

**Universidad de las Ciencias Informáticas
Facultad 7**



**Título: Diseño de un Almacén de Datos para
apoyar la toma de decisiones en el
Centro de Informática Médica**

**Trabajo de Diploma para optar por el título
de Ingeniero Informático**

Autores: Libisdey Pedraza Fernández
Alejandro Ramos Hernández

Tutor: Ing. David Barreto Medina

Ciudad de La Habana, 2010

“Año 52 de la Revolución”



"Muchas veces la gente no sabe lo que quiere hasta que se lo enseñas."

Steve Jobs

Dedicatoria

Dedico este trabajo:

A mis padres, por darme la vida, por su cariño y su ejemplo, por estar siempre presentes, por darme todo su apoyo; sin ellos este camino hubiese sido muy difícil. A mi familia por su ayuda y preocupación.

Los quiero mucho.

Alejandro Ramos Hernández.

Dedicatoria

Dedico este trabajo a mis padres que son mi razón de ser. A mi mamá, que sin ella no sería nadie. Gracias a su cuidado y dedicación me he convertido en la persona que soy ahora. Es, además de mi mamá, mi mejor amiga.

A mi papá, que me apoyó cuando más lo necesitaba, me brindó las fuerzas necesarias para seguir adelante con mi carrera, a pesar de las dificultades, y me enseñó que rendirse no es una opción cuando quieres llegar a ser alguien en la vida.

A mis hermanitos que son una parte importante en mi vida.

Gracias por su cariño. Los quiero mucho.

Libisdey Pedraza Fernández

Agradecimientos

Siempre que se termina un trabajo de cualquier índole hay terceras personas que influyen de cierta manera. En este caso por ser la culminación de mis estudios universitarios me siento en la obligación y el deber de expresar mis agradecimientos, donde las palabras más rebuscadas no se acercan para compensar la ayuda brindada.

A mis padres por su apoyo, por su ayuda, por su perseverancia ante todo, por reír conmigo ante cada nota satisfactoria y por sentir dolor ante cada situación inesperada, no sólo durante estos 5 años de la carrera sino durante mis 23 años de vida, a ellos el mayor de los agradecimientos.

A mi familia en sentido general, por su preocupación ante mis resultados académicos.

A mi tutor por dedicarme cada segundo cada vez que lo necesité.

A mi compañera de tesis Libisdey por su comprensión y por su paciencia durante todo el tiempo de desarrollo del trabajo.

Agradecimientos

A todos los profesores de la carrera que durante 5 años contribuyeron a mi formación como ingeniero.

A Maylin Campos Núñez que durante mucho tiempo estuvo a mi lado y me ayudó a salir adelante en varios momentos de mis estudios.

A mis compañeros del aula por aclararme cada duda ante cada prueba parcial o final, por estar siempre presente, son como hermanos para mí.

A mis compañeros del apartamento por compartir cada momento de la carrera.

También tengo que agradecer a otros estudiantes y profesores de las facultades 2 y 6 que consultamos en cierto momento de la investigación y que su ayuda fue de vital importancia para la realización de la tesis.

Alejandro Ramos Hernández

Agradecimientos

Agradezco:

A mis padres por apoyarme en todo momento, y darme los ánimos necesarios para seguir adelante. Por ser unos padres tan dedicados y cariñosos conmigo. Sin ellos mi carrera no hubiese significado lo mismo, ni tendría el sentido que tiene hoy.

A mis tres hermanitos por hacerme reír cuando estaba triste, por ser niños con maravillosos sentimientos y por querer tanto a su hermana mayor. Especialmente a mi hermano Jaimito, que no ha dejado de preocuparse por mí, ni de mostrarme lo mucho que me adora en estos cinco años. Y a los pequeñines: Davisito y Alejandro, que aunque no viven conmigo, cada vez que me ven me abrazan y me besan como si no hubiese estado ausente mucho tiempo.

A mi novio Machín por ser mi compañero y amigo en estos tres últimos años de mi vida en la universidad. Por apoyarme cuando más me hizo falta, y por darme tanto amor. Gracias por ser como es, una bella persona y un novio cariñoso. Le dedico un agradecimiento especial por ser mi amor y también una parte importante en mi vida.

A toda mi familia que siempre me han apoyado, en especial a mis tías y mi querida prima Lisbey.

Agradecimientos

A mi tutor David por ayudarnos cuando lo necesitamos, por ser una persona alegre y aunque hace poco que lo conocemos, nos hemos llevado muy bien.

A mi compañero de tesis Alejandro que me ha apoyado en todo, y ha sido un compañero trabajador y entusiasta en todo momento. Gracias por tu ayuda.

A los profesores que de una forma u otra nos han ayudado en la realización de la tesis. Entre ellos: la profesora Yaima por estar siempre disponible para ayudarnos y por su hospitalidad cada vez que íbamos a molestarla. A la profesora Pura que también nos ha ayudado en la revisión del trabajo, gracias por atendernos y tratarnos tan bien.

A todos mis amigos por acompañarme en estos cinco años, donde hemos disfrutado muchos momentos que guardaré siempre como algo preciado.

A Yusita, Lisdany y Yanet que compartimos buenos y malos momentos desde primer año. A mis compañeras de apartamento que se han portado muy bien conmigo, y con los que espero compartir algunos momentos más antes de despedirnos.

A mis amigos de otros años, que aunque no los veo mucho, siempre los recuerdo con cariño.

Libisdey Pedraza Fernández

Declaración de Autoría

Declaración de Autoría

Declaramos que somos los únicos autores de este trabajo y autorizamos al Centro de Informática Médica (CESIM) de la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

Para que así conste firmo la presente a los 23 días del mes de junio del año 2010.

Libisdey Pedraza Fernández

Autora

Alejandro Ramos Hernández

Autor

Ing. David Barreto Medina

Tutor

Datos de Contacto

Datos de Contacto

Síntesis del tutor: Ing. David Barreto Medina.

Profesor graduado de Ingeniería Industrial en el año 2004 en la Universidad de Holguín "Oscar Lucero Moya", posee la categoría docente Instructor. Ha impartido las asignaturas: Administración de Empresas, Contabilidad, Metodología de la Investigación Científica y Comercio Electrónico. Se desempeñó como Asesor de la asignatura Administración de Empresas a nivel de universidad en el curso 2008-2009, Jefe de Colectivo de Ciencias Empresariales en la Facultad 7 en el curso 2008-2009 y Asesor Estratégico para la producción en el curso 2008-2009. Actualmente es el Asesor de Planificación y Control del Centro de Informática Médica (CESIM).

Correo electrónico: dbarreto@uci.cu

Resumen

Resumen

El Centro de Informática Médica (CESIM) debe tomar decisiones orientadas a satisfacer la demanda de servicios, soluciones y productos de alta calidad, por lo que es muy importante mejorar los sistemas de información ligados a los procesos de decisión. El presente Trabajo de Diploma propone el diseño de un Almacén de Datos que servirá de apoyo al proceso de toma de decisiones del CESIM.

La suite de Inteligencia de Negocios proporcionada por Pentaho es la herramienta que se usa para realizar el diseño del Almacén de Datos y para crear los reportes a partir de los indicadores que se calcularon. Es una herramienta libre y completa, lo que garantiza que el CESIM no tenga que destinar costos adicionales por licencias de software.

Un diseño de este tipo permitirá darle seguimiento a los objetivos estratégicos a partir de indicadores. También proporcionará conocer con más precisión los eventos que ocurren en los departamentos de producción, así como mantener un control eficaz en el Grupo de Mercadotecnia. Además, es de vital importancia mantener a los clientes informados de los nuevos productos y servicios que se ofrecen en el Centro y mantener un control sobre la producción.

Palabras Claves

Almacén de Datos

Inteligencia de Negocios

Proceso de toma de decisiones

Índice

Índice

Introducción.....	1
Capítulo 1. Fundamentación Teórica.....	6
1.1. Una tecnología para la toma de decisiones.....	6
1.1.1. Características de un almacén de datos	7
1.2. Almacén de Datos y Base de Datos Corporativa	7
1.3. Conceptos fundamentales.....	8
1.3.1. Datamart	8
1.3.2. Estructura Multidimensional.....	10
1.3.3. Minería de Datos	11
1.4. Estructura de un almacén de datos	11
1.5. Arquitectura de un almacén de datos.....	12
1.5.1. Elementos de una arquitectura para el Almacén de Datos	12
1.6. Tecnologías y Herramientas	15
1.6.1. Sistemas de Gestión de bases de datos	15
1.6.2. Modelado de la base de datos. Visual Paradigm	17
1.6.3. Gestión de Relación con Clientes (CRM en inglés).....	18
1.6.4. Herramientas Pentaho.....	20
1.6.5. Herramienta Redmine.....	22
1.7. Metodología propia para la Construcción de un almacén de datos	23
1.7.1. Metodología Hefesto. Características	23
1.7.2. Pasos para la aplicación de la metodología.....	24
Capítulo 2. Análisis del Almacén de Datos.....	27
2.1. Análisis de los procesos en los departamentos de interés	27
2.2. Estructura del Almacén de Datos	27
2.3. Análisis de requerimientos	29

Índice

2.4. Análisis de los indicadores que se necesitan calcular.....	33
2.4.1. Nivel de granularidad.....	38
2.5. Modelo lógico del Almacén de Datos del CESIM	39
2.5.1. Tipo de modelo lógico del almacén de datos.....	40
2.5.2. Tablas de dimensiones.....	40
2.5.3. Tablas de hechos	41
Capítulo 3. Diseño del Almacén de Datos	45
3.1. Arquitectura para el CESIM.....	45
3.2. Diseño de los procesos ETL	46
3.2.1. Diagrama de actividades para explicar el proceso ETL	48
3.2. Creación de Cubos Multidimensionales.....	48
3.3.1. Publicación de los cubos en Pentaho.	49
Conclusiones	50
Recomendaciones	51
Referencias Bibliográficas	52
Bibliografía	56
Anexos.....	60

Introducción

Introducción

Las Tecnologías de la Información y las Comunicaciones (**TIC**) han aumentado el nivel de vida en el mundo actual, debido a la importancia que se le otorga como precursor de la información. Estas tecnologías han influido de manera decisiva en el ámbito educativo convirtiéndose en herramientas de trabajo que facilitan el acceso a una excelente y variada información.

La expansión de las TIC en todos los ámbitos de la sociedad se han producido a gran velocidad, y es un proceso que continúa ya que nuevos elementos tecnológicos surgen sin cesar. Es de vital importancia este desarrollo para las empresas ya que, la información se ha colocado en un lugar primordial como uno de los principales recursos que estas poseen actualmente.

Las personas que se encargan de la toma de decisiones han comenzado a comprender que la información no es sólo un subproducto de la conducción empresarial. Aquella a la vez alimenta a los negocios y puede ser uno de los tantos factores críticos para la determinación del éxito o fracaso de las empresas.

Los Sistemas de Información (SI) y las Tecnologías de la Información han cambiado la forma en que operan las organizaciones actuales. A través de su uso se logran importantes mejoras, pues automatizan los procesos operativos y suministran una plataforma de información necesaria para la toma de decisiones.

Un Sistema de Información es un conjunto de elementos que interactúan entre sí con el fin de apoyar las actividades de una empresa o negocio. Estos sistemas logran con su implantación ventajas competitivas y la reducción del avance de los competidores. (1)

Los Sistemas de Soporte a la Decisión (DSS por sus siglas en inglés) pueden considerarse como una tercera generación de Sistemas de Información, cuyo objetivo es intentar descubrir qué pasaría si se toma una serie de decisiones, o proporcionar automáticamente las decisiones o sugerencias que asistan al administrador. (2)

Introducción

Todos los conceptos antes expuestos forman una parte muy importante en la Inteligencia Empresarial (**IE**), pues sin los DSS las empresas enfrentarían dificultades si trataran de organizar o mantener toda la información. La IE constituye la base para ayudar a los empleados de todos los niveles de la organización, a obtener la información necesaria para tomar rápidamente mejores decisiones que impulsen el negocio.

En la actualidad, la Inteligencia Empresarial ya no es exclusiva del ámbito de los ejecutivos y analistas; todos dentro de la compañía pueden tener acceso a información crítica de la empresa. El acceso en tiempo a esta información puede dotar a sus trabajadores de los medios para tomar decisiones en tiempo real y emprender acciones con la información necesaria, que reflejen y alcancen los objetivos de la empresa.

El término IE se refiere al uso de datos en una empresa para facilitar la toma de decisiones; abarca la comprensión del funcionamiento actual de la empresa, así como la anticipación de acontecimientos futuros, con el objetivo de ofrecer conocimientos para respaldar las decisiones empresariales. (3)

En muchas empresas se pone en práctica este proceso, pues por motivos de seguridad o comodidad necesitan de una buena organización y análisis de todo tipo de documentos. Asimismo, a veces es necesario controlar información de vital importancia para el buen desarrollo y evolución de la organización. Para esto en la actualidad existen sistemas a través de los cuales se almacenan todo tipo de datos y hacen más factible el manejo de estos, como es el caso de los almacenes de datos.

