UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

FACULTAD 6

TÍTULO: ANÁLISIS DE UN ALMACÉN DE DATOS PARA LA RED NACIONAL DE GENÉTICA MÉDICA.

TRABAJO DE DIPLOMA PARA OPTAR POR EL TÍTULO DE INGENIERO EN CIENCIAS INFORMÁTICAS.

AUTORES:

Imias Fernández Álvarez

Yulienni Hernández Armas

TUTORES:

Ing. Haymee Llerena Esperón

Ing. Yadira Robles Aranda

Ing. Alfonso Claro Arceo

Lic. Yanelis Benítez Fernández

CIUDAD DE LA HABANA, CUBA

JUNIO 2010

LO FUNDAMENTAL ES QUE SEAMOS CAPACES DE HACER CADA DÍA ALGO QUE PERFECCIONE LO QUE HICIMOS EL DÍA ANTERIOR.

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 _____.

Autores:

Yulienni Hernández Armas — _____ — Firma del autor

Imias Fernández Álvarez — _____ — Firma del autor

Tutores:

Ing. Yadira Robles Aranda — _____ — Firma del tutor

Ing. Haymee Llerena Esperón — _____ — Firma del tutor

Ing. Alfonso Claro Arceo — _____ — Firma del tutor

Lic. Yanelis Benítez Fernández — _____ — Firma del tutor

DATOS DE CONTACTO

Tutores:

Ing. Yadira Robles Aranda
Correo electrónico: yrobles@uci.cu
Universidad de las Ciencias Informáticas, Ciudad de La Habana, Cuba.

Ing. Haymee Llerena Esperón
Correo electrónico: hllerena@uci.cu
Universidad de las Ciencias Informáticas, Ciudad de La Habana, Cuba.

Ing. Alfonso Claro Arceo
Correo electrónico: aclaro@uci.cu
Universidad de las Ciencias Informáticas, Ciudad de La Habana, Cuba.

AGRADECIMIENTOS

De Imias Fernández Álvarez:

Quiero expresar un profundo agradecimiento a quienes con su ayuda, apoyo y comprensión me alentaron a lograr esta hermosa realidad.

Primeramente a mis padres, Gisela Álvarez Moreira y Raúl Fernández del Río, muchas gracias por su amor, su preocupación, y confianza y por ser la luz que me ha guiado durante todos estos años. Gracias por haberme educado con esos principios sin los cuales hoy no podría ser lo que soy. Los quiero mucho.

A mis hermanitos queridos que siempre han estado a mi lado Abelito y Gustavo, los adoro, gracias por ser como son.

A mi hermana Ivette que siempre ha estado a mi lado, sin importar lo difícil que haya sido el momento, y si hoy estoy aquí en gran medida te lo debo a ti .Gracias tatica por apoyarme en todo, por los sacrificios hechos, por los consejos dados, por el gran ejemplo que has sido para mí y por compartir mi alegría.

A mis sobrinitos lindos, los quiero mucho.

A mi abuelita Tito, mis tíos, mis primos, a toda mi familia.

A mi novio Luis Karel, por ser mi apoyo estos últimos dos años, por estar ahí cada vez que lo necesito, muchas gracias mi titi.

A mi madrina Carmen, por quererme tanto.

Agradecer a una persona muy especial a Yanelis Benítez Fernández, ha sido mi apoyo durante estos 5 años, gracias por confiar siempre en mí, por ayudarme en todo momento, por ser mi segunda mamá, muchas gracias por todo, este trabajo también es para ti.

Agradecer en especial a mis amigas de toda la vida, Ana Teresa y Laura, simplemente gracias por ser mis amigas, por preocuparse por mí, por mis pruebas, por ayudarme cada vez que lo necesité aun estando

lejos. A José Alberto también muchas gracias por cada repaso dado, por ayudarme siempre. A Caco, Tere, Lourdes, Las cucas, muchas gracias por todo.

A mis otras amigas que no son menos importantes, a Yulita mi compañera de tesis y hermana, que ha estado siempre en las buenas y malas, que a pesar de las discusiones y las diferencias de carácter, siempre hemos estado muy unidas. A Gladys y Dayana que fueron las primeras personas que conocí en esta universidad y siempre estuvieron a mi lado. A Madays, Isis, Oscar, Alex, Keyler, Danoy, Piki, Alicia, Yaneisy, Sulay, Ana María, Mayelín, Jesús, Edilberto, muchas gracias por ser mis amigos, por los momentos vividos y por estar ahí cuando los necesité. Nunca los voy a olvidar!!!

A mis amigas que vienen conmigo del pueblo a Yisel, Anamaris, Alianny, gracias por los buenos ratos, por las noches de helados, gracias por todo.

A las nuevas amistades, la Yadi, el Payo, Ale, Alain, Julito, Yeya, Lester, Oscar, Mata, y los demás, gracias a todos los que me regalaron dulcecitos, torticas y galleticas.

Agradecerles a mis tutores por su apoyo dado, en especial a Alfonso por a última hora correr tanto con nosotras y a Yanelis.

A Leonardo y Daylén, por sacarnos del apuro y ayudarnos tanto.

A la Revolución y en especial a nuestro comandante en jefe Fidel Castro por brindarnos la oportunidad de ser mejores y por confiar su futuro en nosotros.

AGRADECIMIENTOS

De Yulienni Hernández Armas:

En especial quiero agradecer a dos personitas que han sido todo para mí a lo largo de mi vida, mis papitos lindos, por su apoyo incondicional, por haberse sacrificado tanto por mí. Creo que si no hubiese sido por ellos nunca me hubiera graduado.

A mi mamita porque todo lo que he logrado ha sido gracias a ella, que me crió con todo el cariño y el amor del mundo. Por darme siempre su apoyo y seguridad. Por saber guiarme en la vida y ayudarme a tomar esas decisiones tan importantes que muchas veces le han dolido más a ella que a mí. Gracias por tus consejos, tu dedicación, tu paciencia y tu gran amor. Te adoro mi mamiti.

A mi papito lindo por ser mi ángel de la guarda, por darme su apoyo, su amor incondicional y por ser el hombre más dichoso de este mundo al tener una hija que lo adora, pero sobre todas las cosas por no juzgarme nunca, por siempre estar ahí para mí y escucharme cada vez que necesité hablar. Te quiero mucho mi papiti.

A mi hermanita linda que es el regalo más bello que me han dado mis papitos. No importa cuánto discutamos de vez en vez, no sabes como agradezco a la vida y lo orgullosa que estoy de tenerte como hermana. Te quiero un montón tatica.

A tía Panchita y tío Matía por ser mis papitos en estos 5 años, por quererme y confiar siempre en mí, por ser los otros grandes tesoros que tengo en mi vida. Sin ustedes yo creo que jamás lo hubiera logrado.

A papi (abuelo Evelio), mami (abuela Teresa), abuelo (Diego) y abuela (Beba) porque este amor que siento por ellos es tan grande que me da fuerzas para todo.

A mis tíos Dieguito, Teresita, Elisa, Isa y Luciano por tenerme siempre presente y estar ahí en los buenos y malos momentos.

A mis primitos lindos, Miry, Liusbelito, Rudito, Ludmi y Yali que siempre tendrán un pedacito de mi corazón.

Agradecimientos

A mi familia toda que ha estado siempre pendiente y apoyándome ante cada éxito o fracaso que haya tenido en la vida.

A mis amigos del alma, a la Mada, por estar siempre a mi lado desde que llegué aquí, por haber sonreído y secado mis lágrimas tantas veces, por preocuparse por mí y dedicarme su cariño. Se que siempre podré contar contigo.

A la Imi, mi linda compañerita de tesis, tengo que decir que nunca hubiera querido otra, le doy las gracias por su paciencia, por compartir conmigo todos esos momentos de locura y alegría, por hacerme reír tanto. Por lograr que después de todas esas discusiones bobas que tanto hacían reír a Oscar termináramos tan junticas como siempre, por ser mí segunda hermanita.

A Isis porque aunque llegó tarde a mi vida me ha ayudado mucho. Le agradezco por los buenos momentos que compartimos, por su compañía, por preocuparse por mí pero sobre todas las cosas le doy las gracias por regalarme su amistad, por quererme así de gratis y aguantarme siempre.

A Osqui por ser ese gran amigo que siempre quise tener, por ayudarme y hacer que esta cabecita bruta aprendiera todas las P. Por estar siempre que lo necesito.

A todos mis amigos de la UCI, a Gladys, Ana, Alicia, Sula, Yane, Ale, Dayana, Dano, Raiki, Keyler y a Edi.

A la UCI, porque aquí viví gran parte de los momentos más lindos de mi vida.

A mis tutores y a todos los profesores que me ayudaron a lograr este gran sueño. En especial a Yanelis y Alfonso por su paciencia y dedicación. Desde el fondo de mi corazón, muchas gracias.

DEDICATORIA

De Imias Fernández Álvarez:

A mis padres por darme las fuerzas necesarias para luchar y así poder realizar mis sueños, por todo el sacrificio y por lo que han puesto de su parte para que todo esto sea posible, por su gran apoyo incondicional. Espero que se sientan orgullosos de mi, gracias, este trabajo es de ustedes. Son los mejores padres del mundo y mi razón de ser. Los quiero con la vida!!!!!

A mis hermanos y sobrinos, simplemente por existir.

A mi abuelita Tito, por pensar siempre en mi.

A toda mi familia, por su apoyo.

A mi novio, por ayudarme en todo momento.

A mis amigos.

De Yulienni Hernández Armas:

A mis papitos, por nunca haber dudado de mí, por respaldarme y amarme siempre; por eso y por la inmensa admiración que siento por ellos, es que les dedico no solamente mi tesis sino también todo mi cariño y amor.

A mi hermanita linda, te quiero mi niñita.

RESUMEN

El presente trabajo de diploma propone el análisis de un almacén de datos para la red nacional de Genética Médica en Cuba, dado por la necesidad de almacenar grandes volúmenes de información que se aglomeran diariamente. Este constituye el primer paso para la confección del mismo y sienta las bases para un eficiente diseño e implementación.

Para la realización del análisis se utilizan las dos primeras fases de la metodología Hefesto, las que permiten, siguiendo sus pasos, conocer la cantidad y el porcentaje de diferentes tipos de personas, a través de varios criterios, obteniendo como resultado final un modelo conceptual ampliado que permitirá visualizar los resultados obtenidos.

PALABRAS CLAVES:

Almacén de datos

Información

Metodología

Conocimiento

Análisis

ÍNDICE

INTRODUCCIÓN

La genética es la base para el entendimiento de la constitución biológica fundamental del organismo. Esta ciencia, mundialmente ha ido avanzando a pasos agigantados, adquiriendo nuevos y vastos espacios en la investigación, lo cual permite brindar novedosas herramientas en esta rama.

Cuba, en los últimos 51 años no ha estado exenta, desarrollando un grupo de programas sociales para mejorar la calidad de vida del pueblo. El 5 de agosto de 2003, como parte de la Batalla de Ideas, se creó el Centro Nacional de Genética Médica (CNGM), institución científica dedicada a la asistencia, la docencia, la investigación y la promoción de servicios de salud.

El CNGM es el centro rector de la red nacional de Genética Médica constituida por 184 centros, ubicados en las 14 provincias del país en los cuales se desempeñan 453 másteres en asesoramiento genético que facilitan el mejoramiento de los servicios a la comunidad [1].

Su objetivo fundamental es disminuir las enfermedades con implicación genética, desarrollando numerosos estudios que permiten elevar el bienestar del pueblo cubano, además de desarrollar recursos técnicos para la educación y formación de una cultura genética en la población.

El CNGM lleva a cabo disímiles registros, creados con el objetivo de almacenar la información necesaria que permita llevar a cabo estudios en este campo. Para lograr el proceso de gestión de información referente a los registros, se desarrolló un producto en la Universidad de las Ciencias Informáticas (UCI), bajo el nombre alasMEDIGEN el cual consta de los 9 módulos siguientes:

1. Módulo Registro Cubano de Historias Clínicas (RECUHCL).
2. Módulo Registro Cubano de Gemelos (RECUGEM).
3. Módulo Registro Cubano de Discapacitados (RECUDIS).
4. Módulo Registro Cubano de Retrasos Mentales (RECURM).
5. Módulo Registro Cubano de Malformaciones Congénitas (RECUMAC).
6. Módulo Registro Cubano de Enfermedades Genéticas (RECUEGEN).
7. Módulo Teleconsulta Genética.
8. Módulo Registro Cubano de Anomalías Cromosómicas (RECUAC).

9. Módulo Registro Cubano de Enfermedades Comunes (RECUEC).

La red nacional de Genética Médica gestiona una amplia gama de datos asociados a los estudios que realizan, los genetistas para acceder a esta información cuentan con una aplicación informática conformada por una base de datos con 343 tablas. Con el transcurso del tiempo, los datos genéticos aumentarán, lo cual traerá por consecuencia un incremento sustancial de la información en las tablas, por lo que el acceso a la aplicación será más difícil y el tiempo de respuesta será mayor. Este cúmulo de información hace engorroso el trabajo y dificulta la posibilidad de obtener un resultado que sea punto de apoyo para la toma de decisiones efectivas. Con esta aplicación no existe un registro de los datos históricos, sólo cuenta con los datos actuales y no con las modificaciones hechas, por lo que si se llegara a necesitar una información que fue registrada antes de la última actualización no se podría acceder a ésta.

Por lo anteriormente explicado, se concluye que la aplicación no presenta herramientas adecuadas para el análisis de datos que posibilite adquirir conocimiento y tomar decisiones rápidamente. Teniendo en cuenta la situación expuesta se ha identificado el siguiente:

Problema Científico

¿Cómo obtener conocimiento, para la toma de decisiones, a partir de los datos que generan los estudios de la red nacional de Genética Médica?

Objeto de Estudio

Procesos que permitan la extracción de conocimiento para la toma de decisiones.

Campo de Acción

Procesos que permitan la extracción de conocimiento de la base de datos de la red nacional de Genética Médica.

Objetivo General

Realizar el análisis de un almacén de datos para la red nacional de Genética Médica.

Objetivos Específicos

- Identificar los requerimientos del almacén de datos para la red nacional de Genética Médica.
- Elaborar el modelo conceptual del almacén de datos.

- Analizar los On-Line Transaction Processing (OLTP) del almacén de datos para la red nacional de Genética Médica.
- Validar la solución propuesta.

Tareas a Cumplir

- Identificación de las preguntas relacionadas con los procesos identificados.
- Identificación de los indicadores y perspectivas de análisis.
- Realización del modelo conceptual.
- Establecimiento de correspondencia entre los requerimientos y los OLTP.
- Validación de la solución propuesta.

El documento de tesis está estructurado, por la introducción, tres capítulos, conclusiones, recomendaciones, bibliografía y anexos.

En el **Capítulo 1: Fundamentación teórica.** Se realiza la fundamentación teórica donde se aborda y profundiza en el estudio de las herramientas de Inteligencia de Negocios que se pueden utilizar para poder obtener conocimiento a partir de los datos que generan los estudios de la red nacional de Genética Médica, además se define la metodología a utilizar para la construcción de la herramienta.

En el **Capítulo 2: Análisis de un almacén de datos para la red nacional de Genética Médica.** En este capítulo se mostrarán los resultados y pasos a seguir para el análisis del almacén de datos según los describe la metodología utilizada. Bajo esta guía, se construye un modelo conceptual a partir de los indicadores y perspectivas obtenidos de las preguntas identificadas.

El **Capítulo 3: "Validación de la solución":** En este capítulo se confeccionará una lista de chequeo con el objetivo de eliminar posibles errores existentes en el análisis del almacén de datos de la red nacional de Genética Médica. Mediante esta lista se expondrán criterios de expertos en el tema, validando la solución.

