL Extracción, Transformación y Carga (*Extract, Transform and Load*) es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, datamart, o datawarehouse.

Los procesos de Extracción, Transformación y Carga constan de múltiples pasos, cuyo objetivo es transferir datos desde las aplicaciones de producción a los sistemas de Inteligencia de negocio: [11]

- Extracción de los datos desde las aplicaciones y bases de datos de producción (ERP, CRM, RDBMS, archivos, etc.)
- Transformación de estos datos para reconciliarlos en todos los sistemas source, realizar cálculos o análisis sintáctico de cadenas, enriquecerlos con información de búsqueda externa y,

además, adaptarlos al formato preciso por el sistema objetivo (*Third Normal Form, Star Schema, Slowly Changing Dimensions, etc.*)

- Carga de los datos resultantes en las diversas aplicaciones de BI: almacenes de datos históricos generales (datawarehouse) o almacenes de datos empresariales, almacenes de datos históricos individuales (datamart), aplicaciones OLAP (Procesamiento analítico en línea) o “cubos”, etc.

1.6. OLAP

OLAP procesamiento analítico en línea (*On-Line Analytical Processing*). 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. [11] 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, cambiar de datos resumidos a detallados, etc.[12]

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:

1.6.1. 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, etc. [13]

1.6.2. 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 arrays de varias dimensiones, llamados cubos. Estos cubos utilizan índices para optimizar el acceso a los datos. [14]

1.6.3. HOLAP

Herramientas HOLAP (*Hybrid On-line Analytical Process*). Permiten un análisis híbrido de la información, es decir que une lo mejor de los dos tipos anteriores. El análisis HOLAP ayudará 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. [15]

Para la construcción del datamart para la UCCM se utilizará la herramienta ROLAP puesto 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. Además soporta técnicas de optimización de accesos para acelerar las consultas y se accede directamente a los datos del datamart.

1.7. HERRAMIENTAS PARA LA CONSTRUCCIÓN DEL DATAMART

1.7.1. SISTEMAS GESTORES DE BASE DE DATOS



Fig.6 Sistemas gestores de base de datos.

1.7.1.1. 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: [16]

- 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.

1.7.1.2. MYSQL

MySQL es un sistema de gestión de bases de datos relacional, licenciado bajo la GPL (acrónimo en inglés de GNU) de la GNU (*General Public License*). 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: [17]

1. Aprovecha la potencia de sistemas multiprocesador, gracias a su implementación multihilo.
2. Soporta gran cantidad de tipos de datos para las columnas.
3. Dispone de API's en gran cantidad de lenguajes (C, C++, Java, PHP, etc.).
4. Gran portabilidad entre sistemas.
5. Soporta hasta 32 índices por tabla.

6. Gestión de usuarios y passwords, manteniendo un muy buen nivel de seguridad en los datos.

1.7.1.3. 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. A pesar de esto, PostgreSQL no es un sistema de gestión de bases de datos puramente orientado a objetos.

A continuación se enumeran las principales características de este gestor de bases de datos: [18]

1. Implementación del estándar (lenguaje) SQL92/SQL99.
2. 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.
3. Incorpora una estructura de datos array.
4. Incorpora funciones de diversa índole: manejo de fechas, geométricas, orientadas a operaciones con redes, etc.
5. Permite la declaración de funciones propias, así como la definición de disparadores.
6. Soporta el uso de índices, reglas y vistas.
7. 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.
8. Permite la gestión de diferentes usuarios, como también los permisos asignados a cada uno de ellos.
9. Soporta casi toda la sintaxis SQL (incluyendo sub consultas, transacciones, tipos y funciones definidas por el usuario).
10. Cuenta con un amplio conjunto de enlaces con lenguajes de programación como por ejemplo C, C++, Java, Python, PHP y otros.
11. El progreso continuo del gestor de datos de código abierto PostgreSQL brinda a los consumidores la opción de instalar una base de datos no privativa.

Después del estudio realizado sobre los distintos sistemas gestores de base de datos se utilizará el PostgreSQL pues posee licencia GPL por lo que no tiene costo asociado a la licencia del software. 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), contando también con un amplio conjunto de enlaces con lenguajes de programación como por ejemplo C, C++, Java, Python, PHP y otros. Todo esto lo convierte en la herramienta ideal para el desarrollo del datamart para la UCCM.

1.7.2. HERRAMIENTAS DE INTEGRACIÓN DE DATOS

Algunas de las herramientas más populares que apoyan la integración de datos:



Fig.7 Herramientas de integración de datos (ETL)

1.7.2.1. 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: [19]

- Realiza transformaciones gráficas basadas en XML para la descripción de los metadatos de los registros.
- Se distribuye bajo la licencia LGPL.
- CloverETL 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.

1.7.2.2. 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.

Características principales: [20]

- 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.

1.7.2.3. 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: [21]

- 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.

Características principales:

- 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 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 datamart para la UCCM 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.

1.7.3. HERRAMIENTAS PARA OLAP

En la figura se muestran algunos de los proveedores de OLAP más populares que se describen más detalladamente a continuación.



Fig.8 Herramientas para OLAP.

1.7.3.1. 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. [22]

1.7.3.2. 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. [23]

1.7.3.3. MONDRIAN

Mondrian es un motor de ROLAP desarrollado en Java que permite analizar grandes conjuntos de datos almacenados en un datawarehouse. 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). [24]

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) de Mondrian directamente sino que implementa su propio modelo OLAP. [25]

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. [26]

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.

1.7.3.4. 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.

1.7.4. HERRAMIENTAS CASE



Fig.9 Herramientas Case.

1.7.4.1. RATIONAL ROSE

Rational Rose es la herramienta CASE que comercializan los desarrolladores de UML 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. [27]

1.7.4.2. 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. [28]

1.7.4.3. 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. Es una herramienta flexible para la plataforma de Windows y Linux, que será utilizada para realizar el diseño de los diferentes artefactos del almacén. [29]

Entre las herramientas case estudiadas se escoge para el modelado y construcción del datamart al Enterprise Architect pues provee beneficios significativos para ayudar a construir modelos de sistemas de software rigurosos y donde es posible mantener la trazabilidad de manera consistente y soporta este proceso en un ambiente fácil de usar, rápido y flexible. Además, es la herramienta propuesta en el documento de la arquitectura del departamento Sistemas de Apoyo a la Salud (SAS).

1.7.5. SERVIDOR DE APLICACIONES



Fig.10 Servidores de aplicaciones.

1.7.5.1. JBoss

JBoss Application Server es el servidor de aplicaciones de código abierto más ampliamente desarrollado del mercado. Combinando una arquitectura orientada a servicios, revolucionaria con una licencia de código abierto, JBoss AS puede ser descargado, utilizado, incrustado y distribuido sin restricciones por la licencia. [30]

1.7.5.2. TOMCAT

Tomcat es un servidor Web con soporte de servlets y JSPs. Incluye el compilador Jasper, que compila JSPs convirtiéndolas en servlets. El motor de servlets de Tomcat a menudo se presenta en combinación con el servidor Web Apache.

Tomcat puede funcionar como servidor Web por sí mismo, que puede ser usado como servidor Web autónomo en entornos con alto nivel de tráfico y alta disponibilidad. Dado que Tomcat fue escrito en Java, funciona en cualquier sistema operativo que disponga de la máquina virtual. [31]

Se selecciona como servidor de aplicaciones para la construcción del datamart el JBoss ya que es un servidor de aplicaciones de código abierto implementado en Java. Al estar basado en Java, puede ser utilizado en cualquier sistema operativo que lo soporte. Combina una arquitectura orientada a servicios revolucionaria con una licencia de código abierto permitiendo ser descargado, utilizado, incrustado y distribuido sin restricciones por la licencia.

1.8. PROPUESTA DE LA PLATAFORMA A UTILIZAR



Fig.11 Plataforma Pentaho.

La plataforma Open Source Pentaho Business Intelligence cubre muy 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. [32]

1.8.1. 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. Pentaho Reporting permite 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. [33]

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 datos, 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). [34]

Integración de Datos se realiza con una herramienta Kettle ETL (*Pentaho Data Integration*) que permite implementar los procesos ETL. Últimamente Pentaho lanzó una nueva versión - PDI 3.0 – que marcó un gran paso adelante en OSBI ETL y que hizo Pentaho Data Integration una alternativa interesante para las herramientas comerciales.[35]

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 seleccionadas durante el estudio de las herramientas de integración de datos y las herramientas para OLAP.

1.9. DATAMARTS EXISTENTES

1.9.1. NIVEL INTERNACIONAL

La mayoría de las empresas del mundo ya cuentan de una manera u otra con diferentes datamarts, esto se debe a que tienen la necesidad constante de consumir información y hacer un mejor uso de los datos ante un mercado más competitivo. Una de estas empresas es Petroecuador que cuenta con un datamart 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.

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 datamart, esto les permite desarrollar reportes sobre precios, descuentos y promociones vigentes además del vencimiento de productos terminados.

Cantv Corporación se beneficia de un datamart de clientes PYME (pequeña y mediana empresa) otorgándole a corto plazo información valiosa que no se encontraba disponible en otro sistema acerca de los clientes, permitiendo establecer prioridades a la hora de la carga de datos de sus diferentes filiales estandarizando los datos de los mismos y actualizándolos de acuerdo a sus prioridades.

1.9.2. NIVEL NACIONAL

Las empresas cubanas en la época actual necesitan obtener de una forma cada vez más rápida la información necesaria para tomar decisiones. Con la utilización de las tecnologías de la información y las comunicaciones trata de alcanzar la mayor productividad posible y que los productos tengan una mejor calidad para obtener ventajas competitivas. El uso de datamarts es una de las estrategias para lograrlo. Ejemplos de tal desempeño:

La Empresa de Proyectos de Arquitectura e Ingeniería (EMPAI) de Matanzas posee un datamart 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 datamart, 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.

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 datamart 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”

1.9.3. UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

En la universidad hasta la fecha se han realizado dos trabajos de tesis relacionados con el tema “Datamarts”. Una de ellas “GESTCON MART, DATA MART PARA LA GESTIÓN DEL CONOCIMIENTO” el mismo sirve de soporte a una base de experiencia para la gestión de proyectos de software, con el propósito de hacer un mejor uso de la información y facilitar su posterior análisis en el proceso de toma de decisiones.

Otro de los trabajos de diploma “Diseño del Datamart del subsistema de Conducidos”, con el objetivo de tener una 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.