Un **Almacén de Datos** (del inglés Data Warehouse) es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), que ayuda a la toma de decisiones en la entidad en la que se utiliza. Se trata de un expediente completo de una organización, almacenada en una base de datos diseñada para favorecer el análisis y la divulgación eficiente de datos. (4)

A pesar de ser un concepto novedoso las empresas han comenzado a utilizar los almacenes de datos, ya que permiten extraer conocimientos de la información histórica almacenada en la organización.

Introducción

El país no está ajeno a los nuevos avances y posee instituciones que desarrollan almacenes de datos con el fin de mantener la información controlada y centralizada. Un ejemplo de esto es el trabajo que se realiza en la Universidad de las Ciencias Informáticas (UCI), la cual está inmersa en una serie de cambios que se rigen por las nuevas tecnologías y la repercusión que tienen estas en el desarrollo de los estudiantes como futuros profesionales.

La UCI es un centro de estudios universitarios nacido como un proyecto de la Revolución Cubana, denominado al principio "Proyecto Futuro", el cual tiene entre sus objetivos: informatizar el país y desarrollar la industria del Software para contribuir al desarrollo económico del mismo. Es la primera universidad cubana creada bajo el calor de la Batalla de Ideas.

En la universidad hay actualmente varios centros especializados en diferentes ramas de la ciencia. Entre sus funciones está la de construir software con fines específicos. Uno de estos centros es el Centro de Informática Médica (CESIM), el cual está dedicado al desarrollo de productos, sistemas y soluciones de alta calidad y competitividad para la optimización del trabajo en el sector de la salud.

Este centro cuenta con varias áreas de trabajo que manejan gran cantidad de información descentralizada, lo que afecta el avance del trabajo que se debe realizar para tomar una decisión en cuanto a los resultados esperados. El CESIM en la actualidad no cuenta con un mecanismo eficiente que permita darle seguimiento a los objetivos estratégicos de manera que: apoye la toma de decisiones a todos los niveles, mejore las relaciones con los clientes y proveedores y mantenga un control efectivo de la actividad productiva.

Esto provoca que los jefes de proyecto, jefes de departamento y directivos del centro tengan dificultades para medir los indicadores de producción y exportación y poder trazar acciones para mejorar el desempeño organizacional.

Dada la situación, se plantea el siguiente **problema a resolver**: ¿Cómo facilitar la toma de decisiones del Centro de Informática Médica (CESIM) en el Grupo de Mercadotecnia y los departamentos de producción?

El **objeto de estudio** consiste en el proceso de gestión de la información para la toma de decisiones del Centro de Informática Médica (CESIM). El **campo de acción** define la gestión

Introducción

de la información para la toma de decisiones relacionada con el Grupo de Mercadotecnia y los departamentos de producción.

El **objetivo de la investigación** es diseñar un Almacén de Datos que facilite la toma de decisiones en el Grupo de Mercadotecnia y en los departamentos de producción.

Para lograr el cumplimiento del objetivo se plantean las siguientes **tareas**:

1. Analizar el estudio del arte sobre los almacenes de datos como herramientas para la toma de decisiones.
2. Definir la metodología para diseñar el Almacén de Datos.
3. Evaluar las tecnologías y herramientas existentes que puedan ser usadas en el trabajo.
4. Realizar el levantamiento de procesos del Grupo de Mercadotecnia y los departamentos de producción.
5. Medir el correcto desempeño de los departamentos de producción y el Grupo de Mercadotecnia a partir de indicadores propuestos.
6. Elaborar la documentación correspondiente a los flujos de trabajo propuestos por la metodología seleccionada.
7. Diseñar el Almacén de Datos.

Se propone la realización del diseño de un Almacén de Datos para el CESIM, ya que es de gran utilidad a la hora de tomar decisiones importantes del negocio. Esto se convierte en un paso de avance para la implementación del almacén en futuras investigaciones.

El trabajo está estructurado de la siguiente forma:

Capítulo 1: “Fundamentación Teórica”

En este capítulo se realiza un estudio de los Sistemas de Soporte a la Decisión existentes en el mundo y en el país. Se hace una profunda investigación sobre los almacenes de datos, y su importancia en el mundo actual. Se definen la metodología y las herramientas a utilizar.

Capítulo 2: “Análisis del Almacén de Datos”

Introducción

En este capítulo se plantea una propuesta de solución para el CESIM. Se investigan sus necesidades y perspectivas en cuanto al negocio en general. Se llevan a cabo los pasos que propone la metodología Hefesto para el diseño del Almacén de Datos.

Capítulo 3: “Diseño del Almacén de Datos”

En este capítulo se define la arquitectura del Almacén de Datos del CESIM y se realiza el diseño de los procesos ETL (Extracción, Transformación y Carga), que son de gran importancia para su futura implementación.

Capítulo 1. Fundamentación Teórica

En este capítulo se hace un estudio de los conceptos fundamentales que conforman un almacén de datos, su importancia en el mundo, así como sus ventajas y desventajas. Se analizan las herramientas relacionadas que pueden ser usadas durante el diseño del mismo, además de la metodología a utilizar para cumplir con el buen funcionamiento y cumplimiento de los objetivos trazados.

1.1. Una tecnología para la toma de decisiones

El Almacén de Datos surgió con el objetivo de hacer consultable la información que se tiene de una empresa tanto de meses como de años anteriores.

“Un almacén de datos se crea al extraer datos desde una o más bases de datos de aplicaciones operacionales. La data extraída es transformada para eliminar inconsistencias y resumir si es necesario y luego ser cargadas en el almacén. El proceso de transformar, crear el detalle de tiempo variante, resumir y combinar los extractos de datos, ayudan a crear el ambiente para el acceso a la información Institucional. Este nuevo enfoque ayuda a las personas individuales, en todos los niveles de la empresa, a efectuar su toma de decisiones con más responsabilidad.” (5)

Los principales objetivos de un almacén de datos son: (6)

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

1.1.1. Características de un almacén de datos

Bill Inmon¹ fue uno de los primeros autores en escribir sobre el tema de los almacenes de datos: define un almacén en términos de las características del repositorio de datos: (7)

- **Orientado a temas.**- Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.
- **Variante en el tiempo.**- Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.
- **No volátil.**- La información no se modifica ni se elimina, una vez almacenado un dato se convierte en información de sólo lectura, y se mantiene para futuras consultas.
- **Integrado.**- La base de datos contiene los datos de todos los sistemas operacionales de la organización. Estos datos deben ser consistentes.

“Otra característica es que contiene metadatos, es decir, datos sobre los datos. Los metadatos permiten saber la procedencia de la información, su periodicidad de refresco, su fiabilidad, forma de cálculo y otros.” (8)

Los metadatos son los que permiten simplificar y automatizar la obtención de la información desde los sistemas operacionales a los sistemas informacionales.

1.2. Almacén de Datos y Base de Datos Corporativa

“Se puede caracterizar un Data Warehouse haciendo un contraste de cómo los datos de un negocio almacenados en este, difieren de los datos operacionales usados por las aplicaciones de producción.” (9)

¹ Bill Inmon, considerado actualmente como el padre del Data Warehouse.

Base de Datos Operacional	Data Warehouse
Datos operacionales	Datos del negocio para información
Orientado a la aplicación	Orientado al sujeto
Actual	Actual + Histórico
Detallada	Detallada + Resumida
Cambia continuamente	Estable

Figura 1.1 Tabla comparativa.

“El ingreso de datos en el DW viene desde el ambiente operacional en casi todos los casos. El mismo es siempre un almacén de datos transformados y separados físicamente de la aplicación donde se encontraron los datos en el ambiente operacional.” (10)

A menudo se almacena gran cantidad de información, la cual está a veces subdividida en pequeñas unidades lógicas. Periódicamente, se importan los datos de otros sistemas de información dentro del almacén de datos, para realizar sobre ellos un procesamiento posterior.

Además está preparado para ser explotado mediante herramientas específicas que permiten la extracción de información significativa y patrones de comportamiento que permanecen ocultos en un enorme repositorio de datos.

1.3. Conceptos fundamentales

1.3.1. Datamart

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

Data Warehouse, o integrar por sí mismo un compendio de distintas fuentes de información.” (11)

Por tanto, para crear el Datamart de un área funcional de la empresa es preciso encontrar la estructura óptima para el análisis de su información, estructura que puede estar montada sobre una base de datos OLTP², como el propio almacén de datos, o sobre una base de datos OLAP³ (12). La designación de una u otra dependerá de los datos, los requisitos y las características específicas de cada departamento. De esta forma, se pueden plantear dos tipos de Datamarts:

Datamart OLAP:

“Se basan en los populares cubos OLAP, que se construyen agregando, según los requisitos de cada área o departamento, las dimensiones y los indicadores necesarios de cada cubo relacional. El modo de creación, explotación y mantenimiento de los cubos OLAP es muy heterogéneo, en función de la herramienta final que se utilice.” (13)

Datamart OLTP:

“Pueden basarse en un simple extracto del Data Warehouse, no obstante, lo común es introducir mejoras en su rendimiento (las agregaciones y los filtrados suelen ser las operaciones más usuales) aprovechando las características particulares de cada área de la empresa. Las estructuras más comunes en este sentido son las tablas de reportes y las vistas materializadas, que se construyen con la misma estructura que las anteriores, pero con el objetivo de explotar la reescritura de consultas (queries).” (14)

² Los sistemas OLTP son bases de datos orientadas al procesamiento de transacciones. Una transacción genera un proceso atómico y que puede involucrar operaciones de inserción, modificación y borrado de datos. El proceso transaccional es típico de las bases de datos operacionales.

³ Los sistemas OLAP (Procesamiento Analítico en Línea) son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejos... etc. Este sistema es típico de los Datamarts.

1.3.2. Estructura Multidimensional

Normalmente, una base de datos tiene una estructura bidimensional, es decir, cada columna puede proporcionar un dato para un elemento de cada fila. Así pues, se pueden tener bastantes bases de datos para equiparar toda la información que guarda una empresa. Sin embargo, resulta engorroso trabajar con tanta información, que finalmente, a pesar de estar organizada, puede no ser útil al usuario.

“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: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejo, etc. Este sistema es típico de los Datamarts.” (15)

“Los cubos son elementos claves en OLAP (online analytic processing), una tecnología que provee rápido acceso a datos en un almacén de datos (Data Warehouse). Los cubos proveen un mecanismo para buscar datos con rapidez y tiempo de respuesta uniforme independientemente de la cantidad de datos en el cubo o la complejidad del procedimiento de búsqueda.” (16)

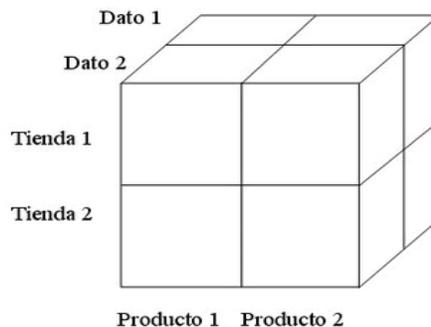


Figura 1.2 Cubo multidimensional.

Su objetivo es agrupar los datos con el propósito de facilitar su posterior análisis, de forma que sean útiles para acceder y analizar información sobre la propia empresa.

1.3.3. Minería de Datos

“La minería de datos (en inglés, data mining) se define como la extracción no trivial de información implícita, previamente desconocida y potencialmente útil, a partir de datos.” (17)

Es el proceso de descubrir conocimiento interesante de grandes cantidades de datos contenidas en bases de datos, almacenes de datos u otro repositorio de información.

1.4. Estructura de un almacén de datos

Los almacenes de datos tienen una estructura distinta de las bases de datos operacionales. A continuación se explica detalladamente lo antes expuesto. (Ver Anexo 1)

Los diferentes componentes del almacén de datos son: (18)

- Detalle de datos actuales.
- Detalle de datos antiguos.
- Datos ligeramente resumidos.
- Datos completamente resumidos.
- Metadata.

Detalle de datos actuales: Obtenidos directamente del procesado de los datos.

- Forman el nivel más bajo de detalle.
- Ocupan mucho espacio.
- Se almacenan en disco, para facilitar el acceso.

Detalle de datos antiguos: Igual que los anteriores, pero con datos correspondientes al pasado.

- Se suelen almacenar en un medio externo, ya que su acceso es poco frecuente.

Datos ligeramente resumidos: Primer nivel de agregación de los datos detallados actuales.

- Corresponden a consultas habituales.

- Se almacenan en disco.

Datos completamente resumidos: Son el nivel más alto de agregación.