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

En el presente capítulo se exponen conceptos asociados a la Inteligencia de Negocios. Se estudian las herramientas, técnicas y tecnologías de Inteligencia de Negocios que se pueden utilizar para poder obtener conocimiento a partir de los datos que generan los estudios de la red nacional de Genética Médica. Además, se realiza un estudio de la metodología que guiará el análisis de la herramienta a utilizar.

1.1. INTELIGENCIA DE NEGOCIOS

La Inteligencia de Negocios o Business Intelligence (BI, por sus siglas en inglés) se puede definir como: el conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización [2]. Este conjunto de herramientas y estrategias tienen en común las siguientes características:

- **Apoyo en la toma de decisiones:** Se busca ir más allá en la presentación de la información, de manera que los usuarios tengan acceso a herramientas de análisis que les permitan seleccionar y manipular sólo aquellos datos que les interesen.
- **Convertir los datos en información:** Para este proceso de conversión de la información se acude a la llamada pirámide informacional que muestra el proceso de conversión de la información (Fig. 1). Para la mejor comprensión de este proceso se definen 4 conceptos:

  Datos: Los datos son la mínima unidad semántica, y se corresponden con elementos primarios de información que por sí solos son irrelevantes, como apoyo a la toma de decisiones. También se pueden ver como un conjunto discreto de valores, que no dicen nada sobre el porqué de las cosas y no son orientativos para la acción.

  Información: La información se obtiene al unir y estructurar los datos, de manera que la composición de estos tome una nueva dimensión y sea de utilidad para quien deba tomar decisiones.

  Conocimiento: Es una mezcla de información que sirve como marco para la incorporación de nuevas experiencias y es útil para la acción.

Inteligencia: La Inteligencia, como actividad, es el conocimiento anticipado logrado a través del procesamiento de las informaciones. La difusión de la Inteligencia debe ser oportuna para contribuir a la toma de decisiones y así poder alcanzar objetivos de seguridad y bienestar.

Fig. 1 Pirámide informacional.

- **Orientación al usuario final:** Se busca independencia entre los conocimientos técnicos de los usuarios y su capacidad para utilizar estas herramientas.

Teniendo en cuenta las bibliografías consultadas es importante señalar que la Inteligencia de Negocios, además de en un conjunto de herramientas, se apoya en un grupo de técnicas y tecnologías que almacenan y procesan grandes cantidades de datos e información para transformarla en conocimiento, permitiendo así una eficaz toma de decisiones.

1.1.1. PROCESO DE INTELIGENCIA DE NEGOCIOS

A fin de comprender cómo es que una organización puede crear inteligencia de sus datos, para proveer a los usuarios finales oportuna y acertadamente acceso a esta información, se describirá a continuación el proceso de Inteligencia de Negocios. El mismo está dividido en cinco fases, las cuales serán explicadas teniendo como referencia el siguiente gráfico, que sintetiza todo el proceso (Fig. 2).

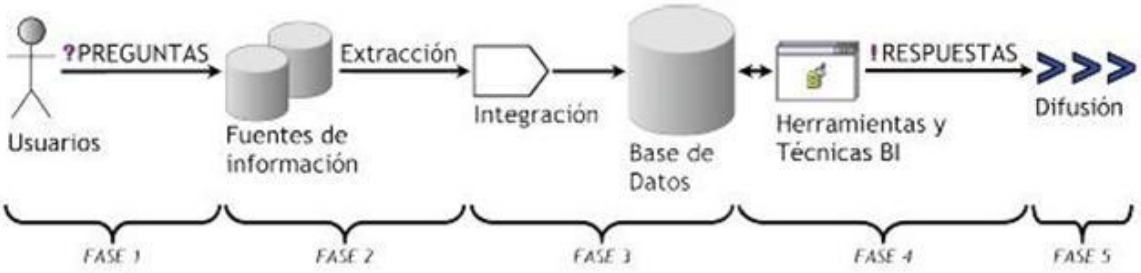

Fig. 2 Fases del proceso de Inteligencia de Negocios

- **Fase 1: Dirigir y Planear.** En esta fase inicial es donde se deberán recolectar los requerimientos de información específicos de los diferentes usuarios, así como entender sus diversas necesidades, para que luego en conjunto con ellos se generen efectivamente las preguntas que les ayudarán a alcanzar sus objetivos.
- **Fase 2: Recolección de Información**. Es aquí en donde se realiza el proceso de extraer desde las diferentes fuentes de información de la empresa, tanto internas como externas, los datos que serán necesarios para encontrar las respuestas a las preguntas planteadas en el paso anterior.
- **Fase 3: Procesamiento de Datos.** En esta fase es donde se integran y cargan los datos en un formato utilizable para el análisis. Esta actividad puede realizarse mediante la creación de una nueva base de datos.
- **Fase 4: Análisis y Producción.** Ahora, se procederá a trabajar sobre los datos extraídos e integrados, utilizando herramientas, técnicas y tecnologías de Inteligencia de Negocios, para crear conocimiento. Como resultado final de esta fase se obtendrán las respuestas a las preguntas, mediante la creación de reportes, indicadores, entre otros.
- **Fase 5: Difusión.** Finalmente, se les entregará a los usuarios que lo requieran las herramientas necesarias, que les permitirán explorar los datos de manera veloz y sencilla [3].

1.1.2. TÉCNICAS, TECNOLOGÍAS Y HERRAMIENTAS DE UN SISTEMA DE INTELIGENCIA DE NEGOCIOS

La Inteligencia de Negocios se apoya en un conjunto de técnicas, tecnologías y herramientas que facilitan la extracción, el análisis y el almacenamiento de los datos generados en una organización, permitiendo generar conocimiento y apoyar la toma de decisiones. A continuación se muestran las características fundamentales de un conjunto de ellas.

- **Sistemas de Soporte a la Decisión (DSS)**

Los Sistemas de Soporte a las Decisiones (DSS) conforman una herramienta enfocada al proceso de análisis de los datos de una organización.

Este proceso no es sencillo, debido a que los DSS suelen disponer de una serie de informes predefinidos en los que presentan la información de manera estática, pero no permiten profundizar en los datos, navegar entre ellos ni manejarlos desde distintas perspectivas.

DSS es una de las herramientas más emblemáticas de la Inteligencia de Negocios ya que, entre otras características, permiten resolver gran parte de las limitaciones de los programas de gestión con una interfaz gráfica.

El principal objetivo de los DSS es explotar al máximo la información residente en un almacén de datos, mostrando informes muy dinámicos y con gran potencial de navegación, pero siempre de forma amigable, vistosa y sencilla [4].

Para una mejor comprensión de esta herramienta, se exponen las características siguientes:

- **Informes dinámicos, flexibles e interactivos:** El usuario no tiene que regirse por los listados predefinidos que se configuraron en el momento de la implantación, y que no siempre responden a sus dudas reales.
- **No requiere conocimientos técnicos:** Un usuario no técnico puede crear nuevos gráficos e informes y navegar entre ellos. Por tanto, para examinar la información disponible o crear nuevas métricas no es imprescindible tener conocimientos previos.
- **Rapidez en el tiempo de respuesta:** Los datos suelen ser extraídos de un almacén de datos, pues están optimizados para el análisis de grandes volúmenes de información.
- **Cada usuario dispone de información adecuada a su perfil:** El usuario va a tener acceso a la información que necesita, para que su trabajo sea lo más eficiente posible.
- **Disponibilidad de información histórica:** En estos sistemas se puede comparar datos actuales con información de otros períodos, con el fin de analizar tendencias, fijar la evolución de parámetros de negocio, entre otros.

- **Almacén de datos (datawarehouse, por sus siglas en inglés)**

Un almacén de datos se caracteriza por integrar y depurar información de una o más fuentes de datos distintas para luego procesarlos, permitiendo su análisis con grandes velocidades de respuesta. La creación de esta herramienta representa, en la mayoría de las ocasiones, el primer paso para implantar una solución completa y fiable de Inteligencia de Negocios [5].

Entre sus beneficios principales se encuentran:

- Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.
- Facilita la aplicación de técnicas, estadísticas de análisis y modelado para encontrar relaciones ocultas entre los datos del almacén; obteniendo un valor añadido para el negocio de dicha información.
- Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.

- **Minería de datos**

La minería de datos, es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.

Es una tecnología de soporte para el usuario final, cuyo objetivo es extraer conocimiento útil a partir de la información contenida en las bases de datos de las empresas.

Básicamente, surge para intentar ayudar a comprender el contenido de un almacén de datos. Con este fin, hace uso de prácticas estadísticas y en algunos casos, de algoritmos de búsqueda próximos a la Inteligencia Artificial y a las redes neuronales [6]. Aunque en la minería de datos cada caso concreto puede ser radicalmente distinto del anterior, el proceso común a estos casos se compone de cuatro etapas principales:

- **Determinación de los objetivos:** Trata de la delimitación de los objetivos que el cliente desea bajo la orientación del especialista en minería de datos.

- **Pre-procesamiento de los datos:** Se refiere a la selección, la limpieza, el enriquecimiento, la reducción y la transformación de las bases de datos. Esta etapa consume generalmente alrededor del 70% del tiempo total de un proyecto de minería de datos.
- **Determinación del modelo:** Se comienza realizando un análisis estadístico de los datos, y después se lleva a cabo una visualización gráfica de los mismos para tener una primera aproximación. Según los objetivos planteados y la tarea que debe llevarse a cabo, pueden utilizarse algoritmos desarrollados en diferentes áreas de la Inteligencia Artificial.
- **Análisis de los resultados:** Verifica si los resultados obtenidos son coherentes y los compara con los obtenidos por el análisis estadístico y la visualización gráfica. El cliente determina si le aportan un nuevo conocimiento que le permita considerar sus decisiones.

- **OLAP (del español, Procesamiento analítico en línea)**

Los sistemas OLAP son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil, como son: tendencias de ventas, patrones de comportamiento de los consumidores y elaboración de informes complejos [3].

Características de los sistemas OLAP:

- El acceso a los datos suele ser sólo de lectura. La acción más común es la consulta, con muy pocas inserciones, actualizaciones o eliminaciones.
- Los datos se estructuran según las áreas de negocio, y sus formatos están integrados de manera uniforme en toda la organización.
- El historial de datos es a largo plazo, normalmente, de dos a cinco años.
- Se suelen alimentar de información procedente de los sistemas operacionales existentes, mediante un proceso de extracción, transformación y carga (ETL).

JUSTIFICACIÓN DE LA HERRAMIENTA A UTILIZAR

Después de realizar un análisis de las principales herramientas, técnicas y tecnologías para la realización de un sistema de Inteligencia de Negocios, se determina, que lo primario para la extracción de conocimiento a partir de los datos que generan los estudios de la red nacional de Genética Médica, es un

almacén de datos. Esta elección se debe a la necesidad de obtener una herramienta que facilite la toma de decisiones y que brinde la posibilidad de realizar comparaciones de estudios en diferentes períodos de tiempo.

Un almacén de datos es la base de un sistema de Inteligencia de Negocios, teniendo en cuenta que las demás herramientas, técnicas y tecnologías, basan su funcionamiento, a partir de los datos almacenados en él.

1.2. ALMACÉN DE DATOS

Se puede definir un almacén de datos como la herramienta de Inteligencia de Negocios que posibilita la extracción de datos de sistemas operacionales y fuentes externas, permite la integración y homogeneización de los datos de toda la empresa, provee información que ha sido transformada y resumida, para que ayude en el proceso de toma de decisiones estratégicas y tácticas [3].

A raíz del surgimiento del almacén de datos, dos grandes personalidades de la informática lo han definido desde diferentes puntos de vistas. A continuación se presentan los conceptos más relevantes.

Inmon, considerado el padre del almacén de datos lo define como una colección de datos orientada al sujeto, integrados, variantes en el tiempo y no volátiles que soportan el proceso administrativo de soporte a las decisiones. Su definición es útil porque utiliza atributos que pueden medirse [7].

Ralph Kimball, menciona que el almacén de datos es el lugar en donde la gente puede archivar su información. Lo define como: una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis [8]. También fue él quien determinó que esta herramienta no era más que: la unión de todos los Datamarts de una entidad [8].

Analizando estas definiciones se concluye que un almacén de datos es la combinación de tecnologías y procesos orientados a la obtención de conocimiento y toma de decisiones, donde lo más importante es la historicidad de los datos, que provienen de diversas fuentes.

1.2.1. CARACTERÍSTICAS DE UN ALMACÉN DE DATOS

Las definiciones expuestas anteriormente permiten enunciar las principales características de un almacén de datos. Éstas son:

- **Integrado:** los datos almacenados deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios (Fig. 3).

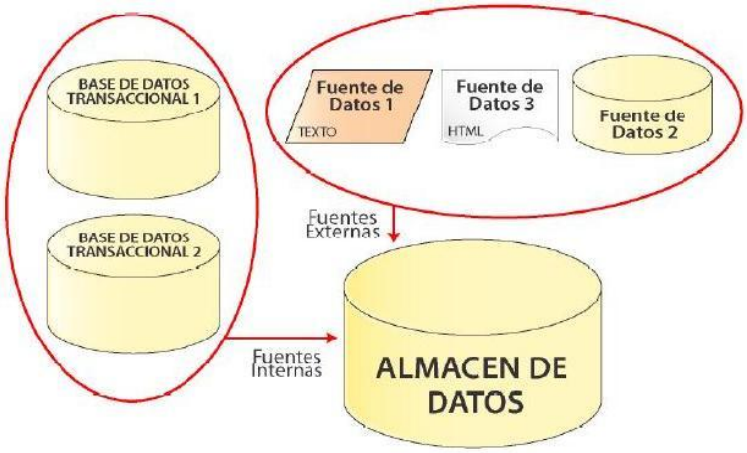

Fig. 3 Datos Integrados

- **Temático:** Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales (Fig. 4). Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla del almacén de datos. De esta forma, las peticiones sobre clientes serán más fáciles de responder, dado que toda la información reside en el mismo lugar.

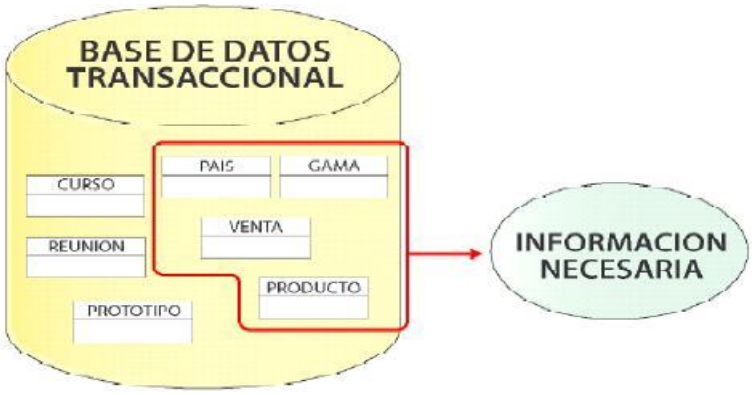

Fig. 4 Orientados a temas

- **Histórico:** Un factor importante en la toma de decisiones, es contar con información histórica para comparar datos en distintos períodos de tiempo e identificar tendencias. El tiempo debe

estar en todos los registros del almacén de datos, de manera que, cuando un dato entra en éste se sepa en qué momento tenía ese valor (Fig. 5).