Conclusiones del capítulo:

En el capítulo se realizó un estudio sobre almacenes de datos, sus ventajas y beneficios para los directivos de la UCCM a la hora de tomar decisiones. Se escogieron las herramientas y la metodología a utilizar como guía para la construcción del datamart. Además, se realizó un estudio sobre los datamarts existentes a nivel internacional, nacional y en la universidad.

CAPÍTULO 2. ANÁLISIS Y DISEÑO

Se describen todos los pasos de la metodología utilizada, con el objetivo de facilitar el trabajo que significa la construcción de un datamart. En principio se identifican las necesidades de información del cliente para el diseño de un modelo conceptual, así como los pasos para realizar extracción, transformación y carga de los datos en el datamart.

2. PASOS Y APLICACIÓN DE LA METODOLOGÍA HEFESTO

2.1.1. ANÁLISIS DEL PROCESO

En el capítulo anterior se propuso trabajar con la metodología Hefesto a la cual se le dará seguimiento paso por paso en este capítulo. Para obtener una mayor visión y entendimiento de lo que se realiza, se construye con la herramienta Enterprise Architect un flujo de procesos que representa toda una serie de pasos o iteraciones que se deben seguir hasta concluir con el funcionamiento datamart para la UCCM (vale aclarar que este modelo no pertenece a ninguno de los pasos de la metodología Hefesto).

En la figura 11 se muestra el análisis del conjunto de procesos a seguir para la construcción del datamart. El flujo de trabajo comienza cuando el usuario, en este caso uno de los directivos de la UCCM solicita información para la toma de decisiones para iniciar una serie de procesos hasta llegar al resultado final.

Lo primero a analizar será el proceso de extracción, transformación y carga, el cual tiene como entrada fundamental los recursos que se consumirán durante el mismo. Estos recursos son los sistemas operacionales involucrados, en este caso la base de datos de colaboración médica, la cual almacena los datos que se extraerán para su transformación y posterior carga. Luego se realiza el análisis en línea, teniendo como entrada el propio almacén con las dimensiones requeridas y atributos necesarios para la construcción del cubo de datos y las sentencias SQL. La salida es la información solicitada, que será representada en forma gráfica o tabla, para su mejor comprensión y estudio a la hora de tomar decisiones y la información consolidada que representa la meta a cumplir según la información solicitada.

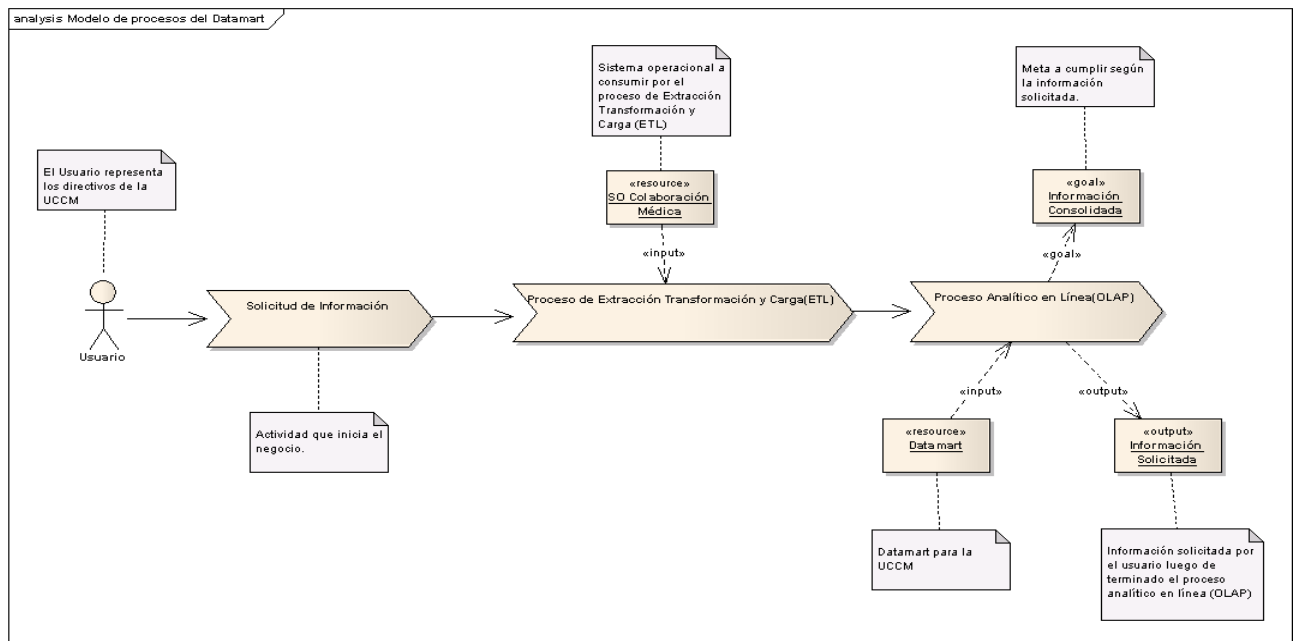


Fig.12 Modelo de procesos.

2.2. PASO1. ANÁLISIS DE REQUERIMIENTOS

2.2.1. IDENTIFICAR PREGUNTAS

Para realizar el primer paso de la metodología se comienza recolectando las necesidades de información de los directivos de la UCCM y se obtienen las preguntas claves del negocio. Para ello se les realizó una entrevista con el propósito de conocer cuales 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 era difícil manejar.

De la entrevista se obtuvieron los siguientes resultados:

- ✓ Se desea conocer el total de colaboradores que han cumplido misión por países en un período por años.
- ✓ Se desea conocer la cantidad de colaboradores por tipo de misión en un determinado país en un año.

- ✓ Se desea conocer la cantidad de colaboradores con más de 5 años de misión en un determinado país.
- ✓ Se desea conocer el total de colaboradores por especialidades en un período de tiempo dado.
- ✓ Se desea conocer el por ciento de colaboradores femeninas que cumplieron misión en un país en un período de tiempo.
- ✓ Se desea conocer la cantidad de países con más de tres años de misión (países en los cuales Cuba tiene cooperantes hace más de tres años)
- ✓ Se desea conocer el total de colaboradores por área geográfica que obtuvieron evaluación satisfactoria en la misión en un determinado período de tiempo.
- ✓ Se desea conocer el total de colaboradores médicos que cumplieron misión en un área geográfica determinada en un período de tiempo dado.
- ✓ Se desea conocer el importe total por nómina congelada de los colaboradores en X misión en un período de 5 años.
- ✓ Se desea conocer el importe total por nómina ayuda familiar de los colaboradores en X misión en un período de 5 años.
- ✓ Se desea conocer el importe total destinado al pago de los colaboradores por tipo de misión en un período de tiempo (suma de las 3 cuentas).
- ✓ Se desea conocer el total de colaboradores con nómina vitalicia por provincia.
- ✓ Se desea conocer el total de tipos de pasaportes confeccionados en un período de tiempo.
- ✓ Se desea conocer el total de desertores por países en X misión en un período de tiempo determinado.
- ✓ Se desea conocer el total de misiones por área geográfica en un período de tiempo.
- ✓ Se desea conocer el por ciento de misiones de X tipo por país.

2.2.2. INDICADORES Y PERSPECTIVAS DE ANÁLISIS

Indicadores:

Son en general valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: saldos, promedios, cantidades, sumatorias, fórmulas, etc. [36]

Perspectivas:

Se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de responder a las preguntas planteadas. [37]

A partir de las preguntas identificadas en la entrevista realizada a los directivos de la UCCM se identificaron los indicadores y perspectivas de análisis:

Total de colaboradores que han cumplido misión por países en un período de tiempo determinado.

Indicador

Perspectivas

Cantidad de colaboradores por tipo de misión en un determinado país en un período de tiempo.

Indicador

Perspectivas

Cantidad de colaboradores con más de 5 años de misión en un determinado país.

Indicador

Perspectivas

Total de colaboradores por dedicación en un período de tiempo dado.

Indicador

Perspectivas

Por ciento de colaboradoras femeninas que cumplieron misión en un país en un período de tiempo.

Indicador

Perspectivas

Cantidad de países con más de tres años de misión.

Indicador

Perspectivas

Total de colaboradores por área geográfica que obtuvieron evaluación satisfactoria en la misión en un

Indicador

Perspectivas

período de tiempo.

Perspectivas

Total de colaboradores médicos que cumplieron misión en un área geográfica determinada en un

Indicador

Perspectivas

período de tiempo dado.

Total de importe por nómina congelada de los colaboradores en X misión en un período de tiempo.

Indicador

Perspectivas

Total de importe por nómina ayuda familiar de los colaboradores en X misión en un período de tiempo.

Indicador

Perspectivas

Importe total destinado al pago de los colaboradores por tipo de misión en un período de tiempo.

Indicador

Perspectivas

Total de colaboradores con nómina vitalicia por provincia.

Indicador

Perspectivas

Total de tipos de pasaportes confeccionados en un período de tiempo.

Indicador

Perspectivas

Total de desertores por países en X misión en un período de tiempo determinado.

Indicador

Perspectivas

Total de misiones por área geográfica en un período determinado de tiempo.

Indicador

Perspectivas

Por ciento de misiones de X tipo por país.

Indicador

Perspectivas

2.2.3. MODELO CONCEPTUAL

Una vez identificados los indicadores y perspectivas se procede a la confección del modelo conceptual de datos del datamart, en el que se reflejan las perspectivas a la izquierda de la figura 13 y los indicadores a la derecha de la misma figura proporcionando una idea precisa y clara del alcance del almacén.

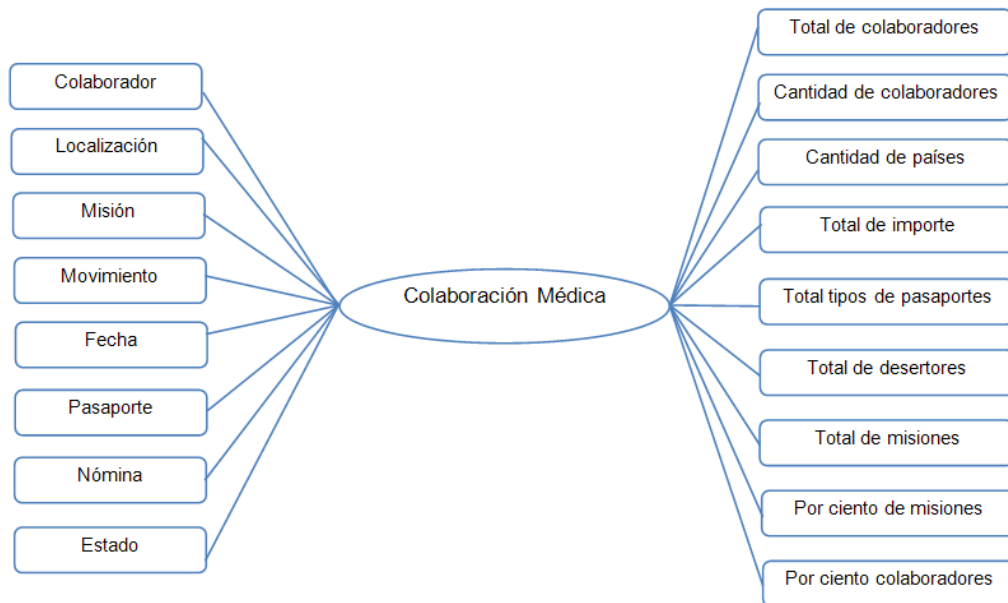


Figura 13. Modelo Conceptual.