- Corresponden a consultas que se realizan muy a menudo y que se deben obtener muy rápidamente.
- Suelen estar separados del Almacén de Datos, formando Supermercados de Datos (Datamarts).

Metadata: Describen la estructura de los datos contenidos en el almacén.

- Están en una dimensión distinta al resto de niveles.

1.5. Arquitectura de un almacén de datos

Una de las razones por las que el desarrollo de un almacén de datos crece rápidamente, es que realmente es una tecnología muy entendible. De hecho, puede representar mejor la estructura amplia de una empresa para administrar los datos e informaciones dentro de la organización. A fin de comprender cómo se relacionan todos los componentes involucrados en el almacén, es esencial tener una arquitectura fiable.

1.5.1. Elementos de una arquitectura para el Almacén de Datos

“Una Arquitectura Data Warehouse (Data Warehouse Architecture - DWA) es una forma de representar la estructura total de datos, comunicación, procesamiento y presentación, que existe para los usuarios finales que disponen de una computadora dentro de la empresa.” (19)

La arquitectura se constituye de un número de partes interconectadas: (20)

- Base de datos operacional / Nivel de base de datos externo.
- Nivel de acceso a la información.
- Nivel de acceso a los datos.
- Nivel de directorio de datos (Metadata).
- Nivel de gestión de proceso.

- Nivel de mensaje de la aplicación.
- Nivel del almacén de datos.
- Nivel de organización de datos.

Base de datos operacional / Nivel de base de datos externo:

Los sistemas operacionales procesan datos para apoyar las necesidades operacionales críticas. Para hacer eso, se han creado las bases de datos operacionales históricas que proveen una estructura de procesamiento eficiente, para un número relativamente pequeño de transacciones comerciales bien definidas.

Ciertamente, la meta del almacén es liberar la información que es almacenada en bases de datos operacionales y combinarla con la información desde otra fuente de datos, generalmente externa.

Nivel de acceso a la información:

El nivel de acceso a la información de la arquitectura del almacén de datos, es el nivel del que el usuario final se encarga directamente. En particular, representa las herramientas que el usuario final normalmente usa día a día. Por ejemplo: Excel, Lotus 1-2-3, Access, SAS, entre otros.

Este nivel también incluye el hardware y el software involucrados en mostrar información en pantalla y emitir reportes de impresión, hojas de cálculo, gráficos y diagramas para el análisis y presentación.

Nivel de acceso a los datos:

El nivel de acceso a los datos está involucrado con el nivel de acceso a la información para conversar en el nivel operacional.

El nivel de acceso a los datos no solamente conecta bases de datos diferentes y sistemas de archivos sobre el mismo hardware, sino también a los fabricantes y protocolos de red. Una de

Capítulo I: Fundamentación Teórica

las claves de una estrategia para el almacén es proveer a los usuarios finales con "acceso a datos universales".

El acceso a los datos universales significa que, teóricamente por lo menos, los usuarios finales sin tener en cuenta la herramienta de acceso a la información o ubicación, deberían ser capaces de acceder a cualquier o todos los datos en la empresa que es necesaria para hacer su trabajo.

El nivel de acceso a los datos entonces, es responsable de las interfaces entre las herramientas de acceso a la información y las bases de datos operacionales.

Nivel de Directorio de Datos (Metadata):

A fin de proveer el acceso a los datos universales, es absolutamente necesario mantener alguna forma de directorio de datos o repositorio de la información metadata. La metadata es la información alrededor de los datos dentro de la empresa.

A fin de tener un depósito totalmente funcional, es necesario tener una variedad de metadata disponible, información sobre las vistas de datos de los usuarios finales e información sobre las bases de datos operacionales.

Los usuarios finales deberían de acceder a los datos desde el almacén (o desde las bases de datos operacionales), sin tener que conocer dónde residen los datos o la forma en que se han almacenado.

Nivel de Gestión de Procesos:

El nivel de gestión de procesos tiene que ver con la programación de diversas tareas que deben realizarse para construir y mantener el almacén de datos. Este nivel puede depender del alto nivel de control de trabajo para muchos procesos (procedimientos) que deben ocurrir para mantener el almacén actualizado.

Nivel de Mensaje de la Aplicación:

El nivel de mensaje de la aplicación tiene que ver con el transporte de información alrededor de la red de la empresa. El mensaje de aplicación se refiere también como "subproducto", pero puede involucrar sólo protocolos de red. Puede usarse por ejemplo, para aislar aplicaciones operacionales o estratégicas a partir del formato de datos exacto, recolectar transacciones o mensajes y entregarlos a una ubicación segura en un tiempo seguro.

Nivel del almacén de datos (Físico):

En el almacén de datos (núcleo) es donde ocurre la data actual, usada principalmente para usos estratégicos. En un almacén físico, se realizan muchas copias de datos operacionales y/o externos, que son almacenadas realmente en una forma que es fácil de acceder y es altamente flexible.

Nivel de Organización de Datos:

El componente final de la arquitectura es la organización de los datos. Se llama también gestión de copia o réplica, pero de hecho, incluye todos los procesos necesarios: seleccionar, editar, resumir, combinar y cargar datos en el depósito y acceder a la información desde bases de datos operacionales y/o externas.

1.6. Tecnologías y Herramientas

Las herramientas de negocio inteligentes se han convertido en los sucesores de los sistemas de soporte de decisión, pero tienen un alcance más amplio. No solamente ayudan en las decisiones de soporte sino, en muchos casos, soportan muchas funciones operacionales y de misión crítica de la compañía.

1.6.1. Sistemas de Gestión de bases de datos

Las principales funciones que debe cumplir un SGBD se relacionan con la creación y mantenimiento de la base de datos, el control de accesos, la manipulación de datos de acuerdo con las necesidades del usuario, el cumplimiento de las normas de tratamiento de datos, evitar redundancias e inconsistencias y mantener la integridad.

1.6.1.1. MySQL Server

MySQL es un sistema de gestión de base de datos relacional, multihilo y multiusuario. Una de las características más interesantes de MySQL es que permite recurrir a bases de datos multiusuario a través de la web y en diferentes lenguajes de programación, que se adaptan a diferentes necesidades y requerimientos: (21)

- Está disponible para la mayoría de las plataformas de sistemas operativos.
- Su bajo consumo lo hacen apto para ser ejecutado en una máquina con escasos recursos sin ningún problema.
- El conjunto de aplicaciones Apache-PHP-MySQL es uno de los más utilizados en aplicaciones en ambiente Web.
- Velocidad a la hora de realizar las operaciones.

1.6.1.2. SQL Server

Microsoft SQL Server es un sistema de gestión de bases de datos relacionales (SGBD) basado en el lenguaje Transact-SQL⁴, capaz de poner a disposición de muchos usuarios grandes cantidades de datos de manera simultánea. (22)

Las utilidades de administración de este gestor son envidiables para muchos de los gestores comerciales existentes, debido a su gran facilidad de configuración e instalación.

Desventajas: (23)

- Sólo permite 16 instancias distintas concurrentes aproximadamente en una máquina.
- No maneja compresión de datos, por tanto ocupa mucho espacio en disco.
- Está atado a la plataforma del sistema operativo sobre la cual puede instalarse.
- Es privativa, lo que provoca gastos excesivos con su aplicación.

⁴ Transact SQL es el lenguaje de programación que proporciona SQL Server para ampliar SQL con los elementos característicos de los lenguajes de programación: variables, sentencias de control de flujo, bucles, etc.

1.6.1.3. PostgreSQL

“**PostgreSQL** es un motor de base de datos, es servidor de base de datos relacional libre, liberado bajo la licencia BSD⁵.” (24)

PostgreSQL provee nativamente soporte para: (25)

- Números de precisión arbitraria.
- Texto de largo ilimitado.
- Figuras geométricas (con una variedad de funciones asociadas).
- Direcciones IP (IPv4 e IPv6).
- Arrays.

Otras características: (26)

- Claves ajenas también denominadas llaves ajenas o claves foráneas.
- Disparadores⁶ (*triggers*): Un disparador o trigger se define en una acción específica basada en algo ocurrente dentro de la base de datos. En PostgreSQL esto significa la ejecución de un procedimiento almacenado basado en una determinada acción sobre una tabla específica.

1.6.2. Modelado de la base de datos. Visual Paradigm

“Visual Paradigm es una herramienta UML profesional que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. Presenta licencia gratuita y comercial. Es fácil de instalar y actualizar y es compatible entre ediciones.” (27)

⁵ La licencia BSD es la licencia original de una distribución de Software: *Berkeley Software Distribution*.

⁶ Los disparadores (*triggers* en inglés) son funciones enlazadas a operaciones sobre los datos.

Características principales: (28)

- Ingeniería de ida y vuelta.
- Ingeniería inversa - Código a modelo, código a diagrama.
- Ingeniería inversa Java, C++, Esquemas XML, ET, exe/dl.
- Generación de código - Modelo a código, diagrama a código.
- Diagramas de flujo de datos.
- Generación de objetos Java desde la base de datos.
- Generación de bases de datos - Transformación de diagramas de Entidad-Relación en tablas de base de datos.
- Ingeniería inversa de bases de datos - Desde Sistemas Gestores de Bases de Datos (DBMS) existentes a diagramas de Entidad-Relación.
- Generador de informes.
- Importación y exportación de ficheros XMI.
- Editor de figuras.

1.6.3. Gestión de Relación con Clientes (CRM en inglés)

La Gestión de Relación con Clientes es básicamente la respuesta de la tecnología a la creciente necesidad de las empresas de fortalecer las relaciones con sus clientes.

“Las herramientas de Gestión de Relación con Clientes (Customer Relationship Management CRM) son las soluciones tecnológicas para conseguir desarrollar la "teoría" del marketing relacional. El marketing relacional se puede definir como "la estrategia de negocio centrada en anticipar, conocer y satisfacer las necesidades y los deseos presentes y previsibles de los clientes.” (29)

“La meta principal de un CRM no se enfoca necesariamente en reducir los costos del negocio. Las soluciones de CRM mejoran los esfuerzos de ventas y mercadotecnia, y permiten a las organizaciones proporcionar un servicio superior a los clientes. Los nuevos clientes se ganan y los ya existentes se retienen y se busca que compren productos o servicios en una mayor cantidad.” (30)

Los clientes finales se benefician al recibir un mejor servicio y obtener los productos y servicios que desean, cuando lo desean. Una empresa que no cuenta con una estrategia de CRM está en una desventaja competitiva.

1.6.3.1. Sugar CRM

“Sugar CRM es un software CRM de código libre (open source). Funciona enteramente sobre un servidor web con soporte para PHP, por ejemplo, Apache. Sugar CRM tiene una base sólida, y se puede expandir mediante módulos. El software CRM permite a las empresas gestionar de manera eficiente la relación con sus clientes, lo que provoca que las acciones de marketing sean más efectivas.” (31)

Tiene tres versiones, una de ellas libre y otras dos versiones con componentes no-libres y con un costo por usuario. Sugar CRM es una aplicación CRM muy completa para negocios de distinto tamaño. Está diseñada para facilitar la gestión de ventas, oportunidades, contactos de negocios y más. (32)

1.6.3.2. Vtiger CRM

“Vtiger CRM es una solución CRM de fuente abierta, distribuido bajo los términos de la licencia MPL versión 1.1 (Mozilla Public License). Está construido sobre las tecnologías de Apache, PHP y MySQL, todas también equipamiento lógico de fuente abierta.” (33)

Vtiger CRM incluye, entre otras cosas, las siguientes características: (34)

- Automatización de fuerza de ventas.
- Servicio de servicio y soporte a clientes.
- Automatización de mercadotecnia.
- Gestión de inventario.
- Soporte para múltiples bases de datos.
- Gestión de seguridad.
- Personalización de productos.
- Calendario.
- Correo electrónico a través de interfaz HTTP.

- Componentes adicionales (plug-ins): soporte para Microsoft Outlook, Soporte para Microsoft Office, Extensión para Mozilla Thunderbird, Portal para clientes, formularios y muchos otros más.

1.6.4. Herramientas Pentaho

“La plataforma de Inteligencia de Negocio Pentaho cubre muy amplias necesidades de Análisis de los Datos y de los Informes empresariales. Las soluciones de Pentaho están escritas en Java y tienen un ambiente de implementación también basado en Java. Eso hace que Pentaho es una solución muy flexible para cubrir una amplia gama de necesidades empresariales, tanto las típicas como las sofisticadas y específicas al negocio.” (35)

Los módulos de la Suite de Pentaho son:

1.6.4.1. Pentaho Analysis Service

“Ahora bautizado como “Pentaho Analysis Service” forma parte del motor OLAP integrado en el SUITE BI de PENTHO.” (36)

Mondrian funciona sobre las bases de datos estándar del mercado: Oracle, DB2, SQL-Server, MySQL, Postgre, etc., lo cual habilita y facilita el desarrollo de negocio basado en la plataforma Pentaho. (Ver Anexo 2)

1.6.4.2. Pentaho Report Design Wizard

Solución proporcionada por Pentaho e integrada en su suite para el desarrollo de informes.