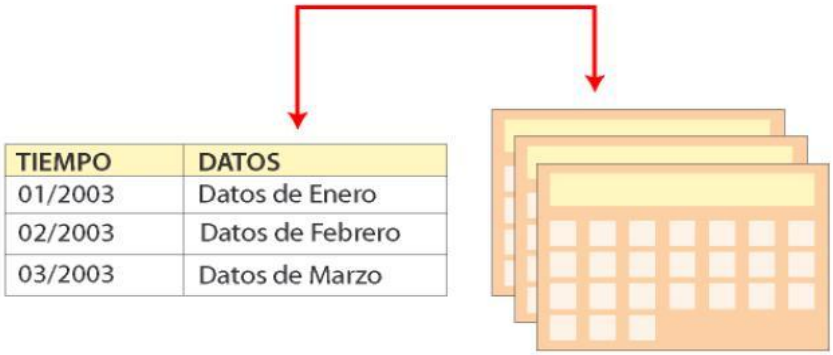

Fig. 5 Datos Históricos

- **No volátil:** La información en un almacén de datos existe para ser leída, pero no modificada, es por tanto permanente, significando que la actualización del almacén de datos es la incorporación de los últimos valores que tomaron las distintas variables contenidas en él, sin ningún tipo de acción sobre lo que ya existía (Fig. 6).

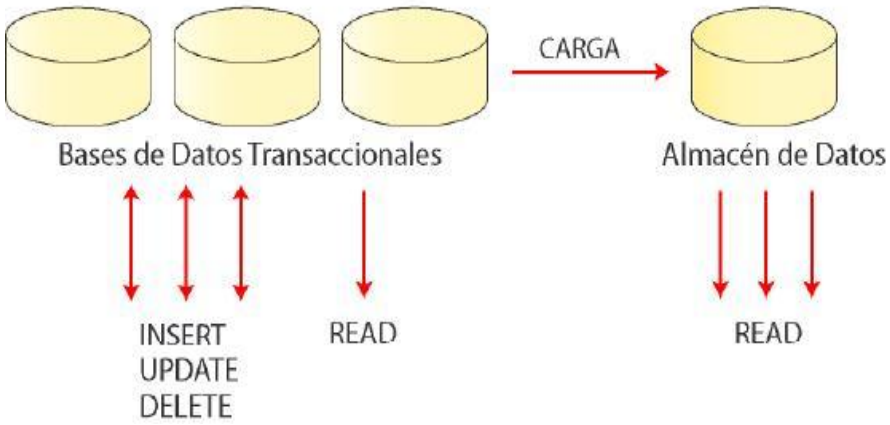

Fig. 6 No volátil

Otra característica del almacén de datos, es que contiene metadatos, es decir, datos sobre los datos. Los metadatos permiten saber la procedencia de la información, su fiabilidad y forma de cálculo. Por ejemplo, el título, tema, autor, tamaño de un archivo, constituye metadatos sobre el archivo. Información acerca de

las propiedades de datos tales como lógica de negocios que definen la estructura y contenido de dimensiones y medidas [9].

Además, un almacén de datos nunca es estático pues debe crecer y cambiar a medida que lo vayan haciendo las necesidades del negocio. Esto significa que deben ser diseñados para cambiar de forma constante ya que no es posible predecir los requerimientos de la información que habrá en el futuro, pues mientras el negocio crezca estos requerimientos cambiarán.

Estas características deben estar presentes en cualquier almacén de datos y serán críticas para medir el éxito o fracaso del mismo.

1.2.2. VENTAJAS DEL ALMACÉN DE DATOS

Los almacenes de datos se han convertido en la forma más utilizada por la gran mayoría de las organizaciones para dirigir la entrega de información y las necesidades analíticas debido a sus numerosos beneficios. Después de conocer sus características fundamentales, se muestran sus ventajas para una mejor comprensión.

- Transforma datos orientados a las aplicaciones en información orientada a la toma de decisiones.
- Integra y consolida diferentes fuentes de datos en una única plataforma sólida y centralizada.
- Provee la capacidad de analizar y explotar las diferentes áreas de trabajo.
- Elimina la producción y el procesamiento de datos que no son utilizados ni necesarios.
- La entrega de información a los usuarios va a ser completa, correcta, consistente, oportuna y accesible, en el momento adecuado y en el formato apropiado.
- Aumento de la competitividad de los encargados de tomar decisiones.
- Los usuarios pueden acceder directamente a la información en línea, lo que contribuye a su capacidad para operar con mayor efectividad en las tareas rutinarias o no. Además, pueden tener a su disposición una gran cantidad de valiosa información multidimensional, presentada coherentemente como fuente única, confiable y disponible en sus estaciones de trabajo. Así mismo, los usuarios tienen la facilidad de contar con herramientas que les son familiares para manipular y evaluar la información obtenida en el almacén de datos, tales como: hojas de

cálculo, procesadores de texto, software de análisis de datos, software de análisis estadístico, reportes, entre otros.

- Permite la toma de decisiones estratégicas y tácticas.

A pesar de las numerosas ventajas y características que posee un almacén de datos, es necesario construirlo, para ello se requiere de una metodología eficaz que sea capaz de guiar este proceso. En lo adelante se estudian un conjunto de las más utilizadas a nivel mundial.

1.3. METODOLOGÍAS PARA EL DESARROLLO DE UN ALMACÉN DE DATOS

La construcción de un almacén de datos, ha alcanzado un grado de madurez considerable a lo largo de estos años y presenta diferentes enfoques. En esta herramienta se han destacado numerosas metodologías que definen y guían todo el ciclo de vida del desarrollo concreto. A continuación se exponen las más reconocidas.

- **CRISP-DM**

Esta metodología consiste en un conjunto de tareas descritas en 4 niveles de abstracción es decir, de lo general a lo específico:

- **1er Nivel**: el proceso es organizado en un grupo de fases, cada una consta de varias tareas de segundo nivel, tarea genérica, tarea especializada, e instancia de procesos.
- **2do Nivel**: aquí se tratan las tareas genéricas, se le llama así porque está destinado a ser bastante general para cubrir todas las situaciones posibles de minería de datos. Éstas están destinadas a ser completas porque cubre tanto al proceso entero de minería de datos como a todas las aplicaciones permitidas y estables porque el modelo es válido tanto para acontecimientos normales como para desarrollos imprevistos como es el caso de las nuevas técnicas de modelado.
- **3er Nivel**: conocido como el de tarea especializada, se describe cómo deberían ser realizadas las acciones en las tareas genéricas en ciertas situaciones específicas.
- **4to Nivel**: la instancia de procesos, es un registro de las acciones, decisiones, y de los resultados de una minería de datos real contratada. Una instancia de procesos está organizada

según las tareas definidas en los niveles más altos, pero representa lo que en realidad pasó en un contrato particular a diferencia de lo que pasa en general.

El modelo provee una representación completa del ciclo de vida de un proyecto de minería de datos dividiéndolo en 6 fases que interactúan entre ellas de forma iterativa durante el desarrollo del mismo (Fig. 7). Las flechas indican las dependencias más importantes y frecuentes entre ellas, el círculo exterior simboliza la naturaleza cíclica de un proyecto de esta característica [10].

Fig. 7 Fases del modelo CRISP-DM

- **Metodología Kimball**

La metodología Kimball se enfoca principalmente en la construcción de un almacén de datos. Esto se basa en la creación de tablas de hechos, las cuales son tablas que contienen la información numérica de los indicadores a analizar, es decir, la parte cuantitativa [11].

En la figura se muestra el ciclo de vida de la metodología Kimball para el desarrollo de un proyecto, ya sea un Datamart o un almacén de datos (Fig. 8).

Está compuesta por 4 fases las cuales son:

1. Requerimientos y gestión de proyectos.
2. Arquitectura técnica.

3. Implementación.
4. Implementación y crecimiento.

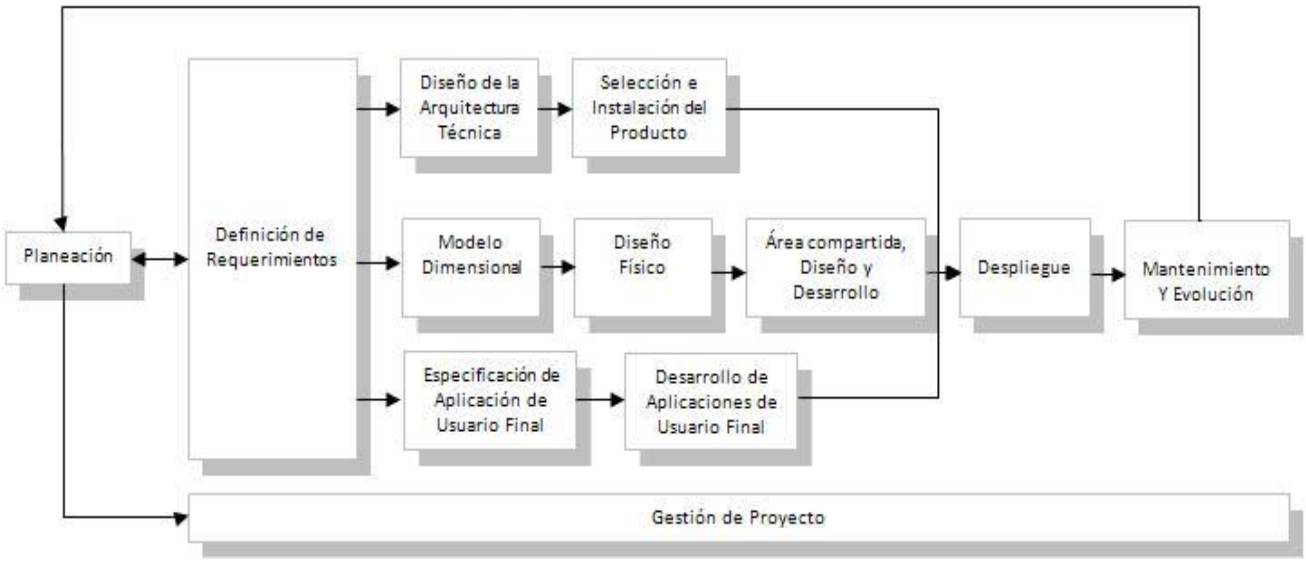

Fig. 8 Ciclo de vida de la metodología Kimball

Como se evidencia en la figura anterior, esta metodología propone dentro de cada una de sus fases la realización de un conjunto de actividades, lo que la convierte en una metodología madura y robusta, por lo que demanda más recursos, tiempo y documentación.

- **Metodología Rapid Warehousing**

El uso de esta metodología permite obtener resultados tangibles en un corto espacio de tiempo. Esta es una metodología iterativa y basada en el desarrollo incremental del proyecto de un almacén de datos dividido en 5 fases [12] (Fig. 9).

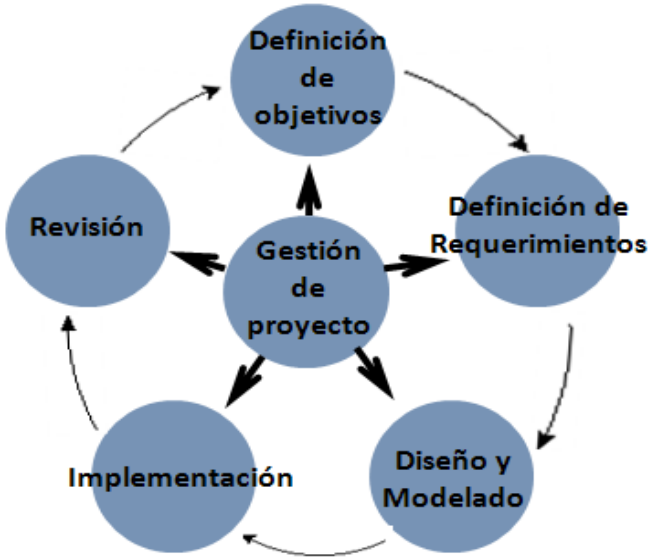

Fig. 9 Fases de la metodología Rapid Warehousing

- **Definición de los objetivos:** En esta fase se especificará el equipo de proyecto, el alcance del sistema y cuáles son las funciones que el almacén de datos realizará como suministrador de información de negocio estratégica para la empresa. Se definirán así mismo, los parámetros que permitan evaluar el éxito del proyecto.
- **Definición de los requerimientos de información:** Tal como sucede en todo tipo de proyectos, sobre todo si se involucran técnicas novedosas como son las relativas al almacén de datos, es importante analizar las necesidades y hacer comprender las ventajas que este sistema puede reportar.
- **Diseño y modelado:** Los requerimientos de información identificados durante la fase anterior proporcionarán las bases para realizar el diseño y el modelado del almacén de datos. En esta fase se identificarán las fuentes de los datos (sistema operacional, fuentes externas) y las transformaciones necesarias para, a partir de dichas fuentes, obtener el modelo lógico del almacén de datos. Este modelo estará formado por entidades y relaciones que permitirán resolver las necesidades del negocio de la organización.
- **Implementación:** La implementación de un almacén de datos lleva implícitos los siguientes pasos:

1. Extracción de los datos del sistema operacional y transformación de los mismos.
2. Carga de los datos validados en el almacén de datos. Esta carga debe ser planificada con una periodicidad que se adaptará a las actualizaciones detectadas durante las fases de diseño del nuevo sistema.
3. Explotación del almacén de datos mediante diversas técnicas dependiendo del tipo de aplicación que se le dé a los datos.

- **Revisión:** La construcción del almacén de datos no finaliza con la implantación del mismo, sino que es una tarea iterativa en la que se trata de incrementar su alcance aprendiendo de las experiencias anteriores. Después de implantarse, se debería revisar, planteando preguntas que permitan, después de los seis o nueve meses posteriores a su puesta en marcha, definir cuáles serían los aspectos a mejorar o potenciar en función de la utilización que se haga del nuevo sistema.

- **Metodología Hefesto**

Esta metodología permite la construcción de un almacén de datos de forma sencilla, ordenada e intuitiva. Hefesto es una metodología bien fundamentada y explícita que permite facilitar un almacén de datos de manera metódica y sencilla, guiándose por pasos lógicos relacionados sólidamente durante todas las etapas del proceso de confección [13].

La metodología Hefesto, comienza recolectando las necesidades de información de los usuarios y se obtienen las preguntas claves del negocio. Luego, se identifican los indicadores resultantes de las interrogantes y sus respectivas perspectivas de análisis, mediante estos se construye el modelo conceptual del almacén de datos. Después, se analizan 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, explicitando las jerarquías que intervendrán. Por último, se definirán los procesos de carga, transformación, extracción y limpieza de los datos fuente (Fig. 10).

Fig. 10 Pasos de la metodología Hefesto

Esta metodología es ágil y madura, propone un conjunto de fases que con pocos recursos, tiempo y documentación, permite realizar un almacén de datos.

JUSTIFICACIÓN DE LA METODOLOGÍA A UTILIZAR

Como se evidencia existen diversas metodologías que pretenden dar un acercamiento a una propuesta para el desarrollo de un almacén de datos. Todas se orientan a la optimización del rendimiento y a su visión de los principales procesos que se deben tener en cuenta para construir esta herramienta de forma flexible y dinámica. A partir del almacén de datos que se desea desarrollar es que se realiza la elección de una u otra metodología.

La red nacional de Genética Médica necesita para la extracción de conocimiento a partir del análisis de datos de los estudios que generan, un almacén de datos en un período de tiempo corto, por lo que requiere la utilización de una metodología sencilla, ágil y madura que garantice el éxito de la integración de la información que actualmente disponen.

Para enfrentar el análisis del almacén de datos, se decide utilizar por las razones anteriormente mencionadas y las siguientes características, la metodología Hefesto.

Características de la metodología Hefesto:

- 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 que su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios del 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 almacén de datos.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente de las herramientas que se utilicen para su implementación.
- Es independiente de las estructuras físicas que contenga el almacén de datos 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.