2.3. PASO2. ANÁLISIS DE LOS OLTP

2.3.1. CORRESPONDENCIA CON LOS REQUERIMIENTOS

El objetivo de realizar este paso es indagar sobre los OLTP disponibles de colaboración médica para poder identificar la correspondencia entre el modelo conceptual y las fuentes de datos.

A continuación se describe como serán calculados cada uno de los indicadores con el objetivo de que todos los elementos del modelo conceptual estén correspondidos en los OLTP.

Cálculo de los indicadores:

- ✓ “Total de colaboradores” representa la suma de todos los id_colaborador de los colaboradores que hayan cumplido misión en x país, por especialidades, por área geográfica

que obtuvieron evaluación satisfactoria en la misión, médicos que cumplieron misión en un área geográfica determinada o colaboradores con nómina vitalicia por provincia en un período de tiempo determinado.

- ✓ “Cantidad de colaboradores” representa el total de los id_colaborador de los colaboradores por tipo de misión o con más de 5 años de misión en un determinado país en un período de tiempo dado.
- ✓ “Por ciento de colaboradores” es el porcentaje del total de colaboradores que representan las colaboradoras femeninas que cumplieron misión en un país determinado en un período de tiempo dado.
- ✓ “Cantidad de países” representa el total de los países en los cuales Cuba tiene cooperantes hace más de tres años.
- ✓ “Importe total” representa la suma del importe de los colaboradores por nómina congelada, la suma del importe por nómina ayuda familiar de los colaboradores; ambas en X misión en un período de 5 años y la suma de las 3 cuentas (congelada, familiar y vitalicio) destinado al pago de los colaboradores por tipo de misión en un periodo de tiempo.
- ✓ “Total de tipos de pasaportes” representa el total de tipos de pasaportes (oficial, diplomático...) confeccionados en x período de tiempo.
- ✓ “Total de desertores” representa el total de colaboradores que han desertado por países en X misión, se obtiene a partir de contar los movimientos cuyo motivo sea la deserción en el período de tiempo dado.
- ✓ “Total de misiones” representa el total de id_misión por área geográfica en un período determinado de tiempo.
- ✓ “Por ciento de misiones” representa el por ciento de un tipo de misión con respecto a las misiones en un determinado país.

2.3.2. CAMPOS QUE INTEGRAN LAS PERSPECTIVAS. NIVEL DE GRANULARIDAD

Luego de haberse recolectado toda la información, los directivos de la UCCM presentaron los campos principales y de interés de cada perspectiva seleccionada de la base de datos de colaboración médica

(Anexo1) que ayudaran a consultar los indicadores, los resultados se exponen en la figura 14 que representa el modelo conceptual con cada uno de los campos o atributos elegidos para cada perspectiva:

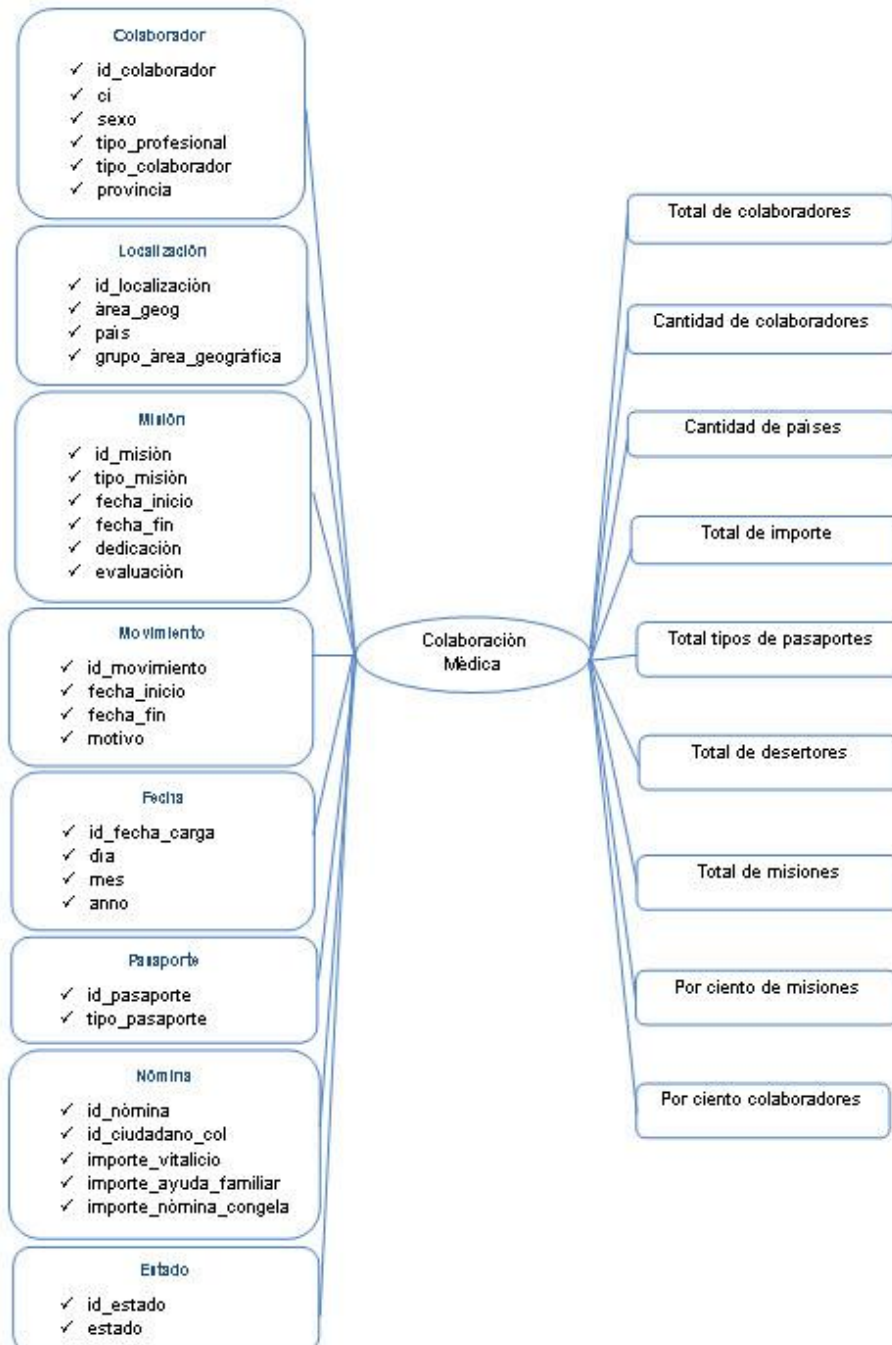


Figura 14. Modelo conceptual con atributos.

2.4. PASO3. MODELO LÓGICO DE LA ESTRUCTURA DEL DATAMART

En este paso se confecciona el modelo lógico de la estructura del datamart. Para obtener la estructura del almacén se selecciona el esquema en estrella por todas las ventajas mencionadas en el capítulo anterior. En la figura 15 se representa según el esquema seleccionado el diseño de las dimensiones y tabla de hechos con sus atributos asociados.

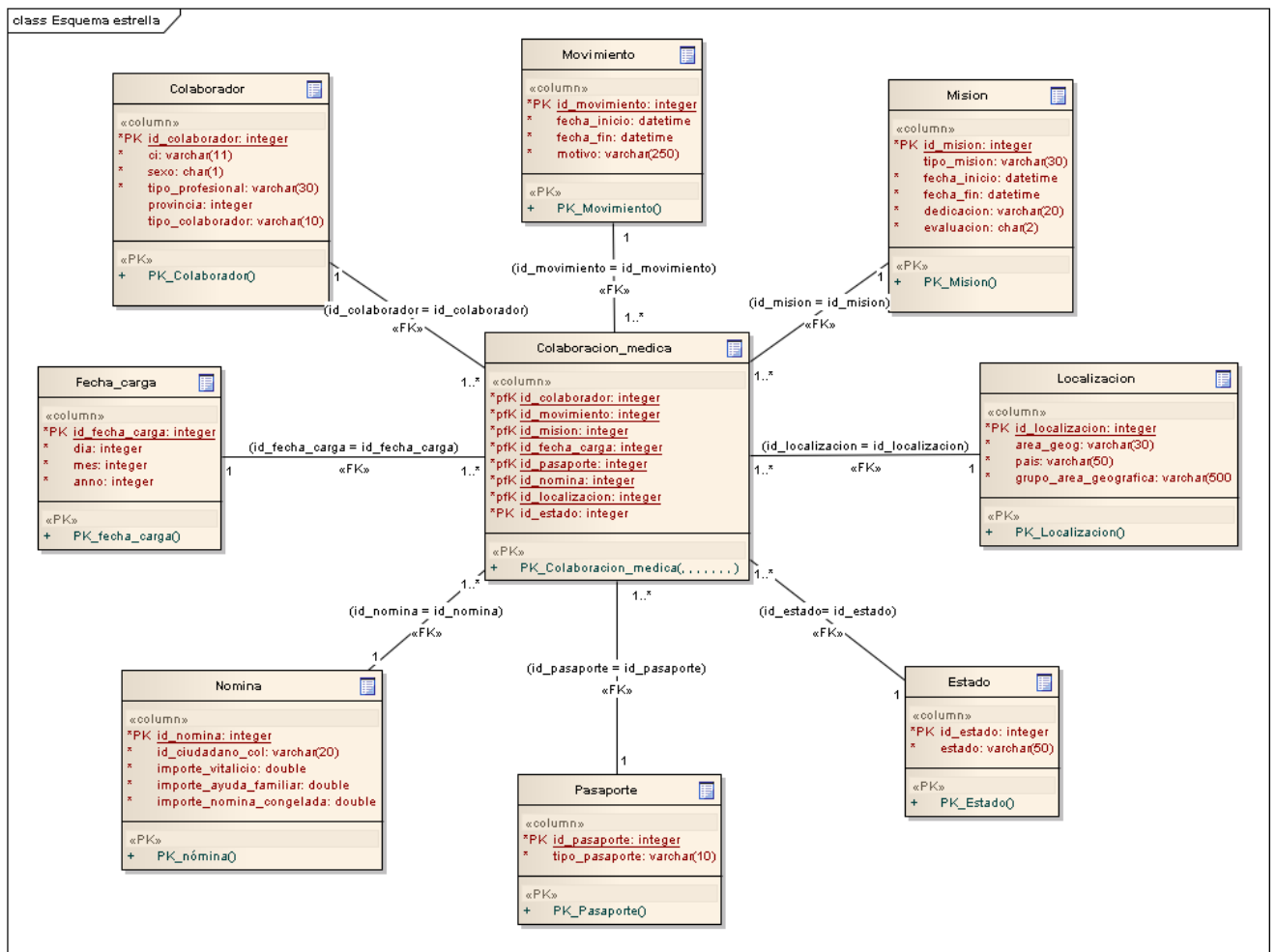


Figura 15. Estructura del datamart para la UCCM.