Existen tres productos con diferentes enfoques y dirigidos a diferentes tipos de usuarios: (37)

1. Pentaho Report Designer: Editor basado en eclipse con prestaciones profesionales y de calidad y con capacidad de personalización de informes a las necesidades de negocio destinado a desarrolladores.

Este módulo incluye asistentes para facilitar la configuración de propiedades. Está estructurado de forma que los desarrolladores pueden acceder a sus prestaciones de forma rápida:

Incluye un editor de consultas para facilitar la confección de los datos que serán utilizados en un informe.

2. Pentaho Report Design Wizard: Herramienta de diseño de informes, que facilita el trabajo y permite a los usuarios obtener resultados de forma inmediata. Está destinada a usuarios con menos conocimientos técnicos.

3. Web ad-hoc reporting: Es el similar a la herramienta anterior pero vía web. Extiende la capacidad de los usuarios finales para la creación de informes a partir de plantillas pre-configuradas y con la ayuda de un asistente de creación.

1.6.4.3. Kettle “Pentaho Data Integration”

Kettle es un proyecto belga que incluye un conjunto de herramientas para realizar ETL (Extracción, Transformación y Carga). Uno de sus objetivos es que el proyecto ETL sea fácil de generar, mantener y desplegar. (38)

Se compone de 4 herramientas:

SPOON: permite diseñar de forma gráfica la transformación ETL.

PAN⁷: ejecuta las transformaciones diseñadas con SPOON.

CHEF: permite, mediante una interfaz gráfica, diseñar la carga de datos que incluye un control de estado de los trabajos.

⁷ PAN es un motor de transformación de datos permitiendo la lectura y escritura sobre diversas fuentes de datos.

KITCHEN⁸: permite ejecutar los trabajos.

1.6.4.4. Weka: Minería de Datos

“Provee un completo conjunto de algoritmos que automatizan los procesos de transformación de datos a la forma en que la minería de datos puede explotarlos. Los resultados pueden ser visualizados en modo grafico ya sea agrupado, segmentado, de árbol de decisión, bosque aleatorio, redes neurales y componentes de análisis.” (39)

Utiliza filtros para la discreción, normalización, re-muestreo, selección y transformación de atributos. Maneja clasificadores y brinda modelos para la predicción nominal o cantidades numéricas.

1.6.4.5. Herramienta para el diseño de los cubos

Mondrian Schema Workbench es una herramienta gráfica para el diseño de los cubos multidimensionales. El motor de Mondrian procesa peticiones de MDX que permiten configurar el cubo en dependencia de los intereses. Genera un archivo XML que registra todas las acciones realizadas con la herramienta.

1.6.5. Herramienta Redmine

“Redmine es una herramienta de Gestión de Proyectos de Software con interface web que funciona bajo el framework de Ruby on Rails. Esta herramienta permite dar de alta proyectos, jefes de proyecto y desarrolladores mediante la interface web. Una vez dados de alta se pueden definir los hitos del proyecto y las tareas que se realizarán para cada uno de estos hitos. Redmine permite obtener el gráfico de Gantt para cada hito, y asignar una tarea para cada miembro desarrollador.” (40)

Dentro de sus características tenemos: (41)

- Soporta múltiples proyectos simultáneamente, seguimiento de tiempos.

⁸ KITCHEN es una aplicación que permite ejecutar “Jobs” planificados en modo “batch” (serie, lote) para correr automáticamente en intervalos de tiempo definidos.

- Así como la integración con herramientas de gestión de versiones.
- En cuanto a gestión de usuarios, el control es basado en roles y permite el auto registro de los mismos.
- Genera información en base a calendarios y diagramas Gantt.
- Permite la creación de una wiki y foros por proyecto.
- Tiene campos ajustados a cada proyecto, cada usuario así como ventanas de tiempo.
- Soporta autenticación a LDAP.
- Puede usar cualquier gestor de bases de datos relacional.
- Envía notificaciones por correo a los desarrolladores cada vez que se asigna una tarea o ante cualquier evento relacionado con el proyecto.

1.7. Metodología propia para la Construcción de un almacén de datos

1.7.1. Metodología Hefesto. Características

Hefesto es una metodología propia, cuya propuesta está fundamentada en una muy amplia investigación, comparación de metodologías existentes y experiencias propias en procesos de confección de almacenes de datos. (42)

La idea principal, es comprender cada paso que se realizará, para no tener que seguir un método al pie de la letra sin saber exactamente qué se hace y por qué.

Esta metodología cuenta con las siguientes características: (43)

- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- Se basa en los requerimientos del usuario, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.
- Reduce la resistencia al cambio, ya que involucra al usuario final en cada etapa para que tome decisiones respecto al comportamiento y funciones del almacén.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- Es independiente de las herramientas que se utilicen para su implementación.

- Es independiente de las estructuras físicas que contengan el 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.
- Se aplica tanto para almacenes de datos como para Datamart.

1.7.2. Pasos para la aplicación de la metodología

Paso 1) Análisis de Requerimientos

Lo primero que se hace es identificar los requerimientos del usuario a través de preguntas que expliquen los objetivos de su organización. Luego, se analizan las preguntas a fin de identificar cuáles son los indicadores y perspectivas tomados en cuenta para la construcción del almacén. Finalmente, se confecciona un modelo conceptual en donde se pueda visualizar el resultado obtenido en este primer paso.

Paso 2) Análisis de los OLTP

Seguidamente, se analizan las fuentes OLTP para determinar cómo serán calculados los indicadores y para establecer las respectivas correspondencias entre el modelo conceptual creado en el paso anterior y las fuentes de datos. Luego, se definen qué campos se incluirán en cada perspectiva. Finalmente, se amplía el modelo conceptual con la información obtenida en este paso.

Paso 3) Modelo Lógico del almacén de datos

A continuación, se confecciona el modelo lógico⁹ de la estructura del almacén de datos, para lo cual se tiene de base el modelo conceptual que ya ha sido creado. Para ello, primero se define el tipo de modelo a utilizar y luego se llevan a cabo las acciones propias al caso, para diseñar

⁹ Modelo Lógico: representación de una estructura de datos, que puede procesarse y almacenarse en algún SGBD.

las tablas de dimensiones y de hechos. Finalmente, se realizan las uniones pertinentes entre estas tablas.

Paso 4) Procesos ETL

Una vez construido el modelo lógico, se procede a probarlo con datos, a través de procesos ETL. Para realizar la compleja actividad de extraer datos de diferentes fuentes, para luego integrarlos, filtrarlos y depurarlos, existen varios software que facilitan estas tareas, por lo cual este paso se centrará sólo en la generación de las sentencias SQL que contendrán los datos que serán de interés.

Antes de realizar la carga de datos, es conveniente efectuar una limpieza de los mismos, para evitar valores faltantes y anómalos. Al generar los ETL, se debe tener en cuenta cual es la información que se desea almacenar en el depósito de datos, para ello se pueden establecer condiciones adicionales y restricciones. Estas condiciones deben ser analizadas y llevadas a cabo con mucha prudencia para evitar pérdidas de datos importantes.

Cuando se haya cargado en su totalidad el almacén de datos, se deben establecer sus políticas de actualización o refresco de datos.

Después del análisis realizado se define como sistema gestor de bases de datos MySQL, ya que es un sistema robusto y de código abierto, además existe una herramienta, llamada phpMyAdmin, que permite administrar las bases de datos mediante una interfaz sencilla. La herramienta para la Gestión de Relación con Clientes que se decide utilizar es Vtiger CRM, ya que está probada y certificada, y es de uso libre.

Como herramienta para el diseño del almacén se precisa utilizar la Suite de Inteligencia de Negocios Pentaho con todos sus módulos integrados por su importancia para el desarrollo empresarial, porque permite el análisis de información histórica y actual, así como los reportes para la ayuda a la toma de decisiones.

Otra herramienta a utilizar es el Redmine la cual fue definida por la dirección técnica de la universidad para el trabajo y la gestión de los proyectos y todas las operaciones que incluyen, además es la más utilizada en la actividad productiva.

Capítulo I: Fundamentación Teórica

Se decide utilizar la metodología Hefesto porque está probada en cuanto a la construcción de los almacenes de datos y a través de la misma se realiza un análisis que ayudará a conocer el funcionamiento y accionar de la empresa. Esto permite examinar e interpretar de forma óptima las necesidades de información de la misma y también apoya una mejor construcción y adaptación del depósito de datos.

Capítulo 2. Análisis del Almacén de Datos

En este capítulo se presenta el análisis realizado a las diferentes áreas del Centro de Informática Médica (CESIM), lo cual se logra a través de la metodología Hefesto, seleccionada anteriormente por ser la más adecuada para dar cumplimiento a los objetivos propuestos y así diseñar correctamente el Almacén de Datos del CESIM. Se lleva a cabo un estudio de los procesos¹⁰ que son de mayor interés para cada una de las áreas establecidas, y con los datos se plantean los indicadores que se analizan para posteriormente hacer el diseño del almacén.

2.1. Análisis de los procesos en los departamentos de interés

Actualmente el Grupo de Mercadotecnia del CESIM realiza un gran esfuerzo para cumplir con los compromisos pactados con los clientes y no descuidar la atención a clientes potenciales, teniendo en consideración que el centro no cuenta con una herramienta para la gestión de la relación con los clientes, que garantice la gestión de la información de los mismos, su seguimiento, atención personalizada y que permita llevar a cabo correctamente la estrategia de soporte de las aplicaciones.

Asimismo en los departamentos de producción no se realiza un seguimiento efectivo de los proyectos productivos, y existen dificultades para medir el desarrollo de las tareas en la producción, por lo que el resultado a veces no es el esperado.

Después de registrar los problemas anteriores se plantea una solución para apoyar la toma de decisiones y así mitigar las dificultades anteriormente relacionadas. El diseño de un almacén de datos brinda un mayor entendimiento de los procesos y actividades que se llevan a cabo para lograr el buen desempeño en los departamentos de producción y el Grupo de Mercadotecnia. Además se puede medir a través de indicadores el cumplimiento de los objetivos estratégicos propuestos y tomar decisiones certeras para cumplirlos.

2.2. Estructura del Almacén de Datos

¹⁰ Un proceso está determinado por algún área en particular, que define una serie de actividades vinculadas entre sí, que dan un resultado final en dependencia de lo que se quiera obtener.

Capítulo II: Análisis del Almacén de Datos

Dentro de la estructura del almacén de datos del CESIM se especifica el tiempo en el cual será evaluada la información y la forma en que se almacena dentro del almacén.

Cuando la información ingresa al depósito de datos se almacena a nivel de Detalle de datos actuales, en este caso particular el Estado de los objetivos estratégicos (2009-2010). Los datos permanecerán allí hasta que ocurra alguno de los tres eventos siguientes:

- Sean borrados del depósito de datos.
- Sean resumidos, ya sea a nivel de Objetivos estratégicos por trimestre o anualmente.
- Sean archivados en el estado de los objetivos estratégicos (2009), que son los sistemas externos en los cuales se guarda toda la información desde los primeros años.

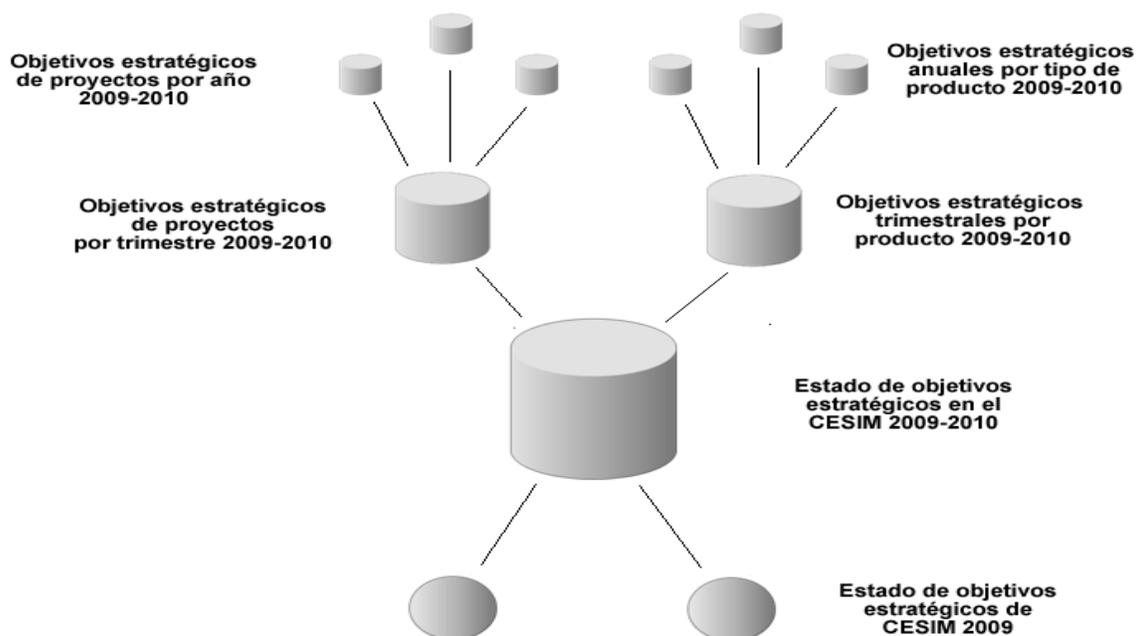


Figura 2.1 Estructura del Almacén de Datos del CESIM.