CONCLUSIONES DEL CAPÍTULO

A partir del estudio realizado sobre las herramientas, técnicas y tecnologías de un sistema de Inteligencia de Negocios se define el uso del almacén de datos para darle solución al problema en cuestión. Se decidió utilizar la metodología Hefesto para el desarrollo del mismo.

CAPÍTULO 2. ANÁLISIS DE UN ALMACÉN DE DATOS PARA LA RED NACIONAL DE GENÉTICA MÉDICA

En este capítulo se mostrarán las fases a seguir para la realización del análisis del almacén de datos y se explican los pasos de cada una de ellas. Se define un modelo conceptual que luego será ampliado a partir del establecimiento de los indicadores y perspectivas identificados de los requerimientos del cliente. Finalmente, se establecerán las correspondencias entre el modelo conceptual y los OLTP.

2.1. ANÁLISIS DE REQUERIMIENTOS

El análisis de los requerimientos del cliente constituye la primera fase de la metodología Hefesto, la cual consta de 3 pasos.

El primer paso comienza con la recopilación de las necesidades de información, el cual puede llevarse a cabo a través de muy variadas y diferentes técnicas, cada una de las cuales posee características inherentes y específicas, como por ejemplo entrevistas, cuestionarios, observaciones, entre otras.

Después de tener las preguntas elaboradas derivadas de los requerimientos del cliente se procede como segundo paso a identificar los indicadores y perspectivas que intervendrán en el análisis.

Finalmente, teniendo en cuenta los indicadores y perspectivas identificados se realiza un modelo conceptual, donde se visualiza de manera clara los primeros resultados alcanzados.

El objetivo principal de esta fase, es obtener e identificar las necesidades de información clave de alto nivel, que facilitará una eficiente toma de decisiones.

2.1.1. PASO 1: IDENTIFICAR PREGUNTAS

En las entrevistas con los genetistas se investigó cuáles eran sus necesidades, los resultados que esperaban y la información clave que considerasen más importante, donde se viesen reflejadas las actividades más relevantes de los estudios que realizan y que estuviese de alguna manera soportado por algún OLTP.

A continuación, se analizó qué era lo que les interesaba conocer de dichos estudios y se obtuvo un proceso, denominado Casos de genética. Esta información está enfocada a las personas clasificadas como: gemelos, discapacitados y retrasados mentales.

A partir del intercambio con los especialistas de genética se definieron como preguntas claves las siguientes:

1. Se desea saber la cantidad de gemelos que padecen de cáncer y que convivan con familiares adictos a la bebida en un período de tiempo determinado.
2. Se desea saber en un determinado período de tiempo la cantidad de gemelos agresivos con buena redacción y antisociales.
3. Se desea saber el porcentaje de gemelos que han fallecido con más de 30 años y cuyos padecimientos hayan sido epilepsia o esquizofrenia en un período de tiempo determinado.
4. Se desea saber la cantidad de gemelos que han nacido en la Provincia de Santiago de Cuba, separados desde los 3 años y con patología de cáncer en un período de tiempo determinado.
5. Se desea saber el porcentaje de gemelos que su edad esté en el rango de 15 a 30 años, de género femenino en un determinado período de tiempo.
6. Se desea saber la cantidad de gemelos indocumentados que han fallecido sin sobrepasar el nivel de escolaridad de 9no grado, dado un período de tiempo determinado.
7. Se desea saber el porcentaje de pacientes gemelos con discapacidad física o retraso mental que padecen de cáncer en un período de tiempo determinado.
8. Se desea saber la cantidad de pacientes gemelos que presentan enfermedades congénitas y alguna discapacidad física o mental en un determinado período de tiempo.
9. Se desea saber la cantidad de gemelos con retraso mental que padezcan de esquizofrenia y sean agresivos durante un determinado período de tiempo.
10. Se desea saber el porcentaje de gemelos que tengan enfermedades genéticas y retraso mental en un determinado período de tiempo.
11. Se desea saber dado un tiempo determinado la cantidad de gemelos con sicklemia, mayores de 60 años con alguna discapacidad física o mental.
12. Se desea saber la cantidad de personas con discapacidad física, que sean gemelos, que consuman alcohol de forma dependiente y que han fallecido en un período de tiempo determinado.

13. Se desea saber la cantidad de discapacitados sordos de la provincia de Villa Clara que vivan en malas condiciones y necesiten ayuda económica en un determinado período de tiempo.
14. Se desea saber cuántos pacientes mayores de 20 años con discapacidad física tienen una escolaridad inferior a 12mo grado en un determinado período de tiempo.
15. Se desea saber la cantidad de pacientes ciegos con amparo filiar que tienen educación especial, en un determinado período de tiempo.
16. Se desea saber la cantidad de pacientes discapacitados menores de 30 años que tengan capacidad laboral y no estén vinculados a ninguna ocupación en un período de tiempo determinado.
17. Se desea saber la cantidad de pacientes con discapacidad física, que convivan con menos de dos personas, que estén postrados y necesiten servicio de limpieza del hogar dado un determinado período de tiempo.
18. Se desea saber la cantidad de pacientes con retraso mental que presentan malformaciones internas cuyas madres durante el embarazo tuvieron hábitos tóxicos en un determinado período de tiempo.
19. Se desea saber la cantidad de pacientes con retraso mental menores de 6 años, por provincia que hayan fallecido en un determinado período de tiempo.
20. Se desea saber el porcentaje de personas con retraso mental que las madres presentaron enfermedades infecciosas durante el embarazo en un determinado período de tiempo.
21. Se desea saber el año en que nacieron mayor número de mujeres con retraso mental que presentaron psicosis primaria.
22. Se desea saber la cantidad de personas con retraso mental, que presentan manchas pequeñas y la madre durante el embarazo recibió radiación en un período de tiempo determinado.

En las preguntas escritas quedan evidenciadas las principales necesidades de información de los especialistas. La identificación de las preguntas conlleva al siguiente paso.

2.1.2. Paso 2: Identificar indicadores y perspectivas

Este es el segundo paso a realizar dentro de la fase Análisis de Requerimientos, el cual tiene como objetivo descomponer las preguntas elaboradas en el paso anterior, para identificar los indicadores que se utilizarán y las perspectivas de análisis que intervendrán.

La bibliografía define el término de indicadores y perspectivas de almacenes de datos de la manera siguiente:

Indicadores: generalmente se expresan a través de un valor numérico. Representan lo que se desea analizar concretamente, ejemplo: saldos, porcentajes, cantidades, entre otros [13].

Perspectivas: se refieren a los objetos mediante los cuales se quiere examinar los indicadores con el fin de responder a las preguntas planteadas [13].

A continuación los indicadores y perspectivas se identificarán con la iconografía siguiente:

- Indicadores
- Perspectivas

1. Cantidad de gemelos con cáncer y que convivan con familiares adictos a la bebida en un período de tiempo determinado.
   - Cantidad de casos
   - Enfermedades, Persona, Fecha
2. Cantidad de gemelos agresivos que sean antisociales y tengan buena redacción en un período de tiempo determinado.
   - Cantidad de casos
   - Conducta social, Persona, Fecha
3. Porcentaje de gemelos fallecidos con más de 30 años y que hayan padecido de epilepsia o esquizofrenia en un período de tiempo determinado.
   - Porcentaje de casos
   - Persona, Enfermedades, Fecha

4. Cantidad de gemelos de la Provincia Santiago de Cuba, separados desde los 3 años y con cáncer en un período de tiempo determinado.
   - Cantidad de casos
   - Lugar, Persona, Enfermedades, Fecha
5. Porcentaje de gemelos femeninos que su edad esté en el rango de 15 a 30 años, en un determinado período de tiempo.
   - Porcentaje de casos
   - Persona, Fecha
6. Cantidad de gemelos fallecidos sin sobrepasar el nivel de escolaridad de 9no grado y que sean indocumentados dado un período de tiempo determinado.
   - Cantidad de casos
   - Persona, Fecha
7. Porcentaje de gemelos con cáncer que tienen discapacidad física o retraso mental en un período de tiempo determinado.
   - Porcentaje de casos
   - Enfermedades, Persona, Fecha
8. Cantidad de gemelos con enfermedades congénitas y alguna discapacidad física o retraso mental en un determinado período de tiempo.
   - Cantidad de casos
   - Enfermedades, Persona, Fecha
9. Cantidad de gemelos con retraso mental que padezcan de esquizofrenia y sean agresivos durante un período de tiempo determinado.
   - Cantidad de casos
   - Persona, Conducta
   - social, Enfermedades, Fecha

10. Porcentaje de gemelos que tengan enfermedades genéticas y retraso mental en un determinado período de tiempo.
    - Porcentaje de casos
    - Enfermedades, Persona, Fecha
11. Cantidad de gemelos con sicklemia, mayores de 60 años con alguna discapacidad física o retraso mental en un período de tiempo determinado.
    - Cantidad de casos
    - Enfermedades, Persona, Fecha
12. Cantidad de discapacitados fallecidos que sean gemelos y dependientes del alcohol en un período de tiempo determinado.
    - Cantidad de casos
    - Persona, Conducta social, Fecha
13. Cantidad de discapacitados sordos de la provincia de Villa Clara que vivan en malas condiciones y necesiten ayuda económica en un período de tiempo determinado.
    - Cantidad de casos
    - Persona, Lugar, Vivienda, Fecha
14. Cantidad de discapacitados físicos con escolaridad inferior a 12 grado, mayores de 20 años en un determinado período de tiempo.
    - Cantidad de casos
    - Persona, Fecha
15. Cantidad de discapacitados ciegos con amparo filiar que tienen educación especial, en un determinado período de tiempo.
    - Cantidad de casos
    - Persona, Fecha

16. Cantidad de pacientes discapacitados menores de 30 años que tengan capacidad laboral y no estén vinculados a ninguna ocupación en un período de tiempo determinado.
    - Cantidad de casos
    - Persona, Fecha
17. Cantidad de pacientes discapacitados físicos, que vivan con menos de 2 personas, que estén postrados y necesiten servicio de limpieza del hogar dado un determinado período de tiempo.
    - Cantidad de casos
    - Enfermedades, Vivienda, Persona, Fecha
18. Cantidad de pacientes con retraso mental que presentan malformaciones internas cuyas madres durante el embarazo tuvieron hábitos tóxicos en un determinado período de tiempo.
    - Cantidad de casos
    - Persona, Fecha
19. Cantidad de pacientes con retraso mental menores de 6 años por provincia que hayan fallecido en un determinado período de tiempo.
    - Cantidad de casos
    - Persona, Lugar, Fecha
20. Porcentaje de pacientes con retraso mental que las madres presentaron enfermedades infecciosas durante el embarazo en un determinado período de tiempo.
    - Porcentaje de casos
    - Persona, Fecha
21. Año en que nacieron mayor número de mujeres con retraso mental que presentaron psicosis primaria.
    - Cantidad de casos
    - Persona, Enfermedades, Fecha

22. Cantidad de pacientes con retraso mental, que presentan manchas pequeñas y que la madre durante el embarazo recibió radiación en un período determinado de tiempo.
    - Cantidad de casos
    - Enfermedades, Persona, Fecha

En la siguiente tabla se relacionan los indicadores y perspectivas que se identificaron entre todas las preguntas. Con esta información se procederá a la realización del modelo conceptual.

Perspectivas	Indicadores
Persona	Cantidad de casos
Enfermedades	Porcentaje de casos
Vivienda	
Lugar	
Conducta_ social	
Fecha	

2.1.3. Paso 3: Modelo Conceptual

Constituye el tercer paso de esta fase. En esta etapa, se construirá un modelo conceptual a partir de los indicadores y perspectivas obtenidas en el paso anterior (Fig. 11).

Como parte del estándar de un modelo conceptual se colocan a la izquierda las perspectivas seleccionadas, que serán unidas a un óvalo central que representa y lleva el nombre de la relación que existe entre ellas. La relación, constituye el proceso o área de estudio elegida. De dicha relación y entrelazadas con flechas, se desprenden los indicadores, estos se ubican a la derecha del esquema.

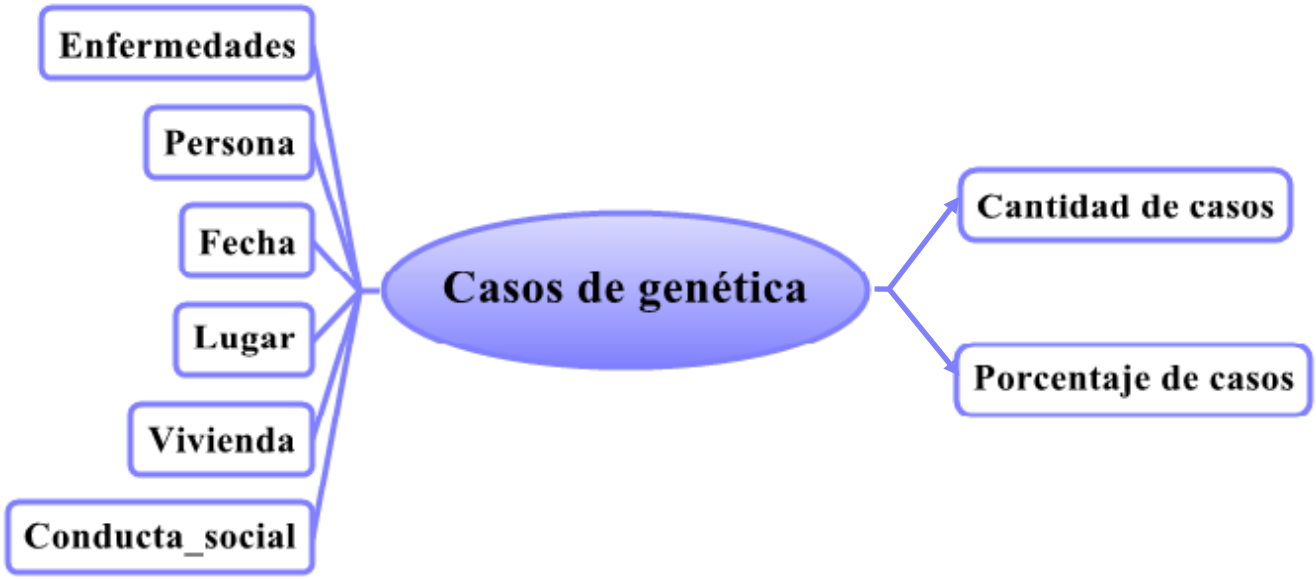

Fig. 11 Modelo conceptual del proceso Casos de genética.

2.2. ANÁLISIS DE LOS OLTP

Con la realización del modelo conceptual se concluye la primera fase propuesta por la metodología Hefesto para el desarrollo de almacenes de datos y se le da inicio a la segunda fase. Ésta tiene como objetivo realizar el análisis de los OLTP el cual se define como una base de datos para soportar procesos transaccionales en línea. Representa toda aquella información transaccional que genera la empresa en su accionar diario. Entre los OLTP más habituales que pueden existir en cualquier organización se encuentran:

- Archivos de textos.
- Hipertextos.
- Hojas de cálculos.
- Informes semanales, mensuales, anuales, entre otros.

La información recopilada de los estudios generados está soportada por un solo OLTP, la base de datos de alasMEDIGEN.

2.2.1. PASO 1: ESTABLECER CORRESPONDENCIA CON LOS REQUERIMIENTOS.

Establecer la correspondencia con los requerimientos es el primer paso de la fase, la cual propone examinar los OLTP disponibles que contengan la información requerida, como así también sus características, para poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos.

En el caso de los indicadores, deben explicarse como se calcularán, y más aún, si son fórmulas u operaciones complejas que dependan de algún atributo del OLTP. La idea es, que todos los elementos del modelo conceptual estén correspondidos en los OLTP.