2.5. PASO4. PROCESO DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA (ETL)

Luego de creado el almacén que soportará las dimensiones que contienen los datos y después de su extracción y posterior transformación desde los sistemas operacionales que se deseen integrar

mediante el proceso ETL se carga el almacén con los datos previstos anteriormente. Para llevar a cabo el soporte de esta tarea se empleó la herramienta antes descrita: la suite Pentaho, Kettle.

En el caso de la dimensión fecha_carga el proceso inicia insertando en la dimensión la fecha (día, mes, año), que es la fecha actual del servidor para así contar con un registro de cuando se realiza la carga del almacén. Para obtener dichos atributos se procedió de la siguiente manera: ver (Figura 16)



Figura 16. Carga de la dimensión fecha_carga.

Se utilizó el componente (Información de sistema) que permite obtener la fecha actual en la que se cargan los datos. A través de un componente Script (Figura 17) se especifica el nombre de las variables que contienen los valores así como el tipo de las mismas para transformarlas y que lleguen al almacén en el formato deseado. Luego para realizar la inserción a la dimensión se utiliza el componente (Selecciona/Renombra) que se encarga de el mapeo de los datos (Figura 18) para mediante el componente (insertar/actualizar) realizar el llenado de la dimensión fecha del datamart (Figura 19).

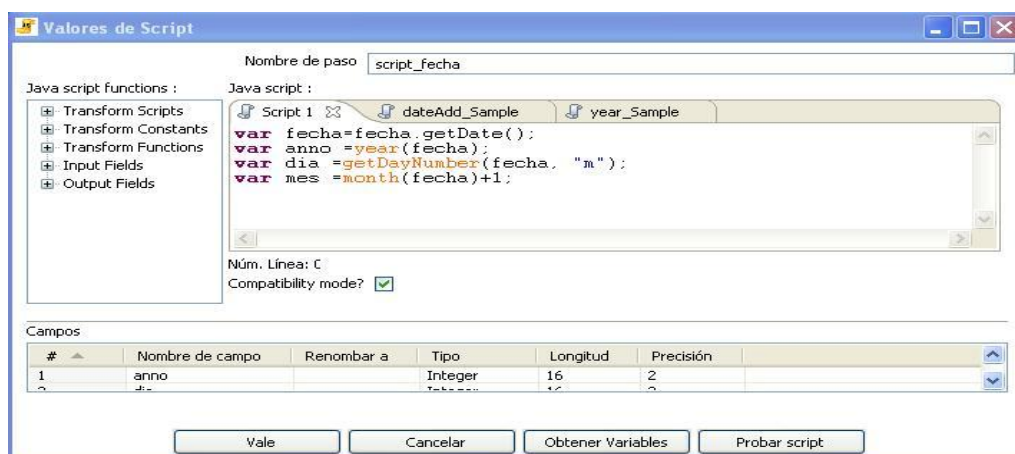


Figura 17. Script de la dimensión fecha.

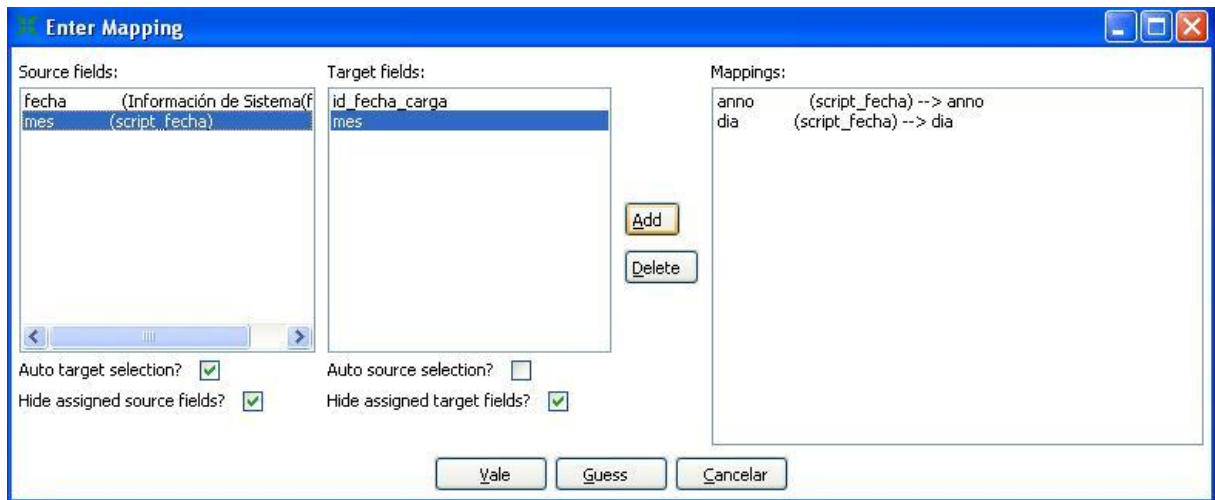


Figura 18. Mapeo de datos de la dimensión fecha.

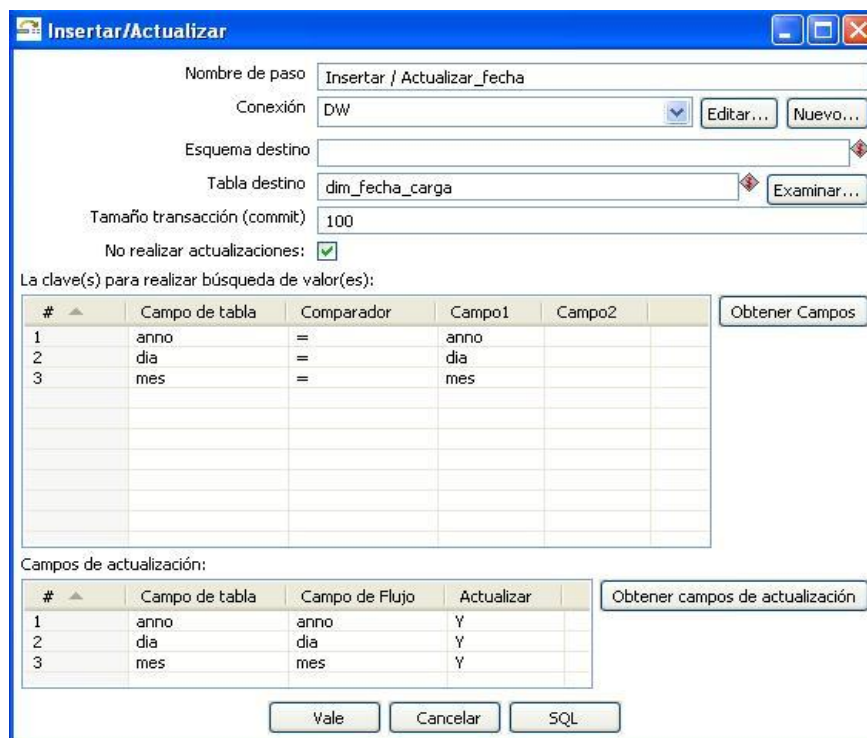


Figura 19. Insertar fecha.

En el caso de la dimensión colaborador, se extraen los atributos mediante el componente (Tabla colaborador) utilizando sentencias sql aunque no todos se encuentran en la base de datos colaboración si no en un repositorio de servicios informatizados de la salud perteneciente a Softel.

Para obtenerlos se consume el servicio registro de ubicación que se almacena en un archivo XML. Con el componente (Entrada/XML) se especifica el archivo XML obtenido, agrupándolos en una misma tupla con el componente (Juntar Filas). Los datos serán transformados utilizando el componente (Script Transformación) para que lleguen a la base de datos en el formato deseado luego con el componente (Filtrar filas) se obtienen las filas con los datos esperados y con los componentes (Selecciona/Renombrar) e (Insertar/Actualizar) se cargan los datos en el datamart. (Figura 20).

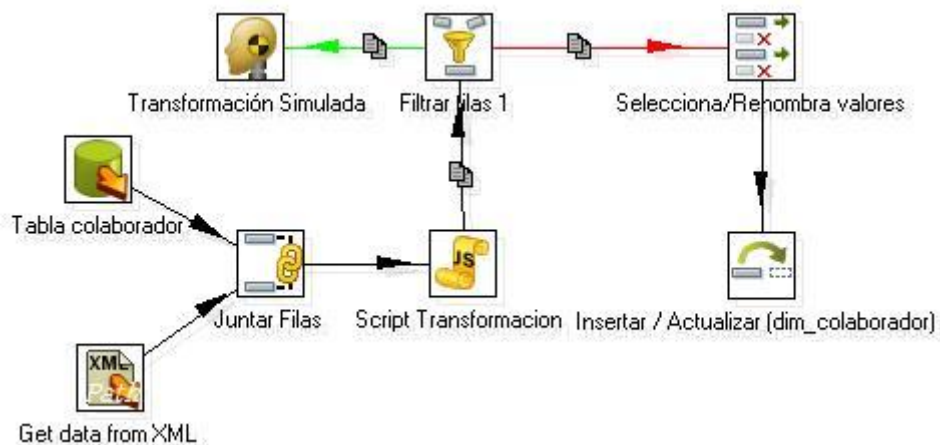


Figura 20. Carga de la dimensión colaborador.