2.3. Análisis de requerimientos

Para dar cumplimiento a los pasos planteados por la metodología Hefesto se analizan los requerimientos que son de gran importancia para el entendimiento del negocio que se evalúa.

Se realiza una entrevista con los usuarios, en este caso el Grupo de Mercadotecnia y los departamentos de producción, en busca de sus necesidades de información. Se analiza cada uno de los procesos que intervienen en cada área y cuáles son los indicadores para identificarlos, los cuales se deben tener en cuenta a la hora de tomar las decisiones.

Además, se procede a identificar lo que les interesaba conocer acerca de los procesos, es decir, saber cuáles eran las variables o perspectivas¹¹ que debían tenerse en cuenta para poder tomar decisiones basadas en ello.

En cuanto a estas perspectivas se decide utilizar las predefinidas por el Cuadro de Mando Integral, con el fin de integrar la totalidad de puntos de vista bajo los que puede observarse la gestión de una empresa. Las perspectivas fundamentales son cuatro:

- Perspectiva financiera.
- Perspectiva del cliente.
- Perspectiva del proceso interno.
- Perspectiva de aprendizaje y crecimiento.

Del análisis realizado se obtiene que las perspectivas que se ponen de manifiesto son: la financiera, la del cliente, y la del proceso interno, las cuales se relacionan en la figura 2.2.

¹¹ La perspectiva sería en el lenguaje del modelo relacional una tabla de la base de datos, que se identifica por atributos o características. En nuestro caso son tablas que se crean en dependencia del proceso al que pertenece para dar cumplimiento a los indicadores propuestos.

Capítulo II: Análisis del Almacén de Datos

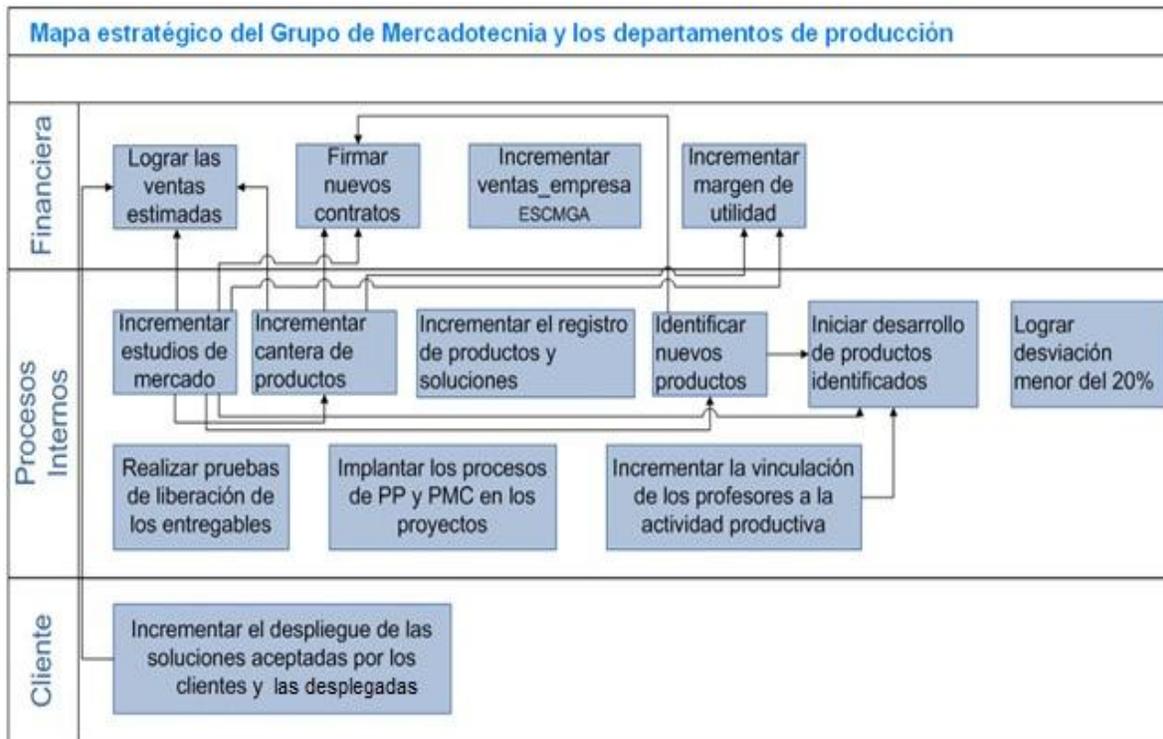


Figura 2.2 Mapa estratégico de las áreas analizadas.

Después del análisis de los requerimientos del Centro de Informática Médica, se deciden presentar los siguientes indicadores que dan solución a varios objetivos estratégicos identificados en las áreas analizadas:

Objetivos estratégicos relacionados con el Grupo de Mercadotecnia

- 1- Realizar estudios de mercado para el 80% de los productos, servicios y soluciones informáticas integrales identificadas.
- 2- Lograr ventas por no menos de 12 millones de USD.
- 3- Firmar nuevos contratos por no menos de 6 MM USD.
- 4- Contar con una cantera de no menos de 4 productos y servicios listos para comercializar.
- 5- Lograr un incremento de las ventas superior a 6 MM USD por concepto de la empresa mixta ESCMGA.
- 6- Conformar el catálogo de productos, servicios y soluciones informáticas integrales.

Capítulo II: Análisis del Almacén de Datos

Objetivos estratégicos relacionados con los departamentos de producción

- 1- Lograr que el 50% de los productos y soluciones tecnológicas se encuentren registrados o en proceso, ante las instituciones estatales correspondientes.
- 2- Identificar al menos cinco nuevos productos de informática médica.
- 3- Iniciar el desarrollo de al menos dos de los productos identificados en el año anterior.
- 4- Lograr que la desviación con respecto al cronograma inicial de los proyectos no sea mayor del 20% por razones imputables a la Universidad.
- 5- Lograr que el 50% de las soluciones aceptadas por los clientes sean desplegadas en el período en cuestión.
- 6- Tener un margen de utilidad de no menos del 30%.
- 7- Ejecutar pruebas de liberación, aceptación y pilotos al 100% de los entregables de todos los proyectos de desarrollo del Centro.
- 8- Implantar los procesos de planificación, seguimiento y control del Programa de Mejora en 40% de los proyectos del centro.
- 9- Mantener una vinculación a la actividad productiva del centro del 95%, como mínimo, para el caso de los profesores de la Especialidad y un incremento del 20% del resto de los profesores.

En síntesis, los indicadores son:

- Por ciento de productos con estudios de mercado.
- Índice de venta.
- Precio del contrato.
- Cantidad de productos listos a comerciar.
- Índice Incremento de venta_ESCMGA.
- Por ciento de productos registrados.
- Cantidad de productos nuevos identificados.
- Cantidad de productos en desarrollo.
- Por ciento de proyectos desviados.
- Por ciento de soluciones aceptadas por clientes y desplegadas.
- Por ciento de margen de utilidad.

Capítulo II: Análisis del Almacén de Datos

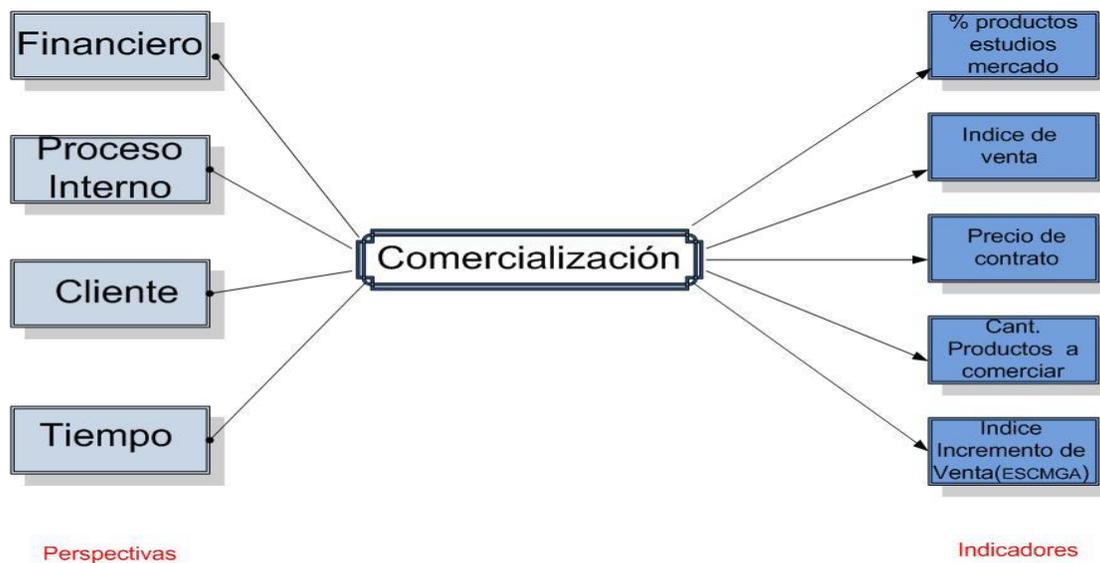
- Por ciento de entregables con PL.
- Por ciento de cumplimiento de hitos del PM.
- Por ciento de proyectos con PP y PMC.
- Por ciento de profesores de la especialidad.
- Incremento de profesores que no son de la especialidad.

Y las perspectivas de análisis son:

- Financiera.
- Cliente.
- Proceso Interno.
- Tiempo.

A partir de los indicadores y perspectivas obtenidas se construye el modelo conceptual¹². A través de este modelo, se podrá observar con claridad cuál es el alcance del proyecto, para luego poder trabajar sobre ellos.

Modelo conceptual del Grupo de Mercadotecnia:



¹² Descripción de alto nivel de la estructura de la base de datos, en la cual la información es representada a través de objetos, relaciones y atributos.

Figura 2.3 Modelo conceptual del Grupo de Mercadotecnia.

Modelo conceptual de los departamentos de producción:

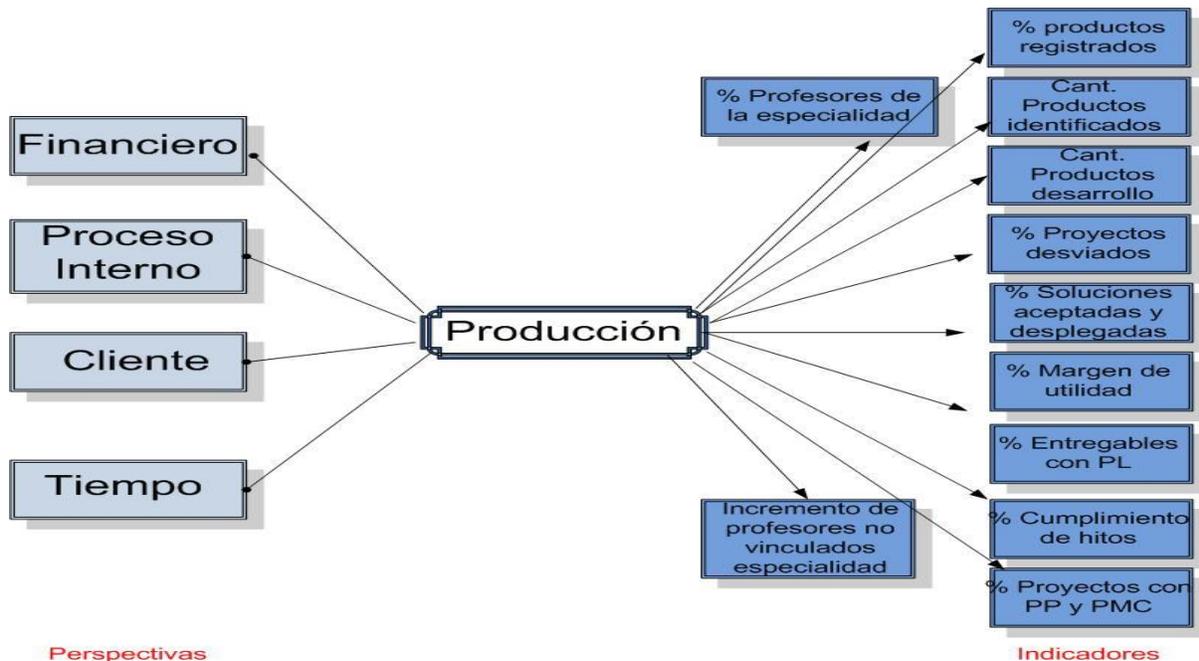


Figura 2.4 Modelo conceptual de los departamentos de producción.