- **Cálculo de los indicadores :**
  - **Cantidad de casos:** se obtiene del conteo de todos los códigos almacenados de los casos genéticos.
  - **Porcentaje de casos:** se obtiene con la multiplicación de los casos que cumplen la condición por 100 y este resultado se divide entre la cantidad de casos.

En las siguientes figuras se expone la correspondencia entre los elementos del modelo conceptual y los OLTP. Esto consiste en identificar las tablas con las cuales se relaciona cada perspectiva.

La perspectiva **Persona** se relaciona con las tablas:

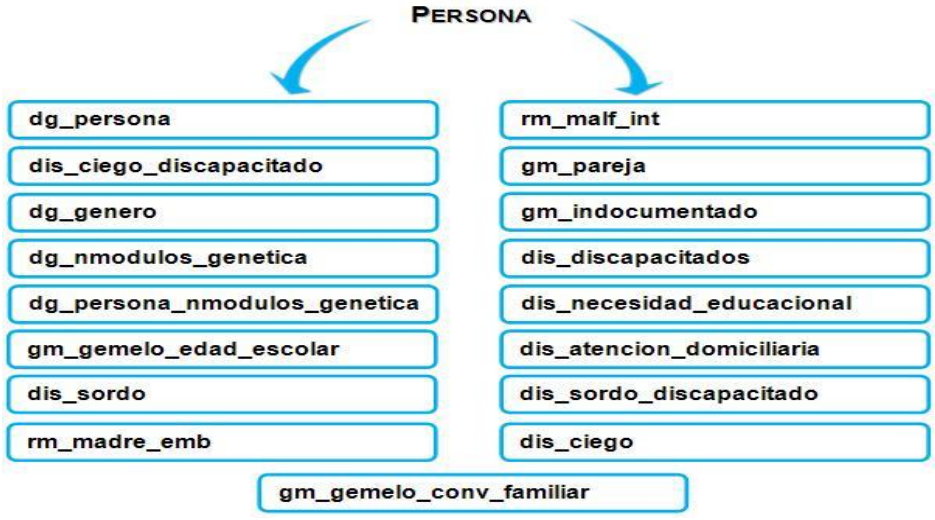

Fig. 12 Perspectiva Persona con las tablas con las que se relaciona.

La perspectiva **Enfermedades** se relaciona con las tablas:

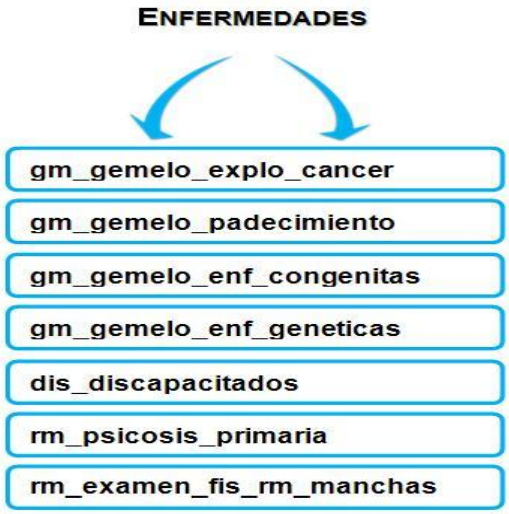

Fig. 13 Perspectiva Enfermedades con las tablas con las que se relaciona.

La perspectiva **Conducta_social** se relaciona con las tablas:

Fig. 14 Perspectiva Conducta_social con las tablas con las que se relaciona.

La perspectiva **Vivienda** se relaciona con las tablas:

Fig. 15 Perspectiva Vivienda con las tablas con las que se relaciona.

La perspectiva **Lugar** se relaciona con las tablas:

Fig. 16 Perspectiva Lugar con las tablas con las que se relaciona.

Después de identificadas las tablas que se relacionan por cada perspectiva, se procederá a identificar los atributos o campos de cada una de ellas.

2.2.2. Paso 2: Seleccionar los campos que integrarán cada perspectiva. Nivel de granularidad

Una vez que se han establecido las relaciones con los OLTP, se examinarán y seleccionarán los campos que contendrá cada perspectiva, ya que será a través de estos por los que se manipularán y filtrarán los indicadores.

Es muy importante conocer en detalle qué significa cada campo y/o valor de los datos encontrados en los OLTP, por lo cual, es conveniente investigar su sentido, ya sea a través de diccionarios de datos, reuniones con los encargados del sistema, análisis de los datos propiamente dichos, entre otros.

Finalmente, se ampliará el modelo conceptual expuesto anteriormente, a través de una imagen que visualiza los resultados obtenidos (Fig. 39).

- **Descripción de los campos:**

Primero se examinó la base de datos y su descripción para comprender los significados de cada campo. Los nombres de estos son bastantes explícitos y se deducen con facilidad, pero aún así fue necesario investigarlos para evitar cualquier tipo de inconvenientes. Es necesario aclarar que los nombres de los campos se escriben con minúsculas y no se acentúan, además si existe alguno compuesto por dos

palabras, se le agrega entre ellas un guión bajo o se unen, la idea es que entre las dos no puede haber un espacio.

A continuación se describen los campos que componen cada perspectiva y se representa mediante una figura su relación con las tablas de la base de datos de alasMEDIGEN.

Con respecto a la perspectiva **"Persona"**, los datos disponibles son los siguientes:

1. **id_persona**: es la clave primaria de la tabla "Persona", y representa unívocamente a una persona en particular.
2. **tipo**: clasificación de la persona, como: gemelos, discapacitados o retrasados mentales.

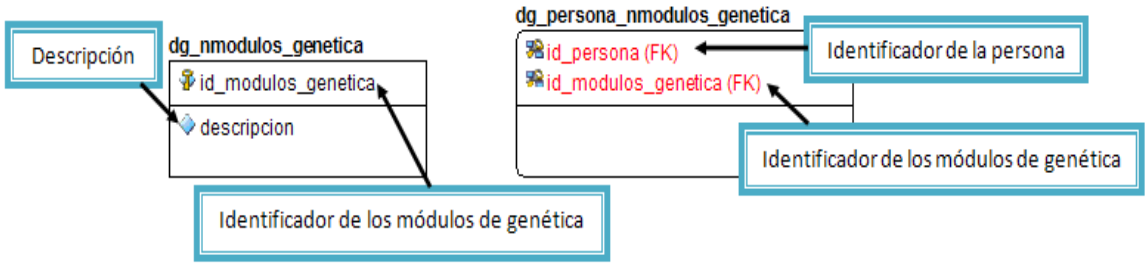

Fig. 17 Tablas dg_nmodulos_genetica y dg_persona_nmodulos_genetica de la base datos de alasMEDIGEN.

3. **edad**: muestra la edad de la persona.

Fig. 18 Tabla dg_persona de la base de datos de alasMEDIGEN.

4. **fallecido**: indica si la persona es fallecida o no.

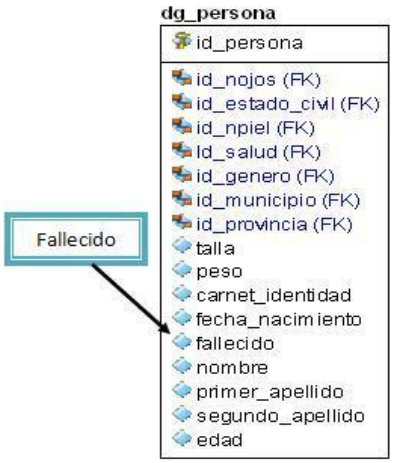

Fig. 19 Tabla dg_persona de la base de datos de alasMEDIGEN.

5. **genero**: representa el sexo de la persona (femenino o masculino).

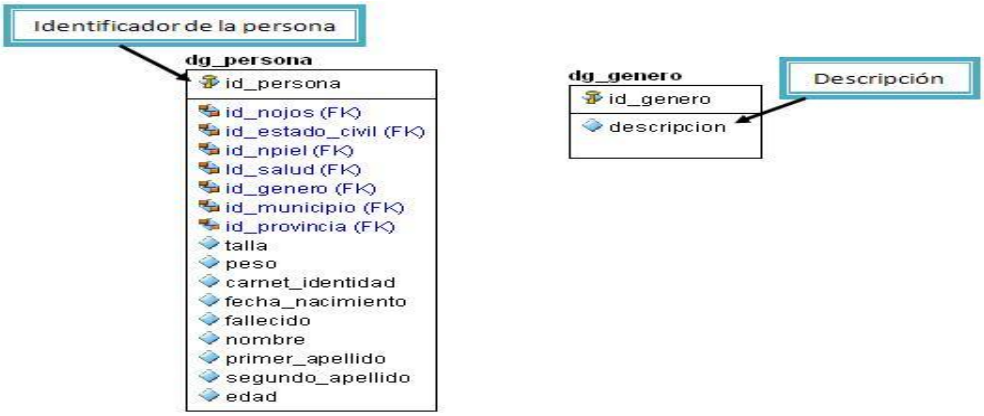

Fig. 20 Tablas dg_persona y dg_genero de la base de datos de alasMEDIGEN.

6. **familiares_adictos**: indica si la persona posee un familiar adicto a la bebida o no.

Fig. 21 Tabla gm_gemelo_conv_familiar de la base de datos de alasMEDIGEN.

7. **buena_redaccion**: especifica si la persona tiene o no buena redacción.

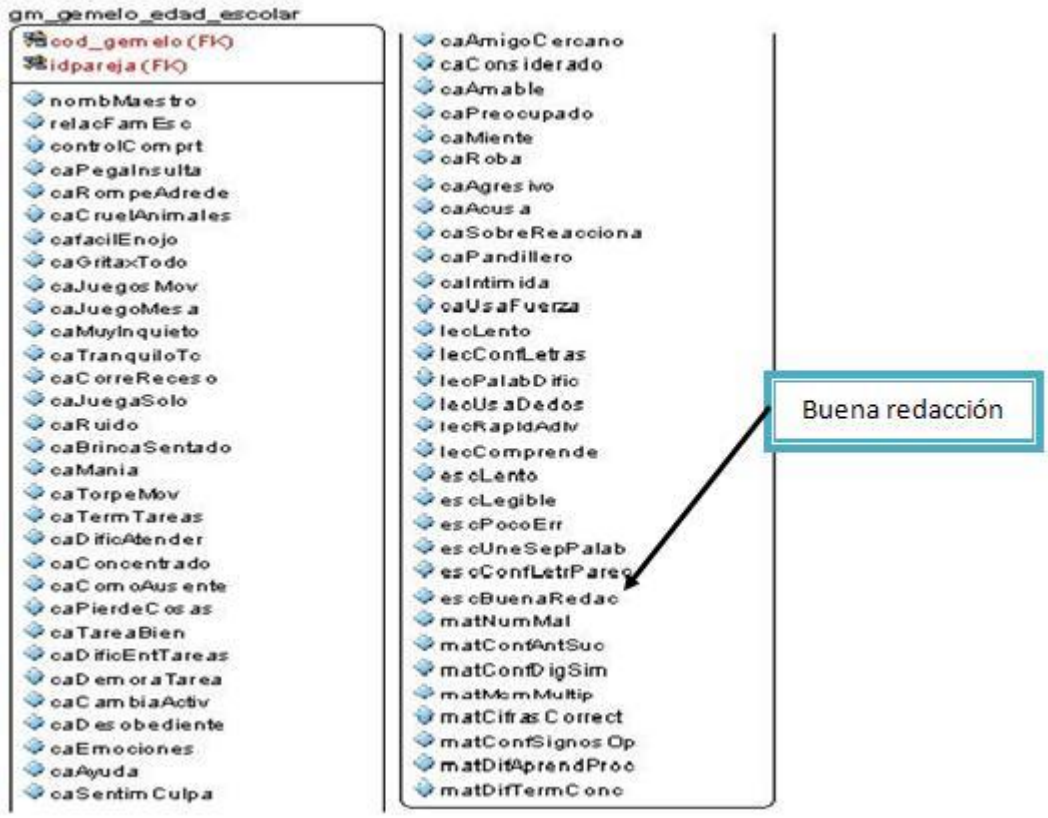

Fig. 22 Tabla gm_gemelo_edad_escolar de la base de datos de alasMEDIGEN.

8. **edad_de_separacion**: define la edad en que los gemelos fueron separados. En caso de ser otro tipo de persona el campo será null.

Fig. 23 Tabla gm_pareja de la base de datos de alasMEDIGEN.

9. **indocumentados**: verifica si una persona tiene problema en la documentación. En caso de no ser gemelos, el campo será null.

Fig. 24 Tabla gm_indocumentado de la base de datos de alasMEDIGEN.

10. **ocupacion**: este campo indica la ocupación de la persona, como: ama de casa, trabajador, estudiante, jubilado, pensionado, desempleado, asistenciado.

Fig. 25 Tabla dis_discapacitado de la base de datos de alasMEDIGEN.

11. **escolaridad**: este campo almacena el nivel escolar de la persona, como: preescolar, primaria, secundaria, preuniveritario y universitaria.

Fig. 26 Tabla dis_discapacitado de la base de datos de alasMEDIGEN.

12. **capacidad_laboral**: este campo especifica si la persona está apta o no laboralmente. En caso de no ser discapacitado el campo será null.

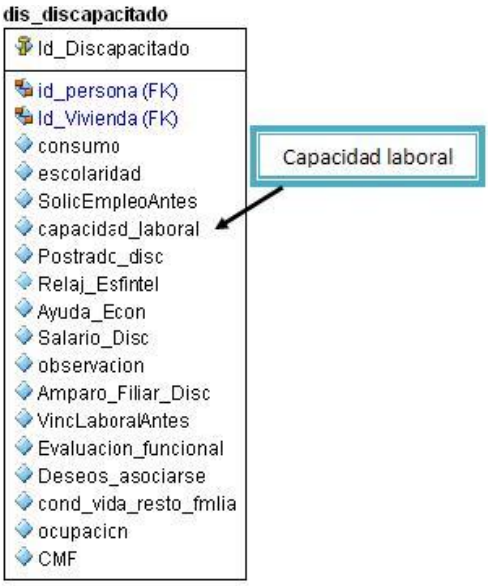

Fig. 27 Tabla dis_discapacitado de la base de datos de alasMEDIGEN.

13. **malformaciones_internas**: especifica el tipo de malformaciones internas, como: sistema nervioso, columna, digestivas, renales y corazón.

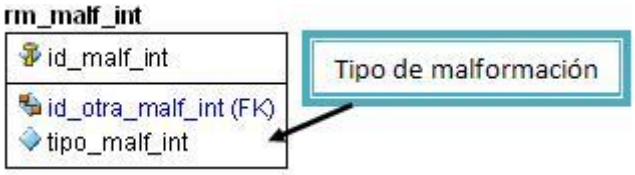

Fig. 28 Tabla rm_malf_int de la base de datos de alasMEDIGEN.

14. **tipo_de_discapacidad**: indica el tipo de discapacidad de la persona, como: sordo y ciego. En caso de no ser discapacitado el campo será null.