El proceso de carga (Anexo 2) es similar para las 6 dimensiones restantes, por lo cual no se tiene en cuenta la explicación de las mismas

Para la carga de datos de la tabla de hechos, que tiene como objetivo relacionar todas las dimensiones del almacén, se realiza una búsqueda en la base de datos de colaboración médica para a través de sentencias sql extraer los indicadores comunes y relacionar los datos entre todas las dimensiones y unirlos mediante la realización de un producto cartesiano, contando con la fecha de carga de los datos (Figura 21).

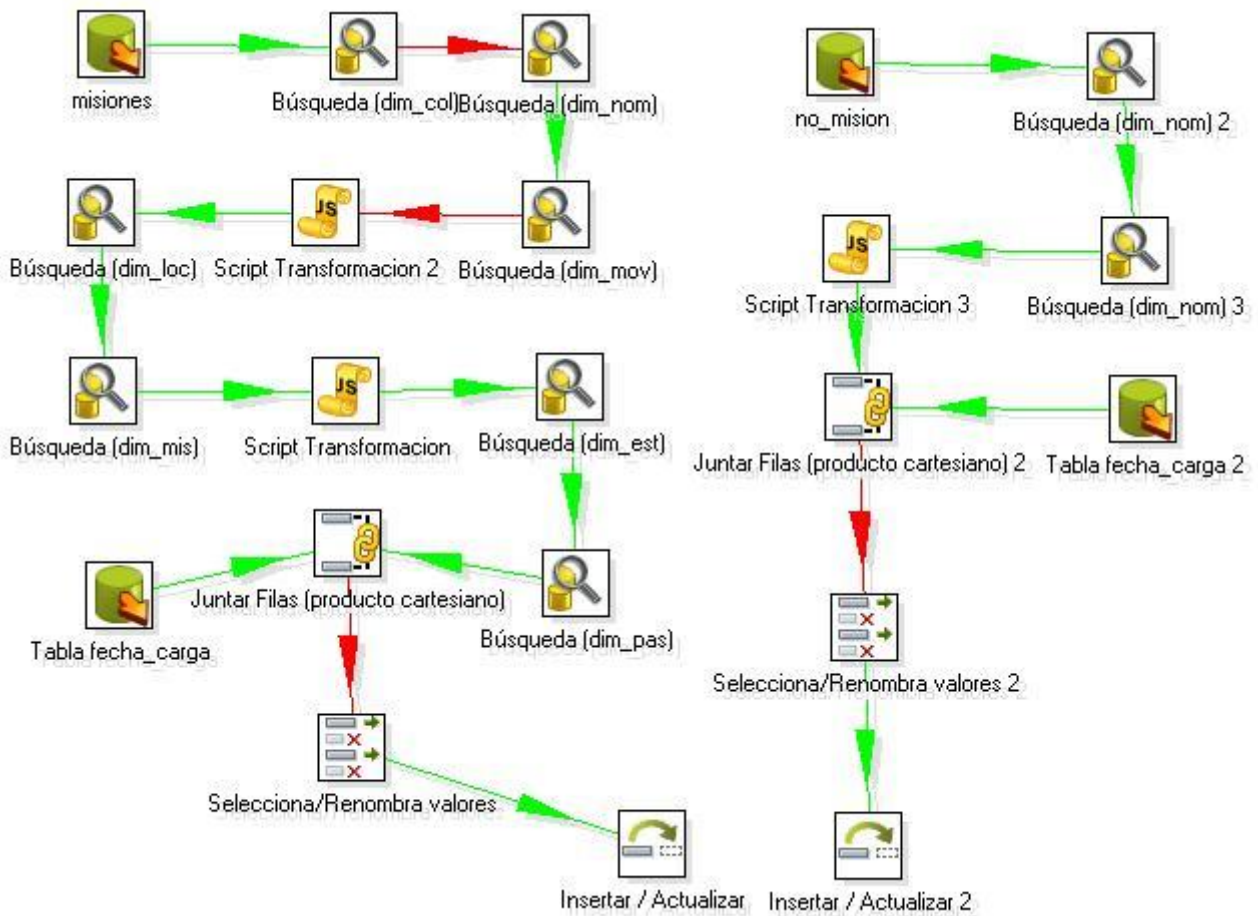


Figura 21. Carga de la tabla de hechos.

A continuación de la carga de la tabla de hechos se crea un trabajo (*job*), mediante el mismo se ejecutan todas las cargas de las dimensiones del almacén. Para iniciar este proceso se utiliza el componente (*Start*) el cual da comienzo a la extracción, transformación y carga por la dimensión fecha_carga y culminando con la tabla de hechos, contando para el manejo de errores en cualquier etapa con el componente (*Dummy*) que informa de errores mediante el componente (*Display MsgBox Info*) al usuario (Figura 22).

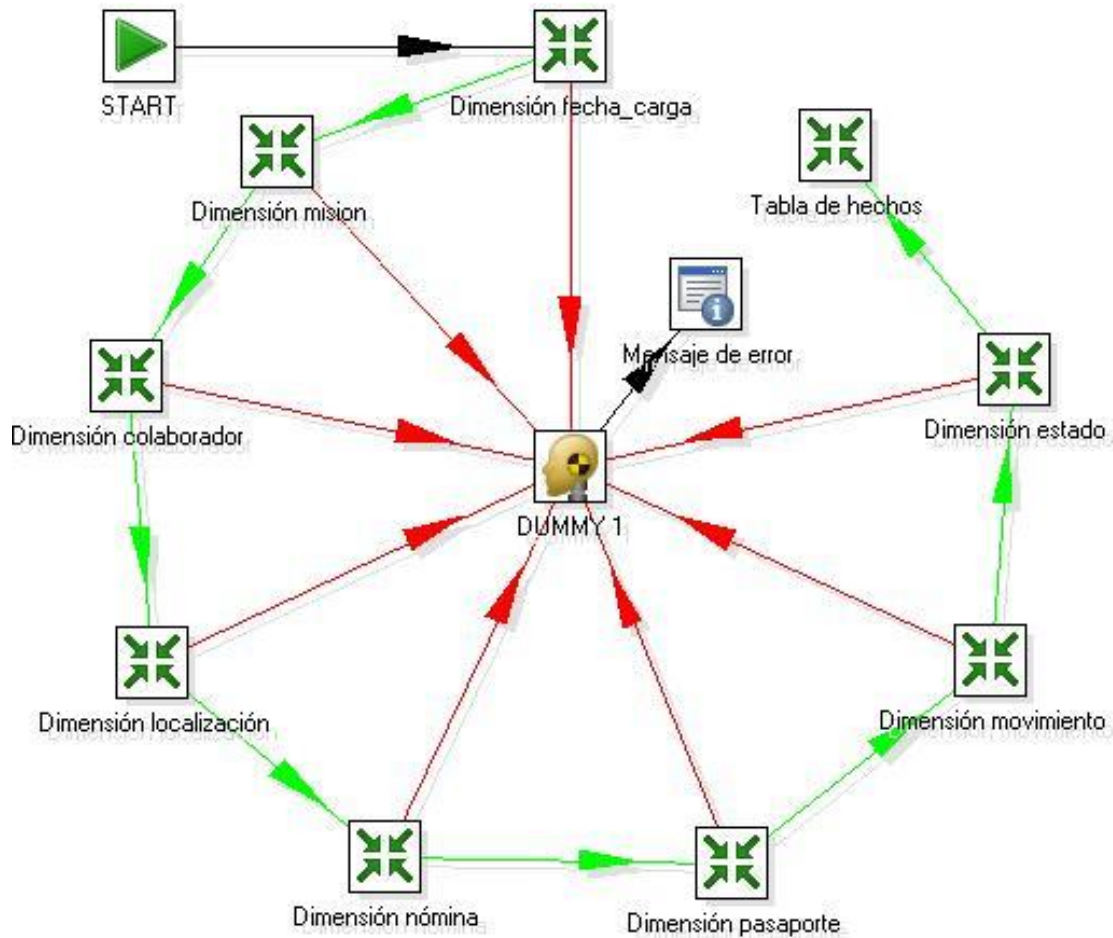


Figura 22. Ejecución del trabajo.

Conclusiones del capítulo:

En el capítulo se desarrollaron cada uno de los pasos de la metodología Hefesto. Se obtuvieron las preguntas claves del negocio, identificándose los indicadores y perspectivas del análisis. Se elaboró el modelo conceptual de datos proporcionando una idea clara del alcance del almacén. Además, se confeccionó el modelo lógico de la estructura del datamart para luego realizar el proceso de extracción, transformación y carga de los datos en el almacén.

CAPÍTULO 3. PROCESO ANALÍTICO EN LÍNEA

En el presente capítulo se muestra la construcción del cubo de datos, donde se definen las jerarquías de los atributos de las dimensiones a las que pertenecen, y la implementación de los indicadores o medidas. Se muestran las herramientas clientes disponibles para que los usuarios puedan consultar la información almacenada, y es representado mediante el diagrama de despliegue los requerimientos tecnológicos y la integración del almacén y las base de datos del Sistema de Colaboración Médica.

3. PROCESOS DE ANÁLISIS DE LA INFORMACIÓN CONTENIDA EN EL DATAMART

3.1. JERARQUÍA DE LAS TABLAS DE DIMENSIONES

Determinar la jerarquía de cada una de las tablas de dimensiones resulta de gran importancia para organizar los niveles dentro de una dimensión ya sea de un nivel general a uno específico o viceversa.

A continuación se muestran (Figura 23) las jerarquías de los atributos de algunas de las dimensiones que componen al datamart:

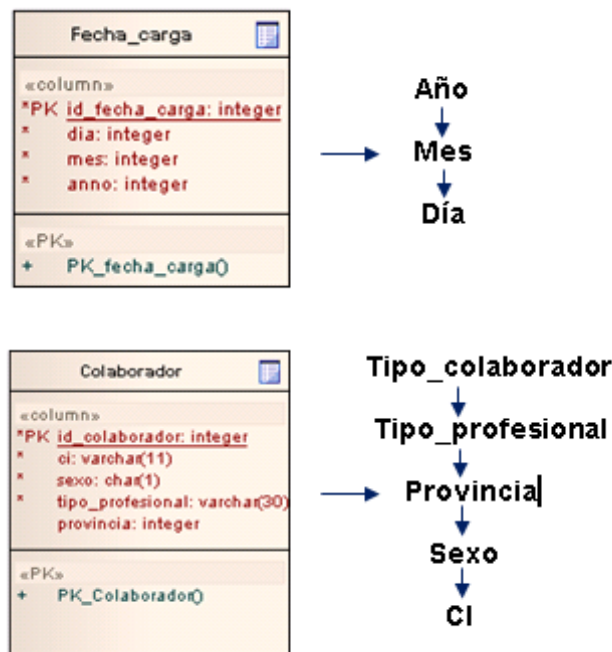


Figura 23. Jerarquía de los atributos de algunas de las dimensiones.