Como puede observarse, la relación mediante la cual se unen las diferentes perspectivas, para obtener como resultado los indicadores requeridos por los usuarios, es precisamente el proceso o área que se analiza.

2.4. Análisis de los indicadores que se necesitan calcular

Luego de realizar el modelo conceptual se prosigue con el análisis de los indicadores requeridos para dar cumplimiento a los objetivos estratégicos, y se define cómo se calcularían para poder tener un resultado concluyente de cómo se comportan las áreas en relación con cada objetivo planteado.

Los indicadores se calculan de la siguiente manera:

1- “Por ciento de productos estudios de mercado”:

Cpem: Cantidad de productos con estudios de mercado.

Capítulo II: Análisis del Almacén de Datos

Csem: Cantidad de servicios con estudios de mercado.

Csiem: Cantidad de soluciones integrales con estudios de mercado.

Tpssi: Total de productos, servicios y soluciones integrales.

- Hechos: $(Cpem + Csem + Csiem) / Tpssi * 100$.
- Función de sumarización: %.

Aclaración: el indicador “Por ciento de productos estudios de mercado” representa el por ciento de los productos, servicios y soluciones con estudios de mercado.

2- “Índice venta”:

- Hechos: precio de venta.
- Función de sumarización: SUM.

Aclaración: el indicador “Índice venta” representa el precio total de las ventas con un valor especificado por el proveedor.

3- “Precio de contrato”:

- Hechos: precio del contrato.
- Función de sumarización: SUM.

Aclaración: el indicador “Precio de contrato” representa el precio de un contrato con un valor especificado por el proveedor.

4- “Cantidad contratos internacionales”:

- Hechos: contratos internacionales.
- Función de sumarización: SUM.

Aclaración: el indicador “Cantidad contratos internacionales” representa la cantidad de contratos internacionales fuera del Venezolano existen en el CESIM.

5- “Cantidad de productos listos a comerciar”:

- Hechos: cantidad productos + cantidad servicios.

- Función de sumarización: SUM.

Aclaración: el indicador “Cantidad de productos listos a comerciar” representa la cantidad de los productos y servicios listos para comercializar.

6- “Índice Incremento de venta_ESCMGA”:

- Hechos: cantidad de ventas.
- Función de sumarización: SUM.

Aclaración: el indicador “Índice Incremento de venta_ESCMGA” representa un incremento de las ventas superior a un precio específico.

7- “Por ciento de productos registrados”:

Cpr: cantidad de productos registrados.

Cppr: cantidad de productos en proceso de registro.

Tp: Total de productos.

- Hechos: $(Cpr + Cppr) / Tp * 100$.
- Función de sumarización: %.

Aclaración: el indicador “Por ciento de productos registrados” representa el por ciento de los productos registrados o en proceso de registro.

8- “Cantidad de productos nuevos identificados”:

- Hechos: cantidad productos identificados.
- Función de sumarización: SUM.

Aclaración: el indicador “Cantidad de productos nuevos identificados” representa la cantidad de productos nuevos identificados.

9- “Cantidad de productos en desarrollo”:

- Hechos: cantidad productos en desarrollo.
- Función de sumarización: SUM.

Capítulo II: Análisis del Almacén de Datos

Aclaración: el indicador “Cantidad de productos en desarrollo” representa el desarrollo al menos mínimo de dos de los productos identificados en el año anterior.

10- “Por ciento de proyectos desviados”:

Cpa: cantidad de proyectos atrasados en un 20% del plan original.

Tp: Total de proyectos.

- Hechos: $(Cpa / Tp) * 100$.
- Función de sumarización: SUM.

Aclaración: el indicador “Por ciento de proyectos desviados” representa la suma de la cantidad de proyectos atrasados y la suma del total de proyectos.

11- “Por ciento de soluciones aceptadas por clientes”:

Csad: cantidad de soluciones aceptadas por los clientes y desplegadas.

Tsa: Total de soluciones aceptadas.

- Hechos: $(Csad / Tsa) * 100$.
- Función de sumarización: %.

Aclaración: el indicador “Por ciento de soluciones aceptadas por clientes” representa el por ciento de soluciones aceptadas y desplegadas por los clientes.

12- “Por ciento de margen de utilidad”:

X: Gastos.

Y: Monto total del Proyecto.

- Hechos: $(Y - X) / Y * 100$.
- Función de sumarización: %.

Aclaración: el indicador “Por ciento de margen de utilidad” representa el por ciento de margen de utilidad de los proyectos.

13- “Por ciento de entregables con PL”:

Cepl: cantidad de entregables con pruebas de liberación, aceptación y pilotos.

Te: Total de entregables.

- Hechos: $Cepl / Te * 100$.

Capítulo II: Análisis del Almacén de Datos

- Función de sumarización: %.

Aclaración: el indicador “Por ciento de entregables con PL” representa el por ciento de entregables con pruebas de liberación.

14- “Por ciento de cumplimiento de hitos del PM”:

Chc: cantidad hitos cumplidos.

Thp: Total de hitos planificados.

- Hechos: $Chc / Thp * 100$.

- Función de sumarización: %.

Aclaración: el indicador “Por ciento de cumplimiento de hitos del PM” representa el por ciento de cumplimiento de los hitos del programa de mejoras.

15- “Por ciento de proyectos con proceso de mejoras”:

Cpp, pmc: cantidad de proyectos con los procesos de PP y PMC implantado.

Tp: Total de proyectos.

- Hechos: $(Cpp, pmc / Tp) * 100$.

- Función de sumarización: %.

Aclaración: el indicador “Por ciento de proyectos con proceso de mejoras” representa el por ciento de los procesos de planificación, seguimiento y control del Programa de Mejora implantados.

16- “Por ciento de profesores de la especialidad”:

Cpev: cantidad de profesores de la especialidad vinculados a la producción.

Tpe: Total de profesores de la especialidad.

- Hechos: $Cpev / Tpe * 100$.

- Función de sumarización: %.

Aclaración: el indicador “Por ciento de profesores de la especialidad” representa el por ciento de los profesores de la especialidad vinculados a la producción.

17- “Por ciento de incremento de profesores”:

Cpnevaa: cantidad de profesores que no son de la especialidad vinculados a la producción del año anterior.

Cpnevac: cantidad de profesores que no son de la especialidad vinculados a la producción año en curso.

- Hechos: $(Cpnevaa - Cpnevac) / Cpnevaa * 100$.
- Función de sumarización: %.

Aclaración: el indicador “Por ciento de incremento de profesores” representa el por ciento de incremento de los profesores que no son de la especialidad vinculados a la producción.

2.4.1. Nivel de granularidad

Después de establecidas las relaciones con las bases de datos, se examinan y seleccionan los campos que contendrá cada perspectiva, ya que será a través de ellos por los que se manipularán y filtrarán los indicadores.

Perspectiva “Financiera”:

- *Id_financiera*: es la clave primaria de la tabla “Producto” y representa unívocamente a un producto particular.
- *producto*: nombre o descripción del producto.
- *estado_producto*: indica si el producto está terminado o con atraso.
- *producto_registrado*: expresa si el producto está actualmente registrado o en proceso.
- *estudio_mercado*: es para saber si el producto tiene realizado un estudio de mercado.
- *precio_venta*: precio de venta del producto.
- *Id_identificación*: es para conocer los productos que han sido identificados.
- *pruebas_aceptación*: para conocer qué productos tienen realizadas las pruebas de aceptación.
- *servicio*: nombre o descripción del servicio.
- *estudio_mercado_serv*: expresa si el servicio tiene estudios de mercados realizados.
- *Solución*: nombre o descripción de la solución.
- *estudio_mercado_sol*: es para conocer si se realizaron estudios de mercado a las soluciones.

- *solución_aceptada*: representa si la solución fue aceptada por el cliente.

Perspectivas “Cliente”:

- *id_cliente*: es la clave primaria de la tabla “Cliente”, representa un cliente en particular.
- *nombre*: representa al cliente en el marco nacional e internacional.
- *cta_habilitada*: indica si el cliente tiene una cuenta habilitada.
- *Id_cuentacontable*: representa la cuenta contable asociada al cliente, la cual se utiliza para imputar los movimientos contables que este genere.
- *Id_clasificación*: representa a través de una clave la clasificación del cliente: Muy Bueno, Bueno, Regular, Potencial.
- *eliminado*: indica si el cliente fue eliminado o no. Si fue eliminado no figura en las listas de clientes actuales.

Perspectiva “Proceso Interno”:

- *id_proceso*: es la clave primaria de la tabla “Proceso Interno”, representa a un proceso en particular.
- *proyecto*: representa los datos del proyecto identificándolo en el área de trabajo.
- *hitos_cumplidos*: es para conocer si el proyecto cumplió los hitos del programa de mejoras.
- *proceso_planificación*: es para saber si el proyecto cumple con los procesos de planificación y control del Programa de Mejora implantado.
- *estado_proyecto*: define si el proyecto está atrasado o no.
- *nombre_profesor*: representa una parte de la identificación del profesor más detallada.
- *profesor_vinculado*: es para conocer si el profesor de la especialidad está vinculado a la producción.

Perspectiva “Tiempo”:

- *año*: representa el año en el que se evalúan las informaciones.
- *Trimestre*: representa el tiempo en el que se evalúan las informaciones, en este caso serían tres meses.

2.5. Modelo lógico del Almacén de Datos del CESIM

El modelo lógico¹³ de la estructura del almacén de datos, tiene como base el modelo conceptual que ya ha sido creado. Para ello, primero se define el tipo de modelo que se utiliza y luego se lleva a cabo las acciones propias al caso, para diseñar las tablas de dimensiones y de hechos. Finalmente, se realizan las uniones pertinentes entre estas tablas.

2.5.1. Tipo de modelo lógico del almacén de datos

Existen diferentes tipos de esquemas que se utilizan para contener la estructura del depósito de datos. Los mismos se definen como: esquema en estrella, constelación o copo de nieve.

Esquema en Estrella

El esquema en estrella es el más simple de interpretar y optimiza los tiempos de respuesta ante las consultas de los usuarios. Este modelo es soportado por casi todas las herramientas de consulta y análisis, y los metadatos son fáciles de documentar y mantener, por lo que se utiliza en este trabajo.

2.5.2. Tablas de dimensiones

En este paso se diseñan las tablas de dimensiones que forman parte del almacén de datos. Para los tres tipos de esquemas, cada perspectiva definida en el modelo conceptual constituye una tabla de dimensión.

¹³ Modelo Lógico: representación de una estructura de datos, que puede procesarse y almacenarse en algún SGBD.

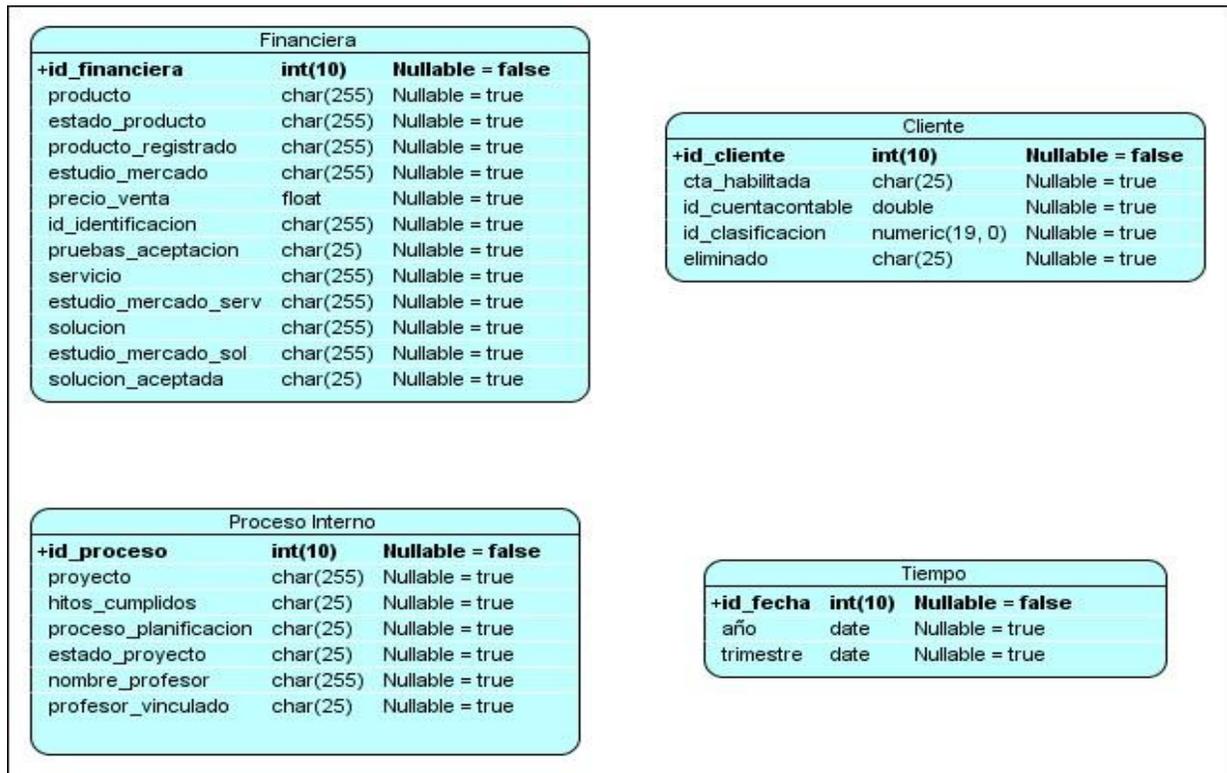


Figura 2.5 Tabla de dimensiones para representar las perspectivas del CESIM.