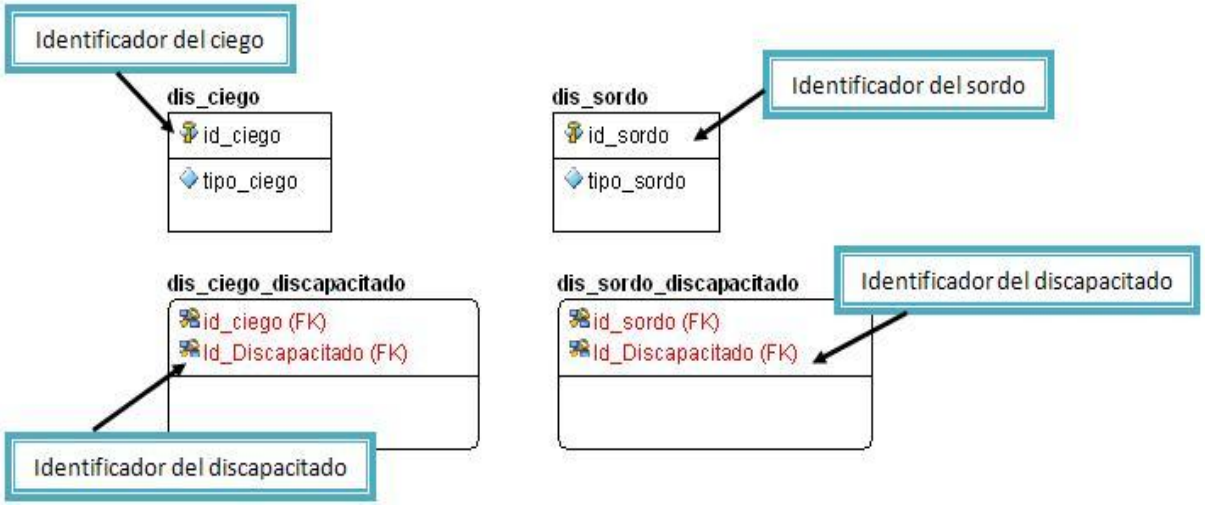

Fig. 29 Tablas dis_ciego, dis_sordo, dis_ciego_discapacitado y dis_sordo_discapacitado de la base de datos de alasMEDIGEN.

15. **causa_del_retraso**: este campo define las causas del retraso, como: madre con hábitos tóxicos o enfermedades infecciosas o recibió radiación durante el embarazo. En caso de no ser retrasado mental, el campo será null.

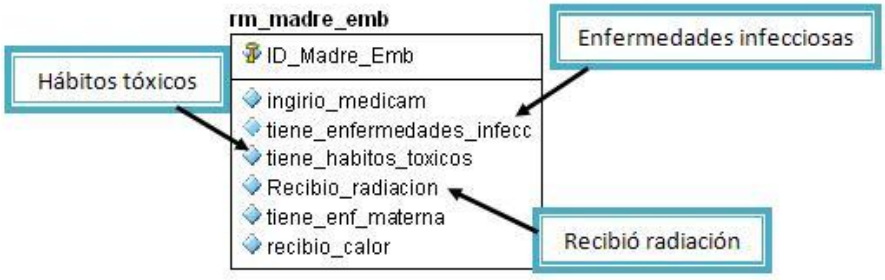

Fig. 30 Tabla rm_madre_emb de la base de datos de alasMEDIGEN.

16. **servicio_recibido**: especifica que servicio necesita una persona discapacitada, como: ayuda económica, amparo filiar, educación especial y servicio de limpieza. En caso de ser otro tipo de persona el campo será null.

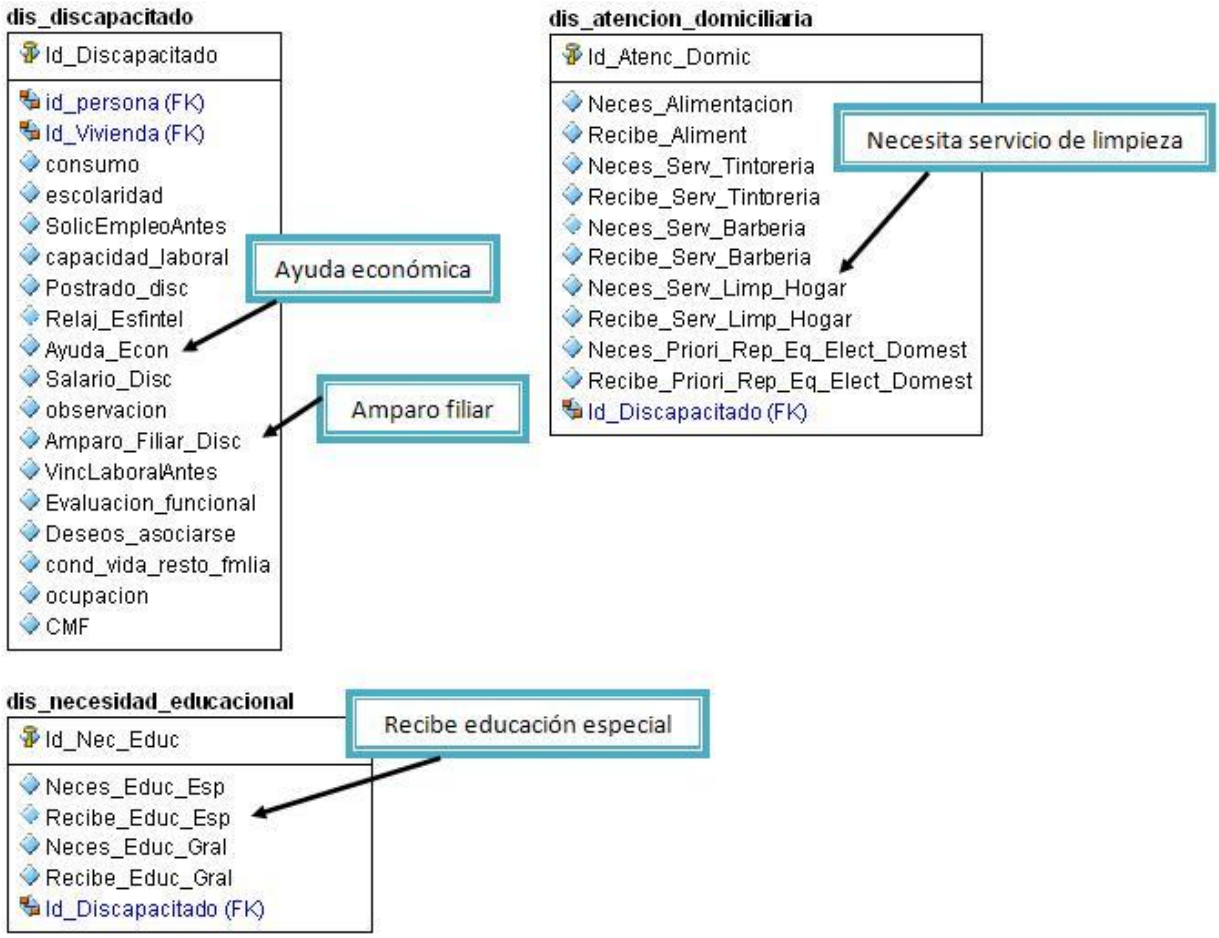

Fig. 31 Tablas dis_discapacitado, dis_atencion_domiciliaria y dis_necesidad_educacional de la base de datos de alasMEDIGEN.

Con respecto a la perspectiva **"Enfermedades"**, los datos disponibles son los siguientes:

1. **id_enfermedad**: es la clave primaria de la tabla "Enfermedades", y representa unívocamente a una enfermedad en particular.
2. **enfermedad**: indica la enfermedad de la persona, como: cáncer, sicklemia, esquizofrenia, epilepsia, psicosis primaria, manchas pequeñas, postrado, enfermedades genéticas y congénitas.

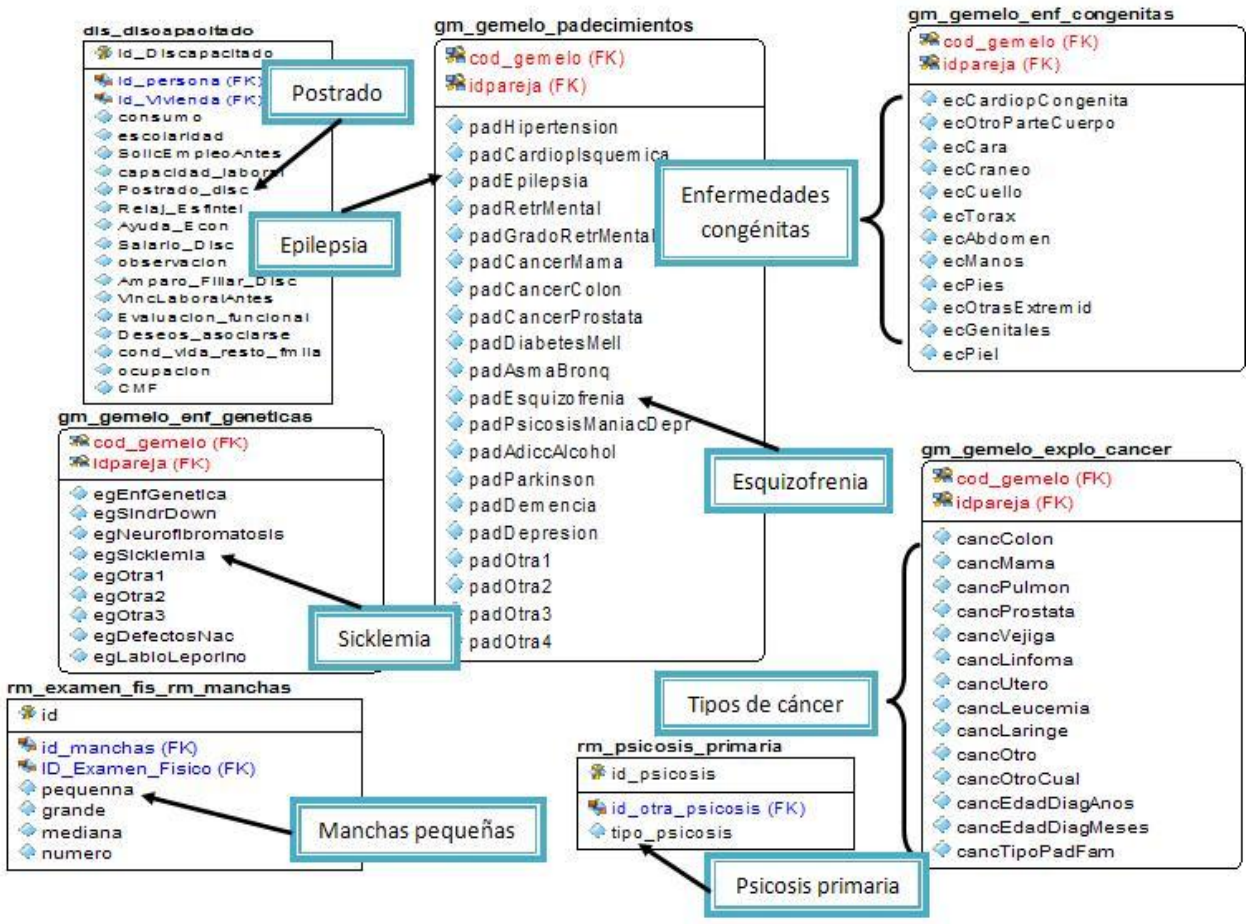

Fig. 32 Tablas rm_psicosis_primaria, rm_examen_fis_rm_manchas, gm_gemelo_explo_cancer, gm_gemelo_enf_geneticas, gm_gemelo_enf_congenitas, gm_gemelo_padecimientos y dis_discapacitado de la base de datos de alasMEDIGEN.

Con respecto a la perspectiva **”Conducta_social”**, los datos disponibles son los siguientes:

1. **id_conducta_social**: es la clave primaria de la tabla ”Conducta_social”, y representa una conducta en específico.
2. **conducta**: indica el tipo de conducta de la persona, como: agresivo, antisocial y alcohólico.

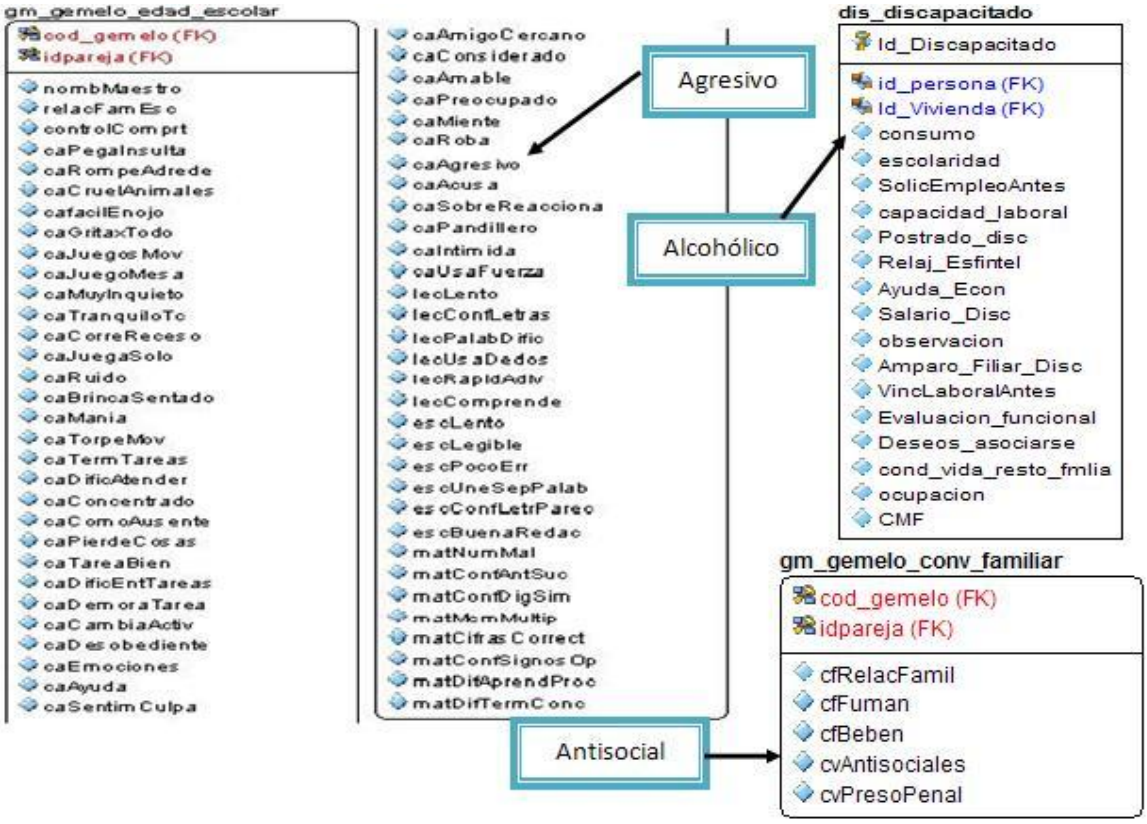

Fig. 33 Tablas gm_gemelo_edad_escolar, gm_gemelo_conv_familiar, y dis_discapacitado de la base de datos de alasMEDIGEN.

Con respecto a la perspectiva **"Vivienda"**, los datos disponibles son los siguientes:

1. **id_vivienda**: es la clave primaria de la tabla " Vivienda", y representa unívocamente a una vivienda en particular.
2. **condicion**: indica las condiciones de la vivienda, como: buena, regular, mala y crítica.

Fig. 34 Tabla dis_vivieda de la base de datos de alasMEDIGEN.

3. **nucleo_familiar**: especifica la cantidad de habitantes del núcleo.

Fig. 35 Tabla dis_vivieda de la base de datos de alasMEDIGEN.

Con respecto a la perspectiva **"Lugar"**, los datos disponibles son los siguientes:

1. **id_lugar**: es la clave primaria de la tabla "Lugar", y representa unívocamente a un lugar en específico.
2. **provincia**: representa la provincia a la que pertenece una persona.

Fig. 36 Tablas dg_persona y dg_nprovincia de la base de datos de alasMEDIGEN.

3. **municipio**: representa el municipio donde reside una persona.