3.2. MEDIDAS

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 de la UCCM. Para ello se utilizan las consultas MDX (*multidimensional expressions* o expresiones multidimensionales) que se ejecutan sobre el datamart para obtener los indicadores que contribuirán a la toma de decisiones.

Desde la base de datos db_datamart se cargan las dimensiones y la tabla de hechos para calcular las medidas. Luego, utilizando la herramienta Workbench se diseña el cubo de datos o Cubo_datamart (Figura 24) agregándole las medidas y formulando las consultas para obtener el total de colaboradores, misiones, importe y desertores además del porcentaje de misiones en dependencia de las necesidades de los directivos de la UCCM.

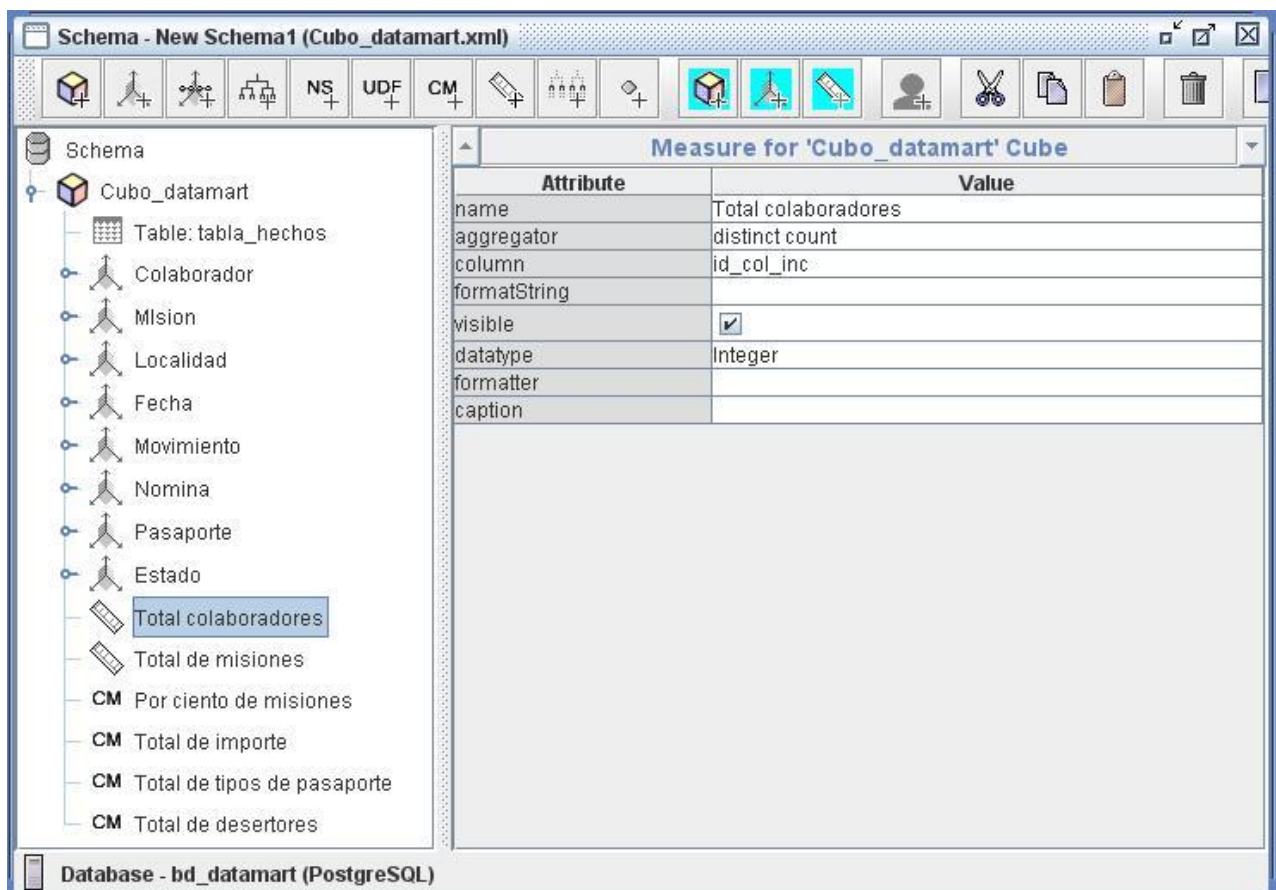


Figura 24. Cubo multidimensional que contiene las dimensiones, jerarquías de los atributos y medidas.

3.3. CONSULTAS

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 (Figura 25). Dichas consultas calcularán los datos numéricos que serán visualizados en los resultados.

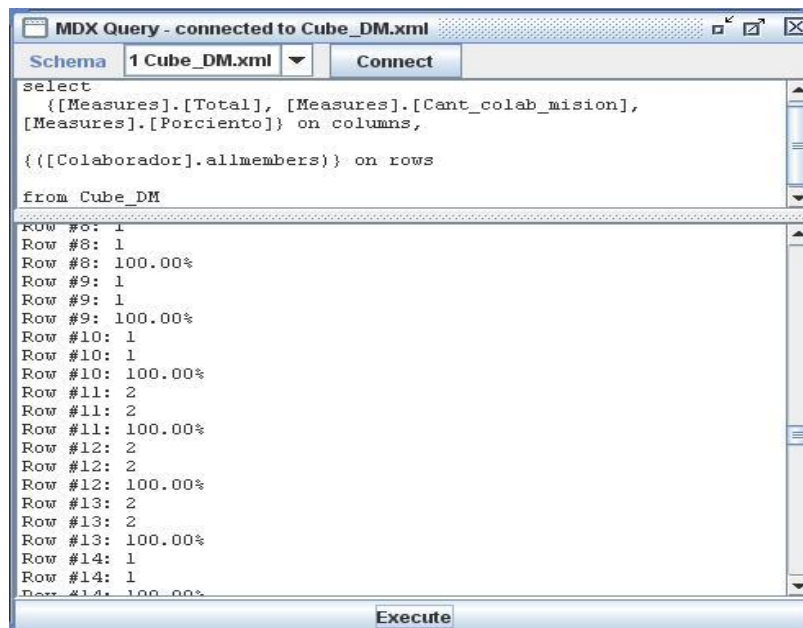


Figura 25. Consultas MDX.

3.4. RESULTADOS DE LOS CLIENTES DE MONDRIAN

3.4.1. JPivot

El cubo es almacenado en un archivo XML para que luego sea cargado en la herramienta Mondrian, herramienta escogida por su alto desempeño para realizar el procesamiento analítico en línea.

El cliente Jpivot estará disponible a través de la web, donde se muestra un menú con las opciones necesarias para obtener la información desde diferentes perspectivas ejemplo filtrar la información según los campos que deseen los directivos (Figura 26) ajustar la visión de los datos, el filtro de visualización y añadir o quitar campos .Los resultados serán mostrados mediante gráficos (Figura 27 y 28) y tablas (Figura 29) y exportados en formato PDF o EXCEL.

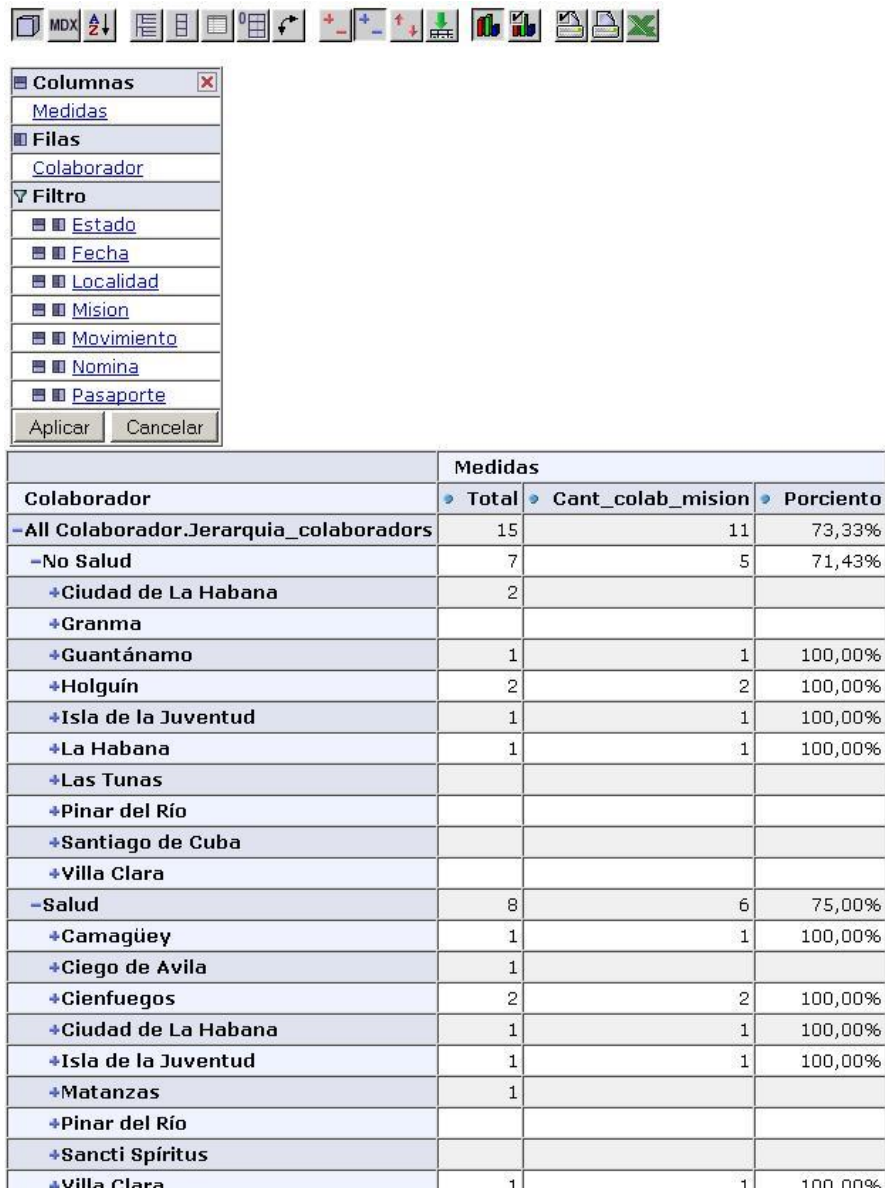


Figura 26. Dimensión colaborador con las medidas e indicadores según el filtrado.