2.5.3. Tablas de hechos

En este paso se definen las tablas de hechos, que son las que contendrán los hechos a través de los cuales se construirán los indicadores de estudio. A continuación se presentan las tablas de hechos para las diferentes áreas de análisis (Grupo de Mercadotecnia y departamentos de producción).

Tablas de hechos para el Grupo de Mercadotecnia:

Capítulo II: Análisis del Almacén de Datos

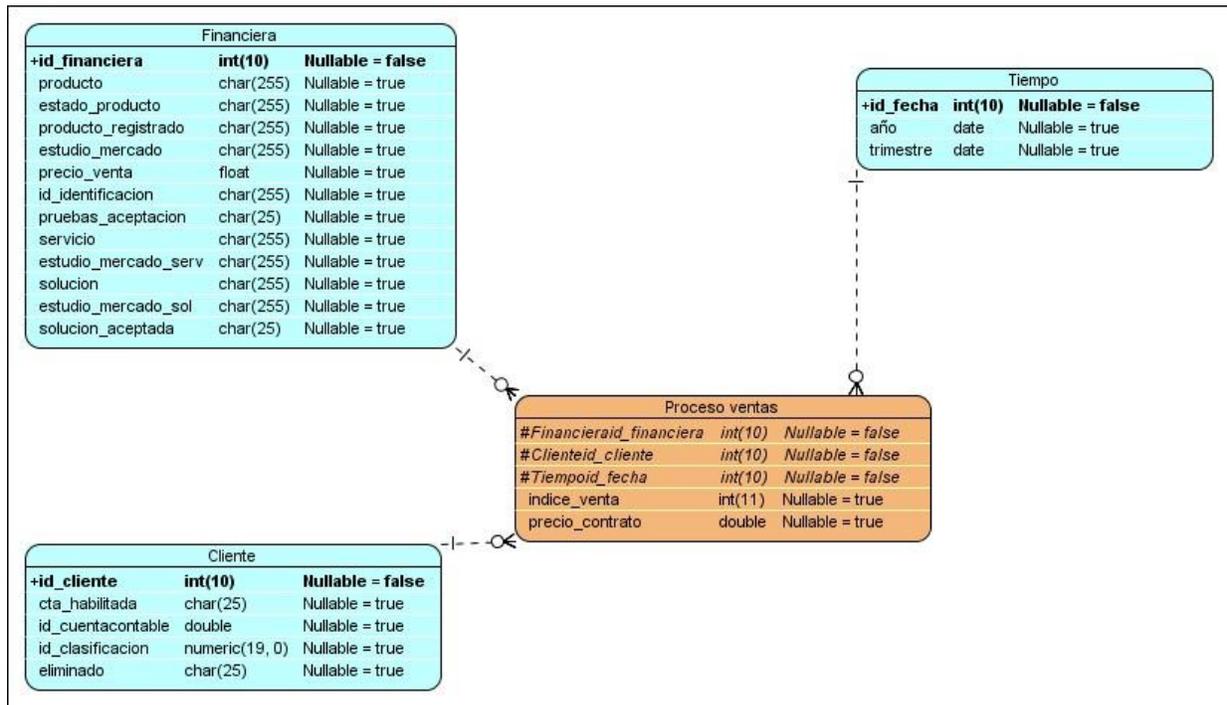


Figura 2.6 Tabla de hecho 1 del Grupo de Mercadotecnia.

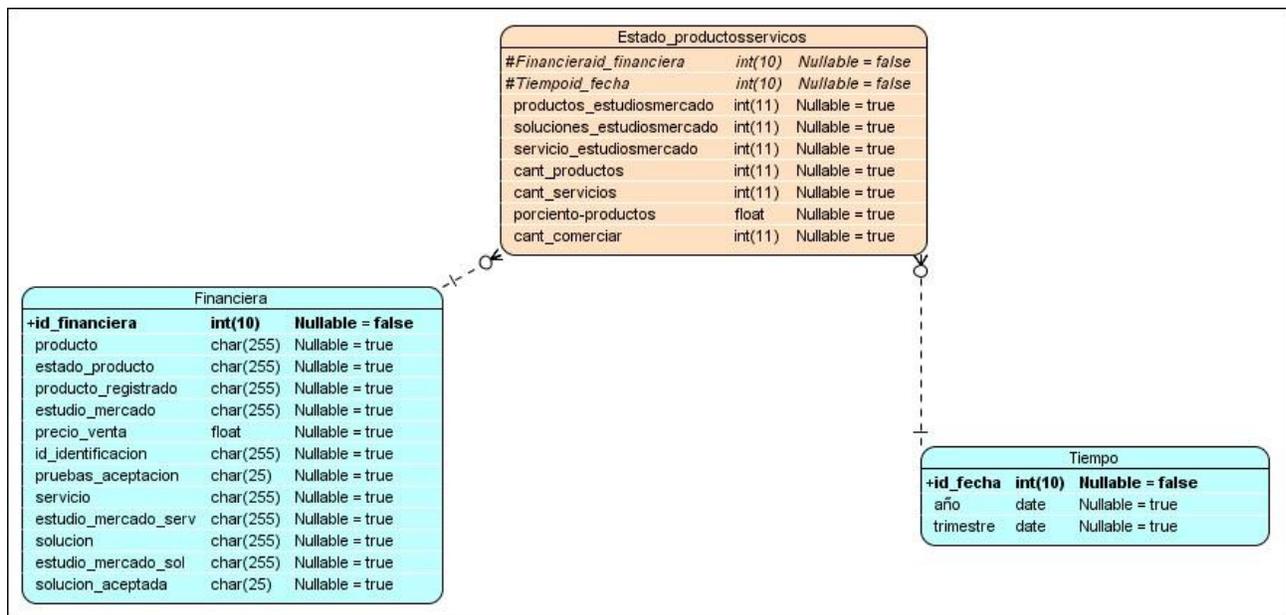


Figura 2.7 Tabla de hecho 2 del Grupo de Mercadotecnia.

Capítulo II: Análisis del Almacén de Datos

Tablas de hechos para los departamentos de producción:

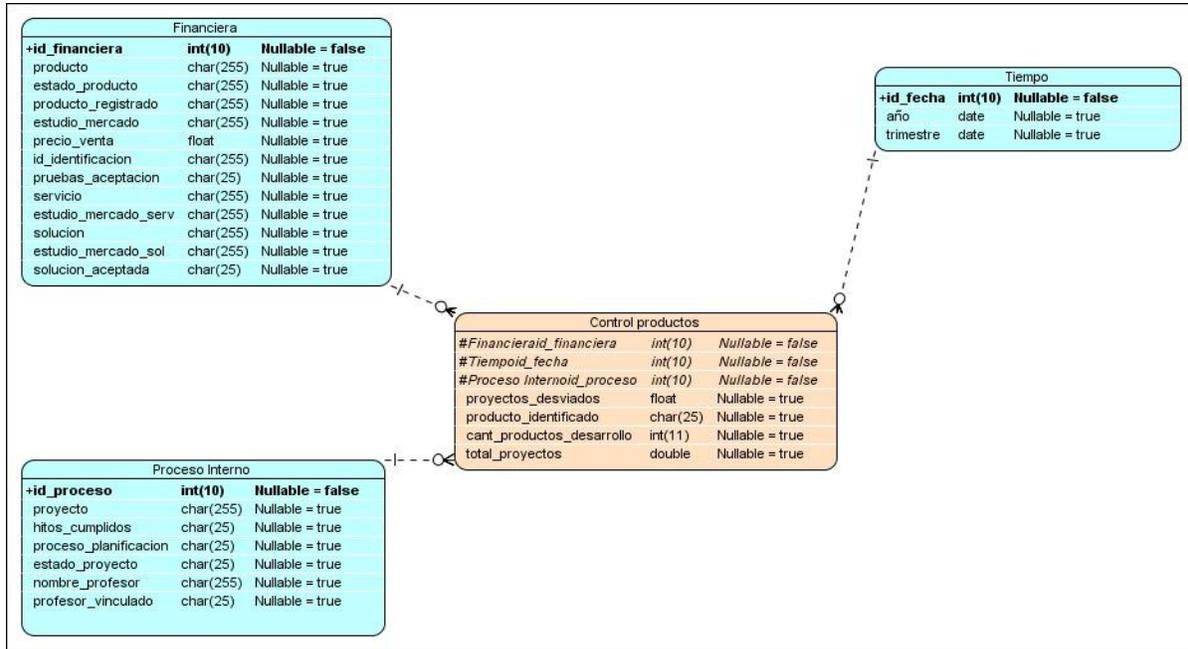


Figura 2.8 Tabla de hecho 1 de los departamentos de producción.

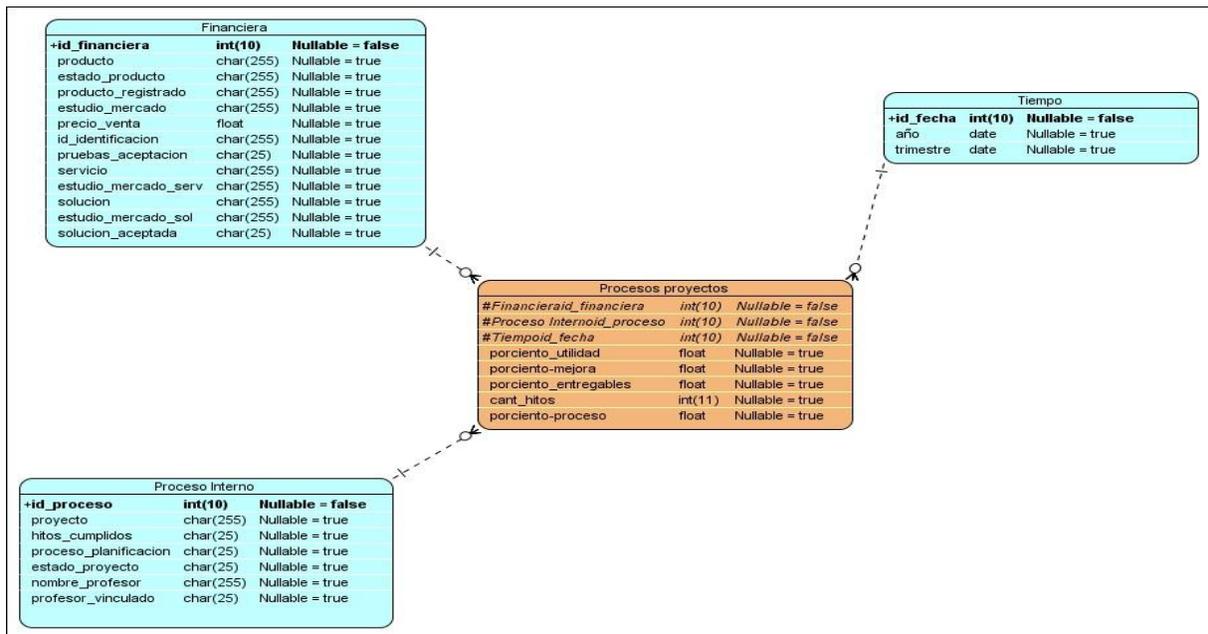


Figura 2.9 Tabla de hecho 2 de los departamentos de producción.