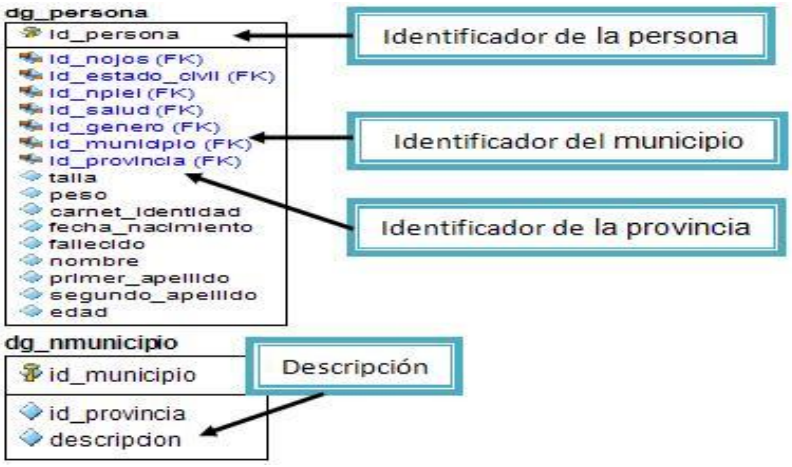

Fig. 37 Tablas dg_persona y dg_nmunicipio de la base de datos de alasMEDIGEN.

Con respecto a la perspectiva “Fecha” los datos disponibles son los siguientes:

1. **id _ fecha**: Es la clave primaria de la tabla ”Fecha”, y representa unívocamente a una fecha en particular.
2. **año**: muestra el año en que se realizó la consulta.
3. **mes**: muestra el mes en que se realizó la consulta.
4. **dia**: muestra el día en que se realizó la consulta.
5. **trimestre**

Esta tabla registra la fecha de la consulta en la que se clasifica un paciente, como: gemelo, discapacitado físico o retrasado mental.

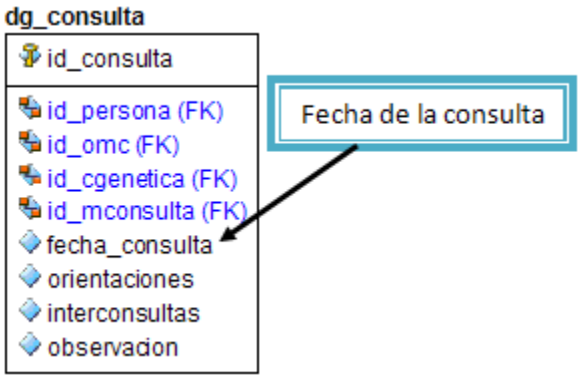

Fig. 38 Tabla dg_consulta de la base de datos de alasMEDIGEN

Después de describir los campos, se procede a ampliar el modelo conceptual, ubicando debajo de cada perspectiva los campos que la componen y de cada indicador la fórmula mediante la cual será calculado. Con este modelo culmina el análisis del almacén de datos.

Como puede apreciarse, el modelo conceptual permite comprender cuáles serán los resultados que se obtendrán, las variables que se utilizarán para analizarlos y la relación que existe entre ellos.

- **Modelo conceptual ampliado**

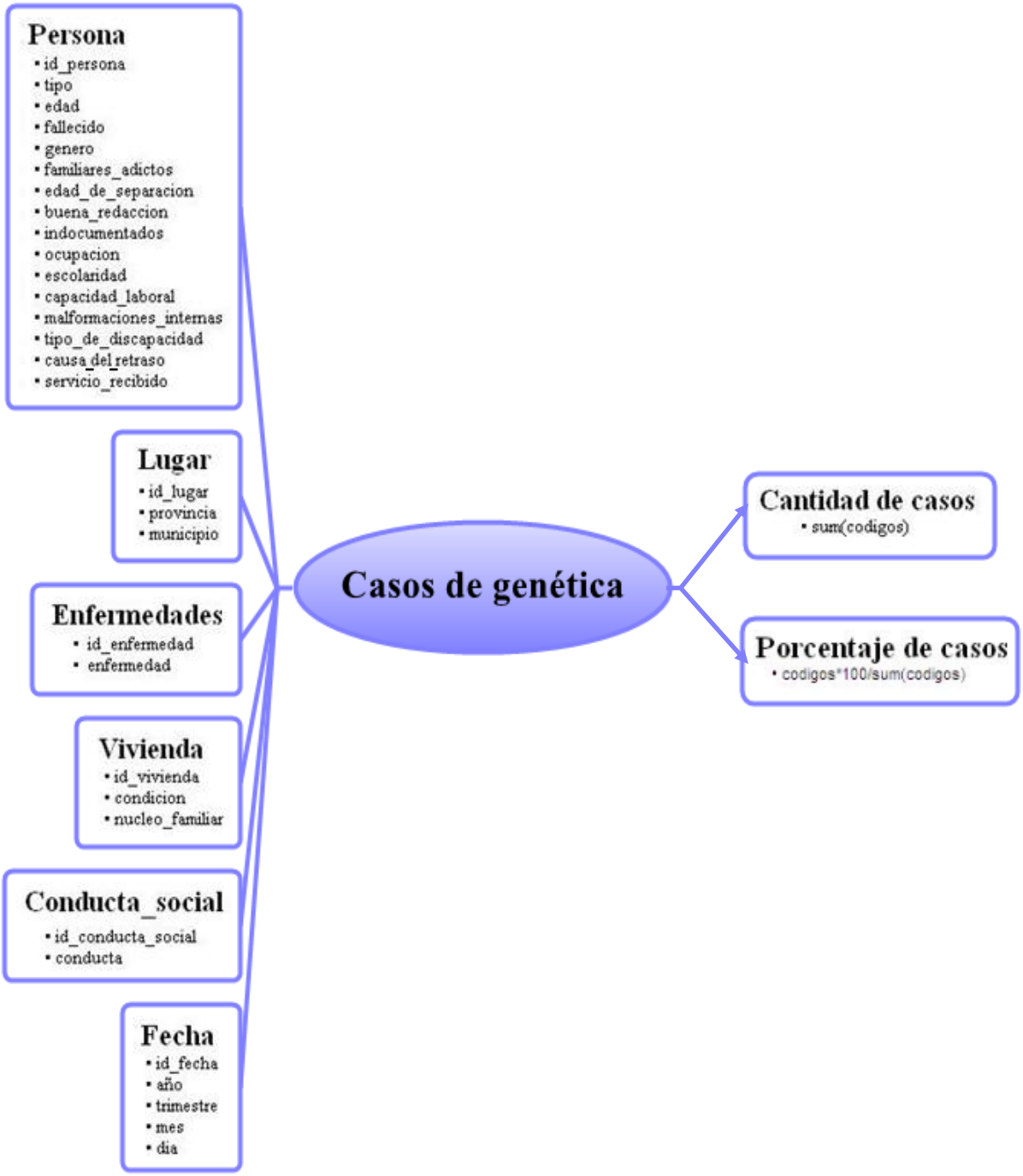

Fig. 39 Modelo conceptual ampliado del proceso Casos de genética.

CONCLUSIONES DEL CAPÍTULO

Con el estudio del sistema alasMEDIGEN, teniendo en cuenta los procesos de las fases y pasos de la metodología seleccionada y atendiendo a los requerimientos del cliente se identificaron los indicadores resultantes de las interrogantes y sus respectivas perspectivas de análisis, mediante las cuales se construyó el modelo conceptual del almacén de datos. Se analizaron los OLTP para señalar las correspondencias con los datos de las fuentes y se seleccionaron los campos de cada perspectiva. Por último, se amplió el modelo conceptual, visualizando los resultados obtenidos.

CAPÍTULO 3. VALIDACIÓN DE LA SOLUCIÓN

Una vez desarrollado el análisis del almacén de datos para la red nacional de Genética Médica, es necesario validar su funcionamiento y comprobar el éxito del mismo. Para cumplir este objetivo se aplica una lista de chequeo, la que constituye un elemento con un elevado nivel de eficiencia que permite validar a través de interrogantes si lo chequeado cumple con un grupo de parámetros. Finalmente, expertos en el tema validarán la solución.

3.1. LISTA DE CHEQUEO

Se entiende por lista de chequeo a una relación de preguntas, en forma de cuestionario que sirve para verificar el grado de cumplimiento de determinadas reglas establecidas con un fin determinado.

Este método utiliza preguntas orientadas a identificar problemas por áreas y sirven para motivar posibles soluciones o la detección de oportunidades de mejora.

3.1.1. ESTRUCTURA DE UNA LISTA DE CHEQUEO

La forma de redactar y confeccionar las Listas de Chequeos es variada.

Uno de los formatos más prácticos y fáciles de usar son aquellos diseñados en forma de cuadro, que permiten un llenado rápido de los distintos casilleros. Se pueden contestar con un Si o un No, o bien tildar los casilleros para los casos en que se verifica la pregunta, dejando el espacio en blanco si no se cumple.

Otra opción de diseño es un listado de preguntas con espacios libres al final, que deben ser respondidas con frases breves y sencillas por parte de aquellos encargados que realizan el control. Están también aquellas que optan por utilizar ambos formatos de manera alternada, colocando en algunas partes casilleros y en otras, espacios libres.

Lo aconsejable es un diseño sencillo, práctico y fácil de visualizar, de manera tal que quien sea el encargado de responderla se familiarice con la lista de manera rápida.

En cuanto al contenido y extensión de las listas, se recomienda redactar un cuestionario breve y fácil de responder. La gran complejidad en algunos casos provoca rechazo de parte de quien debe responderla por el tiempo que invierte hacerlo y comprenderla. La clave del éxito de una lista de chequeo, para su aceptación, es que tenga estas características:

- Fácil de entender.

- Cuestionario breve.
- Buena redacción y caligrafía.

3.2. Lista de chequeo para el análisis del almacén de datos de la Red Nacional de Genética Médica

Esta lista de chequeo fue confeccionada con el fin de identificar errores en el análisis del almacén de datos. Su elaboración está basada en una serie de preguntas enfocadas en los pasos desarrollados para alcanzar la solución.

Para hacer uso de la lista, primeramente se le da un peso al indicador o pregunta a evaluar, seguidamente se evalúa el indicador, marcando con una X la respuesta correcta y emitir si es necesario algún comentario.

Técnicas de Inteligencia de Negocios y Metodología				
Peso	**Indicadores a evaluar**	**Si**	**No**	**Comentarios**
Crítico	¿Atendiendo el problema en cuestión la técnica de Inteligencia de Negocios adecuada es el almacén de datos?			
Crítico	¿La metodología a seguir es la apropiada para realizar el análisis del almacén?			
Paso#1 Metodología Hefesto				
Crítico	¿El proceso obtenido corresponde con las preguntas elaboradas?			
Crítico	¿Al descomponer las preguntas se identificaron correctamente los indicadores y perspectivas?			

Crítico	¿Están correctamente ubicados los indicadores y perspectivas en el modelo conceptual?			
Paso#2 Metodología Hefesto				
Crítico	¿La información requerida por el cliente está disponible en el OLTP?			
Crítico	¿La descripción del cálculo de los indicadores está correcta?			
Crítico	¿Están todos los elementos del modelo conceptual correspondido en el OLTP?			
Crítico	¿Se detallaron los campos que contendrá cada perspectiva?			
Crítico	¿Se amplió el modelo conceptual? ¿Está correcto?			

Una vez definida la lista de chequeo, se procederá a su validación a través de expertos, los cuales especificarán si las interrogantes elaboradas son puntos importantes a medir para el análisis de un almacén de datos.

3.3. VALIDACIÓN DE LA SOLUCIÓN PROPUESTA

Para determinar los expertos que validarán la solución propuesta y la lista de chequeo confeccionada, se tuvo en cuenta:

- Responsabilidad
- Experiencia en el desarrollo de almacenes de datos
- Centro en que laboran

En el proceso de selección se obtuvieron dos expertos profesionales del centro: Unidad de Compatibilización, Integración y Desarrollo de Productos Informáticos para la Defensa (UCID), que han estado vinculados a proyectos productivos y poseen además conocimientos de almacenes de datos.

A continuación se muestran los criterios emitidos por cada experto, que validan la lista de chequeo elaborada y el análisis del almacén de datos para la red nacional de Genética Médica.

- **Validación del experto #1**

Nombre y apellidos: Ing. Leonardo Uria Sánchez.

Proyecto al que pertenece: Herramientas de Sistemas de Apoyo a la Toma de Decisiones.

Rol que desempeña: Arquitecto de datos.

Experiencia en almacenes de datos: Desarrollo del Almacén de Datos del Redmine.

- **Validación de la lista de chequeo**

La lista de chequeo elaborada, especifica los puntos más importantes a tener en cuenta para la realización del análisis de un almacén de datos.

Se enfoca principalmente en los pasos a seguir por la metodología elegida. Las preguntas están bien definidas y se aprecia concordancia en ellas. Establecen una secuencia de pasos a partir de la metodología.

- **Validación del análisis del almacén de datos para la red nacional de Genética Médica, teniendo en cuenta la lista de chequeo**

Analizando el problema planteado la tecnología de Inteligencia de negocios apropiada es un almacén de datos, porque constituye una fuente de información muy potente para ser utilizadas como base de conocimiento de cualquier sistema de análisis de información. Para la construcción de dicho almacén de datos se necesita de una rápida metodología que permita alcanzar el objetivo en el menor tiempo posible, la elegida cumple con dichas necesidades pues Hefesto se caracteriza por su rapidez y eficacia.

El análisis del almacén cumple con los requerimientos de la metodología. Las tesistas obtuvieron un proceso llamado Casos de genética que corresponde íntegramente con las preguntas elaboradas. Establecieron correctamente los indicadores y perspectivas de análisis y a partir de estos, construyeron el modelo conceptual donde tuvieron en cuenta el estándar a seguir para graficar este tipo de modelo. En este paso se puede comprender cuáles serán los resultados que obtendrán y cuáles serán las variables que utilizarán para analizarlos.

Los indicadores fueron descritos explícitamente. Expusieron la correspondencia entre los elementos del modelo conceptual y el OLTP, por cada perspectiva analizaron las tablas con las cuales se relacionan. Un punto muy importante es detallar que significan los campos disponibles para cada perspectiva, pues a través de estos se manejan los indicadores. Los nombres de los campos son fáciles de deducir y bastante sugerentes, detallaron cada uno de estos de manera explícita. Por último, construyeron el modelo conceptual ampliado correctamente, donde colocaron bajo cada perspectiva, los campos seleccionados.

Como conclusión, diría que el análisis expuesto en el trabajo abordado llega a un nivel complejo dentro del grado de especificidad, ya que las perspectivas muestran una buena proyección informativa capaz de soportar lo requerido por el cliente, por lo que cabe destacar como satisfactorio la captura de requisitos y la modelación de la problemática (Ver Anexo 1).

- **Validación del experto #2**

Nombre y apellidos: Ing. Dailen Ramón Zequeira.

Proyecto al que pertenece: Herramientas de Sistemas de Apoyo a la Toma de Decisiones.

Rol que desempeña: Arquitecto de datos.

Experiencia en almacenes de datos: Desarrollo del Almacén de Datos de Fiscalía Militar. Desarrollo del Almacén de Datos de Potencial Humano para el Comité Militar.

- **Validación de la lista de chequeo**

La lista de chequeo está bien elaborada manteniendo una buena ortografía, recoge los puntos necesarios para realizar el análisis de un Almacén de Datos según lo propuesto en la metodología de Hefesto.

Las preguntas realizadas tienen concordancia y están redactadas según los pasos que propone la metodología utilizada, dándoles un orden lógico a las mismas.