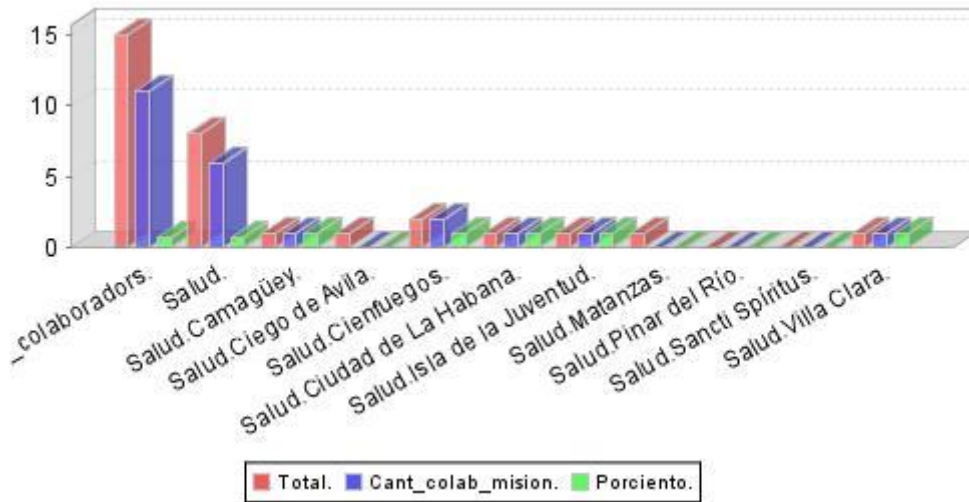


Figura 27. Información gráfica.

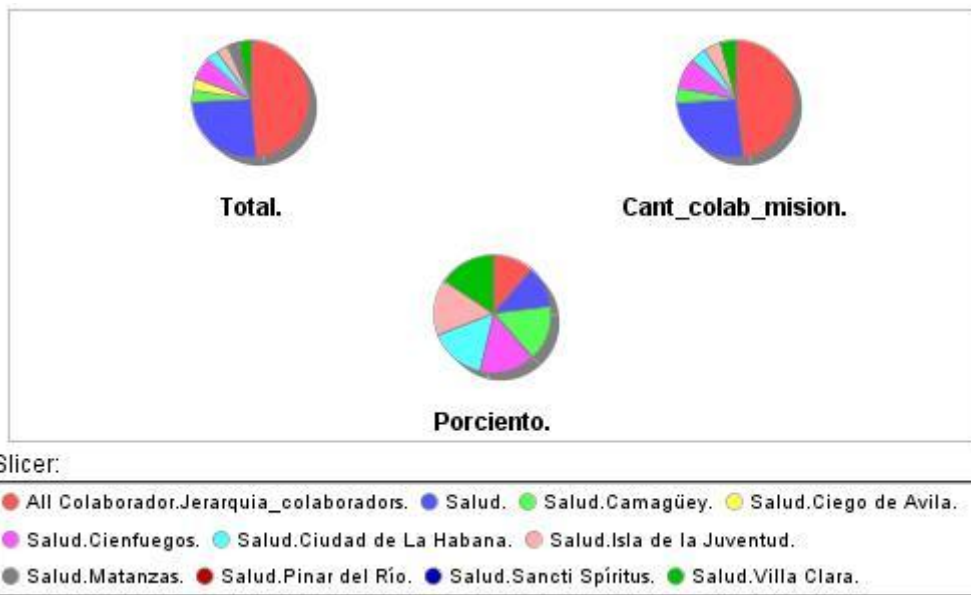


Figura 27. Otra vista de información gráfica.

		Medidas
Colaborador	Estado	Total
- Colaborador.Jerarquia	-Estados	42
	activo	11
	pasivo	4
-No Salud	-Estados	7
	activo	5
	pasivo	2
- Ciudad de La Habana	-Estados	11
	activo	8
	pasivo	3
Holguín	+Estados	2
Isla de la Juventud	+Estados	1
La Habana	+Estados	1

Figura 29. Información tabulada.

3.5. DESPLIEGUE DEL DATAMART

Para realizar el despliegue del datamart para la UCCM se necesitarán un total de 3 servidores, uno para la base de datos de Colaboración Médica implementado sobre el gestor de base de datos MySQL 5.0, otro para el datamart sobre el gestor de base de datos PostgreSQL 8.3 y el servidor de aplicación Web donde se ejecuta el Pentaho Mondrian sobre el JBoss.

Los servidores se comunicarán mediante el protocolo TCP/IP y la PC cliente se conectará mediante el protocolo HTTP para realizar las consultas al datamart a través de la herramienta Mondrian (Figura 30).

Los servidores de base de datos deben de contar con un mínimo de 150 Giga Bytes de capacidad de almacenamiento en disco y 1 Giga Bytes de memoria RAM, para el servidor de aplicación se recomienda que tenga 2 GB de memoria RAM y las PC cliente debe contar con 60 Giga Bytes de capacidad de almacenamiento en disco y 256 Mega Bytes de memoria RAM.

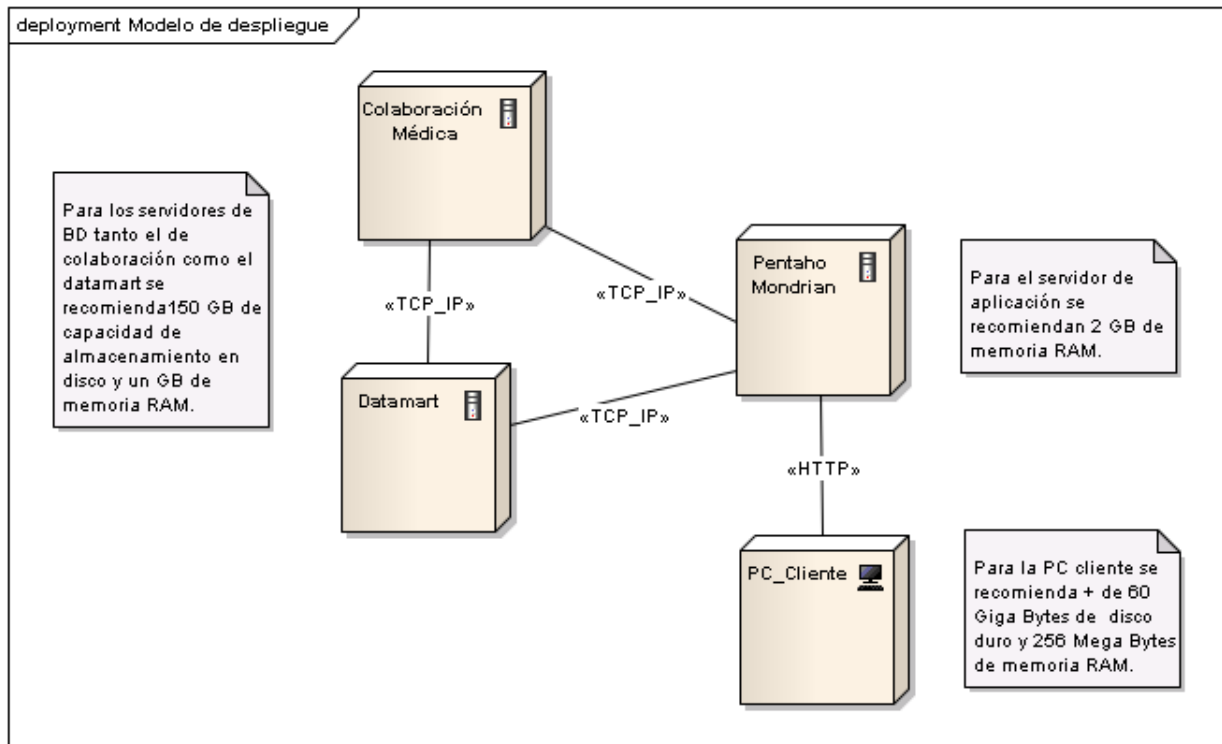


Figura 30. Modelo de despliegue del datamart.

Conclusiones del capítulo:

En el capítulo 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 expuso la forma en que debe realizarse el despliegue de la aplicación en la UCCM.

CONCLUSIONES

A partir de los resultados obtenidos se arribó a la siguiente conclusión:

Con el diseño e implementación del datamart para la UCCM se permite almacenar grandes volúmenes de información relevante solucionando los problemas más comunes que afronta la organización actual: el procesamiento de gran cúmulo de datos, así como la extracción de la información de colaboración médica desde múltiples fuentes para transformarla en el formato deseado, lo que agiliza las consultas requeridas por los directivos de la UCCM para la toma de decisiones.

BENEFICIOS

Con el datamart los directivos de la UCCM dispondrán de una herramienta que les permitirá:

- Tener el control de la cantidad de los colaboradores cumpliendo o que han cumplido misión en cualquier parte del mundo.
- Obtener una representación de los resultados en forma de tabla, gráfica, PDF y Excel.
- Enviar información detallada y consolidada a los niveles superiores, MINSAP, MINREX y al consejo de estado de forma rápida.
- Almacenamiento y análisis de la información histórica de la UCCM.

RECOMENDACIONES

Con el objetivo de mejorar el uso del datamart para la UCCM se recomienda:

- Aplicar técnicas de minería de datos sobre el datamart que les permita a los directivos de la UCCM: analizar factores de influencia en determinados procesos, predecir, estimar variables o comportamientos futuros de los colaboradores y misiones, además de obtener secuencias de eventos que provoquen comportamientos específicos que le ayuden a una mejor toma de decisiones.

REFERENCIAS BIBLIOGRÁFICAS.

1. Dr. Delgado Ramos, Ariel. Informatización en el Sistema Nacional de Salud de Cuba.

[En línea] 10-2009. <http://informatica2009.sld.cu/conferencias/informatizacion-en-el-sistema-nacional-de-salud-de-cuba>.

2. Ing. Bernabeu, Ricardo Darío. Data Warehousing: Investigación y sistematización de conceptos- Hefesto: Metodología propia para la construcción de un Data Warehouse. Argentina: s.n., 2007.