- **Validación del análisis del almacén de datos para la red nacional de Genética Médica, teniendo en cuenta la lista de chequeo**

Luego de analizar el problema que presenta la red nacional de Genética Médica, estoy de acuerdo en que la tecnología de Inteligencia de Negocios más adecuada es el Almacén de Datos porque permite transformar los datos almacenados a través del tiempo, en conocimiento necesario para una buena toma de decisiones, constituyendo así una fuente de conocimiento potente para cualquier Sistema de Información. Una estupenda metodología para desarrollar un Almacén de Datos es Hefesto pues es rápida, sencilla y de buen entendimiento, permitiendo la construcción del mismo en un corto tiempo y con resultados eficaces.

Luego del análisis del negocio y la entrevista con el cliente se identificó un proceso llamado Casos de genética el cual corresponde completamente con las preguntas identificadas y el negocio en cuestión. Las perspectivas y los indicadores fueron identificados correctamente. En la construcción del modelo conceptual se respetó el diseño propuesto, representando correctamente cada perspectiva e indicador, sirviendo este modelo como artefacto para mostrar al cliente.

La información que se requiere según las necesidades del cliente para el Almacén de Datos, está disponible en el OLTP correspondiéndose cada elemento del Modelo Conceptual con las tablas de los OLTP. Se describió correctamente el cálculo de los indicadores propuestos. Es muy importante detallar los campos que contendrán las perspectivas para un mejor entendimiento. En este caso los campos fueron detallados correctamente, a pesar de que estos tienen nombres que pueden deducirse con facilidad. Luego se amplió el Modelo Conceptual escribiendo bajo cada perspectiva los campos que contendrán las mismas y bajo cada indicador la fórmula de cálculo correspondiente.

El análisis del Almacén de Datos desarrollado por las tesistas ha sido satisfactorio, han realizado un correcto análisis según la metodología Hefesto, seleccionaron correctamente las perspectivas e indicadores según la problemática existente, cumpliendo con las necesidades que el cliente planteó en las entrevistas. En general el trabajo ha sido excelente, mostrando alto grado de complejidad y sentando las bases para un correcto diseño e implementación (Ver Anexo 2).

CONCLUSIONES DEL CAPÍTULO

En este capítulo se expuso la validación del análisis del almacén de datos. Expertos en el tema dieron sus puntos de vistas, teniendo en cuenta los aspectos y preguntas elaboradas en la lista de chequeo, la cual fue validada. Luego de haber emitido sus criterios se llega a la conclusión que se ha conseguido un correcto análisis del almacén de datos para la red nacional de Genética Médica.

CONCLUSIONES GENERALES

A partir de los resultados obtenidos se puede concluir que:

- En la primera fase de la metodología aplicada se determinaron 22 preguntas que abarcan los requerimientos más importantes, a partir del intercambio con los especialistas en genética, lo que permitió identificar 6 perspectivas y 2 indicadores.
- Se elaboró el modelo conceptual del almacén de datos que permite visualizar la relación entre las perspectivas y los indicadores.
- Se analizó el OLTP identificado, sistema alasMEDIGEN, para señalar las correspondencias entre los campos de cada perspectiva y las tablas del OLTP.
- Expertos en el tema validaron la solución mediante una lista de chequeo, por lo que se logró un correcto análisis de un almacén de datos para la red nacional de Genética Médica.

RECOMENDACIONES

Los objetivos generales de este trabajo han sido logrados, pero a lo largo de su desarrollo, han ido surgiendo ideas que podrían considerarse en un futuro, para lo cual se recomienda:

- Realizar el diseño e implementación del almacén de datos para la red nacional de Genética Médica.
- Permitir el acceso al documento de la presente investigación como material de consulta para trabajos similares.

REFERENCIAS BIBLIOGRÁFICAS

[1]. **Dra. Norma Elena de León Ojeda.** Genética. *Infomed.* [En línea] 1999. http://www.sld.cu/sitios/genetica

[2]. **Ortiz, Marta Cecilia.** La inteligencia de negocios aplicada a las organizaciones en Latinoamérica. Medellín, Colombia : s.n., 2007.

[3]. **Bernabeu, Ing. Ricardo Dario.** DATA WAREHOUSING: Investigación y Sistematización de Conceptos. Córdoba, Argentina : s.n., 2007.

[4]. Sinnexus Business Intelligence + Informática estratégica. [En línea] 2007. http://www.sinnexus.com/business_intelligence.

[5]. **Instituto Nacional de Estadísticas e Informática(INEI).** Manual para la construcción de un data warehouse. *Conceptos y estrategias de desarrollo.* Perú, Lima : s.n., 1997.

[6]. **CERVANTES, ING. LEOPOLDO ARTURO.** GUÍA PARA OBTENER EL RETORNO A LA INVERSIÓN EN PROYECTOS DE DATA WAREHOUSE. CAMPUS MONTERREY : s.n., DICIEMBRE, 2001.

[7]. **Inmon, W. H.** *Building the Data Warehouse.* New York : John Wiley & Sons, 2002.

[8]. **Kimball, Ralph.** *The Data Warehouse Toolkit: the complete guide to dimensional modeling.* New York : John Wiley & Sons, 2002.

[9]. **Elizabeth Vitt, Michael Luckevich, Stacia Misner.** *Business Intelligence. Ténicas de análisis para la toma de decisiones estratégicas.* España : McGraw- Hill, 2003.

[10]. DataPrix . *La metodología CRISP-DM.* [En línea] http://www.dataprix.com/en/la-metodolog%C3%AD-crisp-dm.

[11]. **Huamantumba, Rayner**. Manual para diseño y desarrollo de Datamart. DATAMART PASO A PASO. 2007.

[12]. **Milla, Ing. Roberto Espinosa.** El Rincon del BI. *Fases en la implantación de un sistema DW. Metodologia para la construcción de un DW.* [En línea] diciembre de 2009. http://churriwifi.wordpress.com/2009/12/05/5-fases-en-la-implantacion-de-un-sistema-dw-metodologia-para-la-construccion-de-un-dw/.

[13]. **Ing. Bernabeu, Ricardo Dario.** HEFESTO: Metodología propia para la Construcción de un Data Warehouse. Córdoba, Argentina : s.n., Noviembre,2007.

BIBLIOGRAFÍA

1. **Dra. Norma Elena de León Ojeda.** Genética. *Infomed.* [En línea] 1999. http://www.sld.cu/sitios/genetica

2. **Ortiz, Marta Cecilia.** La inteligencia de negocios aplicada a las organizaciones en Latinoamérica. Medellín, Colombia : s.n., 2007.

3. **Bernabeu, Ing. Ricardo Dario.** DATA WAREHOUSING: Investigación y Sistematización de Conceptos. Córdoba, Argentina : s.n., 2007.

4. Sinnexus Business Intelligence + Informática estratégica. [En línea] 2007. http://www.sinnexus.com/business_intelligence.

5. **Instituto Nacional de Estadísticas e Informática(INEI).** Manual para la construcción de un data warehouse. *Conceptos y estrategias de desarrollo.* Perú, Lima : s.n., 1997.

6. **CERVANTES, ING. LEOPOLDO ARTURO.** GUÍA PARA OBTENER EL RETORNO A LA INVERSIÓN EN PROYECTOS DE DATA WAREHOUSE. CAMPUS MONTERREY : s.n., DICIEMBRE, 2001.

7. **Inmon, W. H.** *Building the Data Warehouse.* New York : John Wiley & Sons, 2002.

8. **Kimball, Ralph.** *The Data Warehouse Toolkit: the complete guide to dimensional modeling.* New York : John Wiley & Sons, 2002.

9. **Elizabeth Vitt, Michael Luckevich, Stacia Misner.** *Business Intelligence. Ténicas de análisis para la toma de decisiones estratégicas.* España : McGraw- Hill, 2003.

10. DataPrix . *La metodología CRISP-DM.* [En línea] http://www.dataprix.com/en/la-metodolog%C3%AD-crisp-dm.

11. **Huamantumba, Rayner.** Manual para diseño y desarrollo de Datamart. *DATAMART PASO A PASO.* 2007.

12. **Milla, Ing. Roberto Espinosa.** El Rincon del BI. *Fases en la implantación de un sistema DW. Metodologia para la construcción de un DW.* [En línea] diciembre de 2009. http://churriwifi.wordpress.com/2009/12/05/5-fases-en-la-implantacion-de-un-sistema-dw-metodologia-para-la-construccion-de-un-dw/.

13. **Ing. Bernabeu, Ricardo Dario.** HEFESTO: Metodología propia para la Construcción de un Data Warehouse. Córdoba, Argentina : s.n., Noviembre,2007.

14. Data Warehouse para la gestión por procesos en el sistema productivo. Universidad Politécnica de Valencia : s.n., Mayo, 2004.

15. **Pete Chapman, Julian Clinton.** Guía paso a paso de Minería de Datos. *CRISP-DM 1.0.* 1999, 2000. 15.

16. **MSc. Emma R. Rizo Rizo, Dr. Juan Pedro Febles.** Importancia de la utilización de un Data Warehouse (DW) en las empresas. Cuba : s.n.

17. **Sánchez, Leopoldo Zenaido Zepeda.** Metodología para el Diseño Conceptual de Almacenes de Datos. UNIVERSIDAD POLITÉCNICA DE VALENCIA : s.n., 2008.

18. **Vaisman, Alejandro.** La Investigación en OLAP y Data Warehousing: Pasado, Presente y Futuro. Universidad de Buenos Aires : s.n., 2006.

19. **Ruiz, Mario Ramón Mancera.** LISTA DE CHEQUEO.

ANEXOS

Anexo 1. Validación del experto Leonardo Uria Sánchez

Nombre y apellidos: Ing. Leonardo Uria Sánchez.

Proyecto al que pertenece: Herramientas de Sistemas de Apoyo a la Toma de Decisiones.

Rol que desempeña: Arquitecto de datos.

Experiencia en almacenes de datos: Desarrollo del Almacén de Datos del Redmine.

Centro en que labora: UCID

- **Validación de la lista de chequeo**

La lista de chequeo elaborada, especifica los puntos más importantes a tener en cuenta para la realización del análisis de un almacén de datos.

Se enfoca principalmente en los pasos a seguir por la metodología elegida. Las preguntas están bien definidas y se aprecia concordancia en ellas. Establecen una secuencia de pasos a partir de la metodología.

- **Validación del análisis del almacén de datos para la red nacional de Genética Médica, teniendo en cuenta la lista de chequeo**

Analizando el problema planteado la tecnología de Inteligencia de negocios apropiada es un almacén de datos, porque constituye una fuente de información muy potente para ser utilizadas como base de conocimiento de cualquier sistema de análisis de información. Para la construcción de dicho almacén de datos se necesita de una rápida metodología que permita alcanzar el objetivo en el menor tiempo posible, la elegida cumple con dichas necesidades pues Hefesto se caracteriza por su rapidez y eficacia.

El análisis del almacén cumple con los requerimientos de la metodología. Las tesistas obtuvieron un proceso llamado Casos de genética que corresponde íntegramente con las preguntas elaboradas. Establecieron correctamente los indicadores y perspectivas de análisis y a partir de estos, construyeron el modelo conceptual donde tuvieron en cuenta el estándar a seguir para graficar este tipo de modelo. En este paso se puede comprender cuáles serán los resultados que obtendrán y cuáles serán las variables que utilizarán para analizarlos.

Los indicadores fueron descritos explícitamente. Expusieron la correspondencia entre los elementos del modelo conceptual y el OLTP, por cada perspectiva analizaron las tablas con las cuales se relacionan. Un punto muy importante es detallar que significan los campos disponibles para cada perspectiva, pues a través de estos se manejan los indicadores. Los nombres de los campos son fáciles de deducir y bastante sugerentes, detallaron cada uno de estos de manera explícita. Por último, construyeron el modelo conceptual ampliado correctamente, donde colocaron bajo cada perspectiva, los campos seleccionados.

Como conclusión, diría que el análisis expuesto en el trabajo abordado llega a un nivel complejo dentro del grado de especificidad, ya que las perspectivas muestran una buena proyección informativa capaz de soportar lo requerido por el cliente, por lo que cabe destacar como satisfactorio la captura de requisitos y la modelación de la problemática.

Firma del experto

Anexo 2. Validación del experto Dailen Ramón Zequeira

Nombre y apellidos: Ing. Dailen Ramón Zequeira.

Proyecto al que pertenece: Herramientas de Sistemas de Apoyo a la Toma de Decisiones.

Rol que desempeña: Arquitecto de datos.

Experiencia en almacenes de datos: Desarrollo del Almacén de Datos de Fiscalía Militar. Desarrollo del Almacén de Datos de Potencial Humano para los Comité Militar.

Centro en que labora: UCID

- **Validación de la lista de chequeo**

La lista de chequeo está bien elaborada manteniendo una buena ortografía, recoge los puntos necesarios para realizar el análisis de un Almacén de Datos según lo propuesto en la metodología de Hefesto.

Las preguntas realizadas tienen concordancia y están redactadas según los pasos que propone la metodología utilizada, dándoles un orden lógico a las mismas.

- **Validación del análisis del almacén de datos para la red nacional de Genética Médica, teniendo en cuenta la lista de chequeo**

Luego de analizar el problema que presenta la red nacional de Genética Médica, estoy de acuerdo en que la tecnología de Inteligencia de Negocios más adecuada es el Almacén de Datos porque permite transformar los datos almacenados a través del tiempo, en conocimiento necesario para una buena toma de decisiones, constituyendo así una fuente de conocimiento potente para cualquier Sistema de Información. Una estupenda metodología para desarrollar un Almacén de Datos es Hefesto pues es rápida, sencilla y de buen entendimiento, permitiendo la construcción del mismo en un corto tiempo y con resultados eficaces.

Luego del análisis del negocio y la entrevista con el cliente se identificó un proceso llamado Casos de genética el cual corresponde completamente con las preguntas identificadas y el negocio en cuestión. Las perspectivas y los indicadores fueron identificados correctamente. En la construcción del modelo conceptual se respetó el diseño propuesto, representando correctamente cada perspectiva e indicador, sirviendo este modelo como artefacto para mostrar al cliente.

La información que se requiere según las necesidades del cliente para el Almacén de Datos, está disponible en el OLTP correspondiéndose cada elemento del Modelo Conceptual con las tablas de los OLTP. Se describió correctamente el cálculo de los indicadores propuestos. Es muy importante detallar los campos que contendrán las perspectivas para un mejor entendimiento. En este caso los campos fueron detallados correctamente, a pesar de que estos tienen nombres que pueden deducirse con facilidad. Luego se amplió el Modelo Conceptual escribiendo bajo cada perspectiva los campos que contendrán las mismas y bajo cada indicador la fórmula de cálculo correspondiente.

El análisis del Almacén de Datos desarrollado por las tesistas ha sido satisfactorio, han realizado un correcto análisis según la metodología Hefesto, seleccionaron correctamente las perspectivas e indicadores según la problemática existente, cumpliendo con las necesidades que el cliente planteó en las entrevistas. En general el trabajo ha sido excelente, mostrando alto grado de complejidad y sentando las bases para un correcto diseño e implementación.

Firma del experto

GLOSARIO DE TÉRMINOS

CNGM: Centro nacional de Genética Médica.

BI: Inteligencia de negocios (del inglés de Bussines Intelligence).

Datamining: Minería de datos

OLAP: Procesamiento analítico en línea (del inglés de Online Analitical Processing).

OLTP: Procesamiento de Transacciones En Línea (del inglés de On-Line Transaction Processing).

Datamart: es una versión especial de almacén de datos. Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones.

DSS: Sistemas de Soporte a Decisiones.

ETL: Extracción, Transformación y Carga de Datos.

UCI: Universidad de las Ciencias Informáticas