3. 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/>

4. Ídem [3] [En línea]

5. Ídem [3] [En línea]

6. Mg. Oporto Díaz, Samuel. Metodologías para el Data Warehousing.

[Online] 10-2009.

7. Idem [2]

8. Idem [2]

9. Ing. Bernabeu, Ricardo Darío. ARQUITECTURA DEL DATA WAREHOUSING.

[En línea] 07-2009 <http://www.dataprix.com/es/arquitectura-del-data-warehouse>.

10. Ídem [9]. [En línea]

11. ETL para Analytics.

[En línea] <http://es.talend.com/solutions-data-integration/etl-for-analytics.php>

12. Persistencia Molap, Rolap, Holap.

[En línea]. 09-2009 http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx .

13. 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>.

14. Ídem [13] [En línea]

15. Ídem [13] [En línea]

16. Ibarra, María de los Ángeles. IBM DB2 OLAP SERVER.

[Online] 10-2009. <http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/OLAP.pdf>

17. Hernando Velasco, Roberto. El SGBDR Oracle.

[Online] 01-2010. <http://www.rhernando.net/modules/tutorials/doc/bd/oracle.html>

18. Pecos, Daniel PostgreSQL vs. MySQL.

[En línea] http://www.netpecos.org/docs/mysql_postgres/x15.html.

19. Ídem. [18] [En línea]

20. González Morales. 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 Puebla. A. C. 2007.

[En línea] http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/gonzalez_m_ac/capitulo_3.html.

21. Ídem [20] [En línea]

22. Ídem [20] [En línea]

23. Ídem [20] [En línea]

24. Ídem [20] [En línea]

25. Ídem [20] [En línea]

26. Ídem [20] [En línea]

27. Ídem [20] [En línea]

28. Moreno Martínez, Gerardo. Ingeniería de SoftwareUML.

[En línea] 01-2010 <http://www.monografias.com/trabajos5/insof/insof.shtml>

29. [En línea] <http://www.visual-paradigm.com>.

30. Enterprise Architect- Herramienta de diseño UML.

[En línea] 01-2010 <http://www.sparxsystems.com.ar/products/ea.html>.

31. Nuestra Tecnología. [En línea] 12-2009. <http://www.naturasoftware.com/main.php?f=tecnologia>.

32. Tomcat: Introducción al Servidor de Aplicaciones.

[En línea] 01-2010 <http://www.consultec.es/formacion/temarios/TEMCON-307.pdf>.

33. Gutiérrez Kafati, Elizabeth. La plataforma Pentaho Open Source Business Intelligence.

[En línea] 01-2010.

http://egkafati.bligoo.com/content/view/219538/La_plataforma_Pentaho_Open_Source_Business_Intelligence.html.

34. Ídem [33] [En línea]

35. Ídem [33] [En línea]

36. Ídem [2]

37. Ídem [2]

BIBLIOGRAFÍA

1. 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. Casares, Claudio. DataWarehousing. [En línea] 10-2009. <http://www.programacion.com/bbdd/tutorial/warehouse/>
3. 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/>
4. Delgado Ramos, Dr. Ariel. Informatización en el Sistema Nacional de Salud de Cuba. [En línea] 10-2009. <http://informatica2009.sld.cu/conferencias/informatizacion-en-el-sistema-nacional-de-salud-de-cuba>.
5. Enterprise Architect- Herramienta de diseño UML. [En línea] 01-2010 <http://www.sparxsystems.com.ar/products/ea.html>.
6. ETL para Analytics.[En línea] <http://es.talend.com/solutions-data-integration/etl-for-analytics.php>
7. González Morales. 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 Puebla. A. C. 2007. [En línea] http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/gonzalez_m_ac/capitulo_3.html.
8. Gutiérrez Kafati, Elizabeth. La plataforma Pentaho Open Source Business Intelligence. [En línea] 01-2010. http://egkafati.bligoo.com/content/view/219538/La_plataforma_Pentaho_Open_Source_Business_Intelligence.html
9. Hernando Velasco, Roberto. El SGBDR Oracle.

[En línea] 01-2010. <http://www.rhernando.net/modules/tutorials/doc/bd/oracle.html>

10. Ibarra, María de los Ángeles. IBM DB2 OLAP SERVER.

[En línea] 10-2009. <http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/OLAP.pdf>

11. Moreno Martínez, Gerardo. Ingeniería de SoftwareUML.

[En línea] 01-2010 <http://www.monografias.com/trabajos5/insof/insof.shtml>

12. Nuestra Tecnología. [En línea] 12-2009. <http://www.naturasoftware.com/main.php?f=tecnologia>.

13. Oporto Díaz, Mg. Samuel. Metodologías para el Data WareHousing.

[En línea] 10-2009.

14. Pecos, Daniel PostgreSQL vs. MySQL.

[En línea] http://www.netpecos.org/docs/mysql_postgres/x15.html.

15. Persistencia Molap, Rolap, Holap.

[En línea]. 09-2009 http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx .

16. Ramón Cueto, Ariagna y Díaz García, Joannis. Implementación de un Data Warehouse para el control del Recurso Humano de la Salud. Universidad de las Ciencias Informáticas. Ciudad de La Habana, 2009.

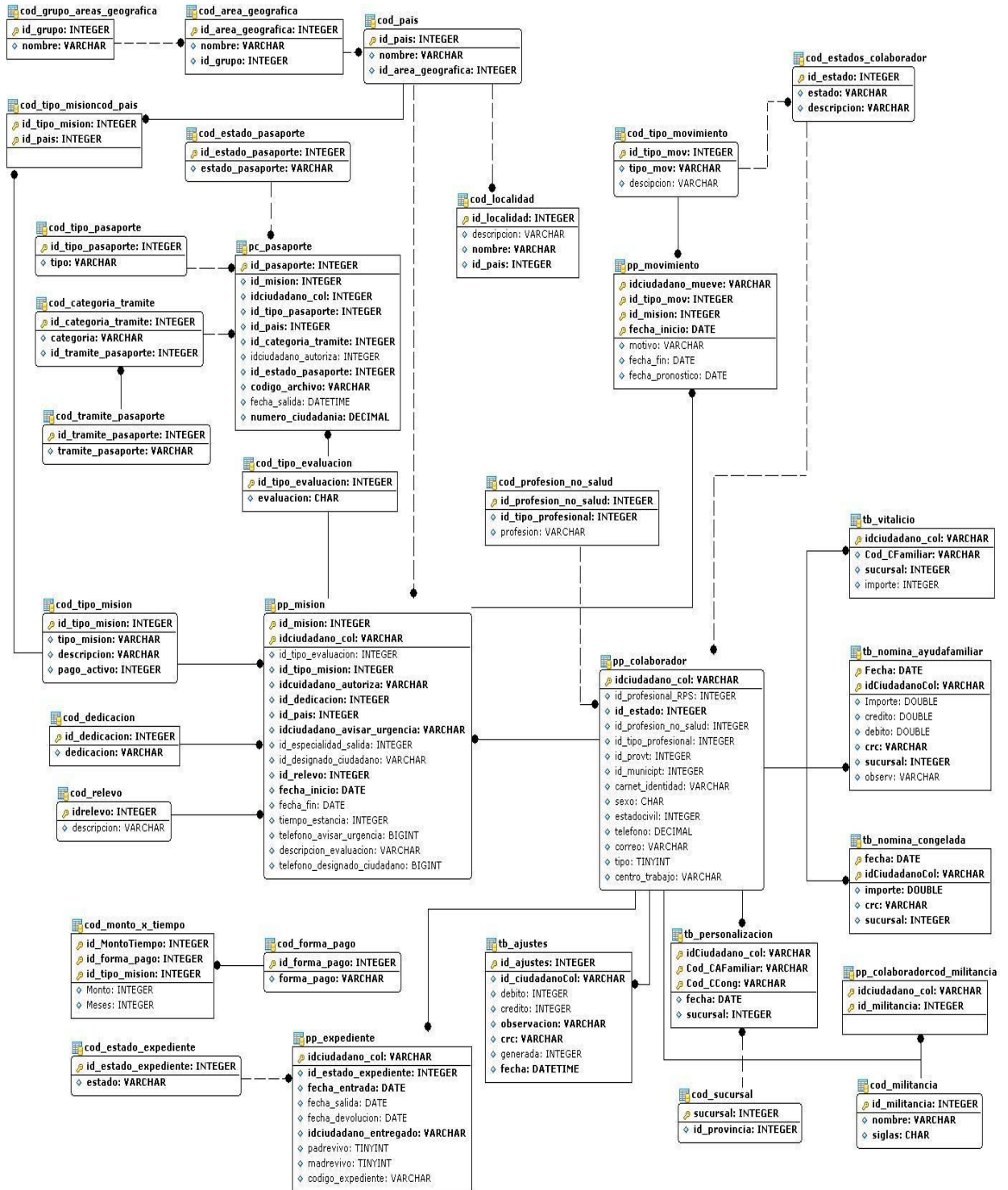
17. La revista de tecnología y estrategia de negocio en Internet. La empresa multidimensional: OLAP. 11/1/2002. Número: 54 [En línea] <http://www.idg.es/iWorld/articulo.asp?id=143456>.

18. Tomcat: Introducción al Servidor de Aplicaciones.[En línea] 01-2010 <http://www.consultec.es/formacion/temarios/TEMCON-307.pdf>.

19. [En línea] <http://www.visual-paradigm.com>.

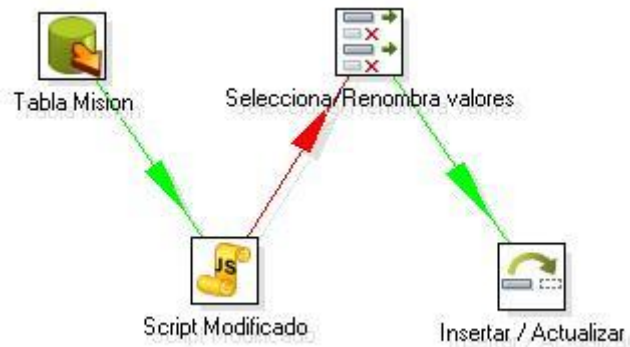
ANEXOS.

1. BASE DE DATOS DE COLABORACIÓN MÉDICA.



2. CARGA DE LAS DIMENSIONES DEL DATAMART.

2.1. DIMENSIÓN MISIÓN.



2.2. DIMENSIÓN LOCALIZACIÓN.



2.3. DIMENSIÓN NÓMINA.



2.4. DIMENSIÓN PASAPORTE.



2.5. DIMENSIÓN MOVIMIENTO.