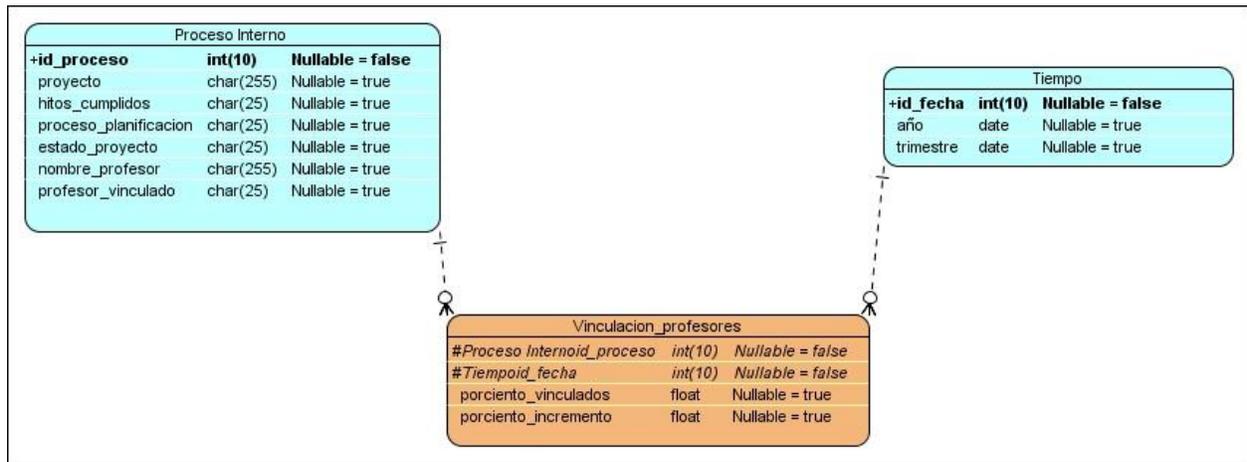


Figura 2.10 Tabla de hecho 3 de los departamentos de producción.

En este capítulo se realiza un análisis crítico de los procesos actuales, lo que permite precisar el estado actual de los procesos que se llevan a cabo en el Grupo de Mercadotecnia y los departamentos de producción del Centro de Informática Médica. Además de poder obtener un mejor resultado de los objetivos expuestos y plantear las bases para la implementación de un sistema capaz de tomar decisiones con eficacia y rapidez.

Capítulo 3. Diseño del Almacén de Datos

En este capítulo se realiza el diseño de los procesos de Extracción, Transformación y Carga (ETL) para el Almacén de Datos del Centro de Informática Medica (CESIM). Se emplea una arquitectura basada en el diseño de Inmon, además del diagrama de actividades que describe cómo se realiza el proceso de extracción, transformación y carga de los datos. Se implementan los cubos multidimensionales que dan como resultado los reportes que muestran el comportamiento de los objetivos estratégicos del centro.

3.1. Arquitectura para el CESIM

Existen dos filosofías conocidas acerca de diseño del almacén de datos: la de Inmon y la de Ralph Kimball. Inmon plantea el siguiente diseño:

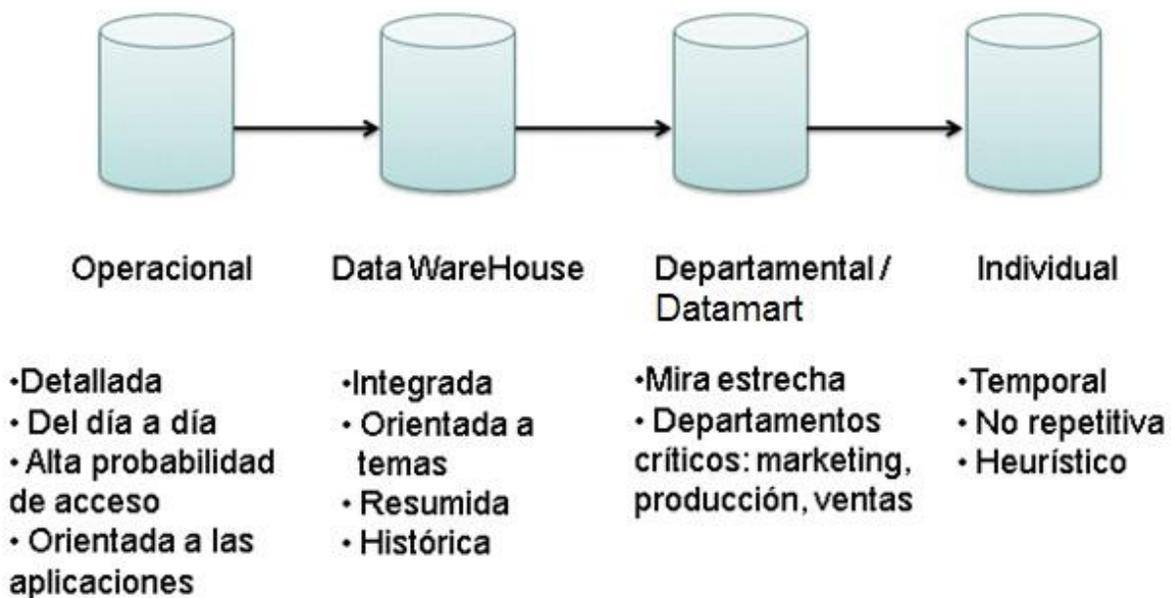


Figura 3.1 Niveles de arquitectura.

Según Inmon, existen cuatro niveles de datos en la arquitectura: el nivel operacional, el atómico o el de almacén de datos, el departamental o nivel de Datamart y el nivel individual. El nivel operacional guarda los datos de las aplicaciones y sirve para los usuarios que necesitan un buen tiempo de respuesta en sus transacciones diarias.

Capítulo III: Diseño del Almacén de Datos

El nivel de almacén guarda datos integrados e históricos que no pueden ser modificados. El nivel departamental contiene datos moldeados según los requerimientos del usuario final, en una forma que satisfaga las necesidades del departamento. En el nivel individual es donde son realizados los procesos heurísticos.

Para este trabajo se decide utilizar la arquitectura planteada por Inmon, ya que se necesita tener los datos, provenientes de las distintas áreas, integrados para asegurar la consistencia de los reportes mostrados. Asimismo, se necesita tener un histórico de los datos del CESIM, por lo que este modelo se adapta mejor a este proyecto.

Después de lo expuesto se propone la arquitectura del Almacén de Datos para el CESIM, en la cual se tiene en cuenta los departamentos analizados.

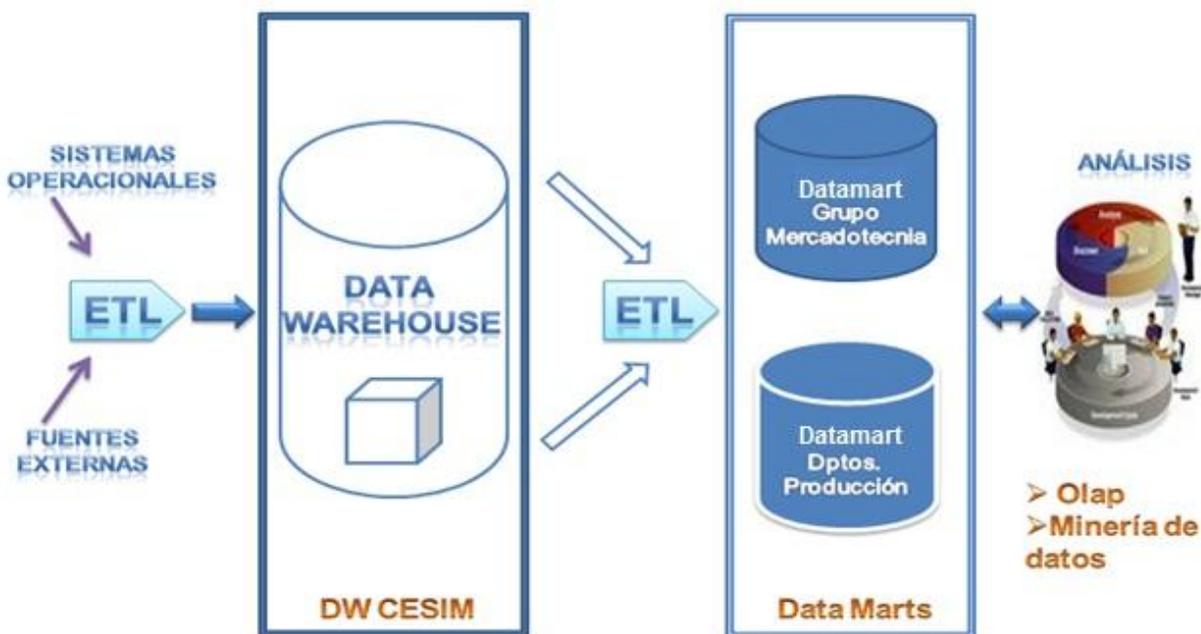


Figura 3.2 Arquitectura del almacén del CESIM.

3.2. Diseño de los procesos ETL

El almacén de datos contiene metadatos sobre los procesos de extracción, transformación y carga de los datos de la base de datos operacional. Los procesos se ejecutarán periódicamente

Capítulo III: Diseño del Almacén de Datos

mediante la actualización del almacén de modo que la información continúe siendo relevante y oportuna para la toma de decisiones.

Los datos fuente para la aplicación Inteligencia de Negocio vendrán de varias plataformas. El propósito de esta etapa es fusionar los datos de las plataformas en un formato para el almacén de datos.

En la **Figura 3.3** se describe el proceso ETL (extracción, transformación y carga), para el almacén del CESIM, donde primeramente se lleva a cabo la extracción de los datos de las fuentes correspondientes y a partir de ahí se realiza la transformación, cuya tarea principal es la limpieza de los datos, para evitar redundancia, y otros problemas que pueden provocar que el almacén se vuelva más lento a la hora de consultar los reportes. Después se realiza la carga de los datos y se guardan en los Datamart creados para cada departamento.

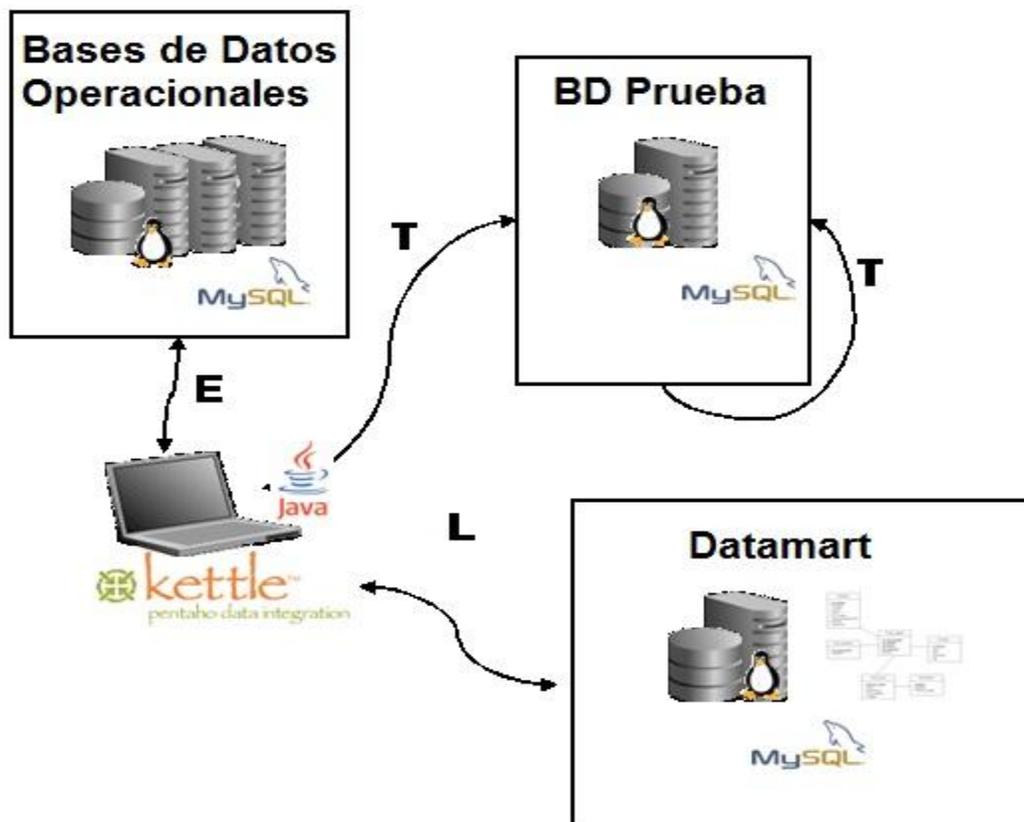


Figura 3.3 Proceso de extracción, transformación y carga de los datos.

3.2.1. Diagrama de actividades para explicar el proceso ETL

El diseño ETL se realiza con la interrelación de las actividades del proceso de extracción, transformación y carga de los datos. El diagrama comienza con la extracción de los datos de las fuentes externas y sucesivamente continúa con una serie de pasos que finalizan en la carga de los datos en los Datamart.

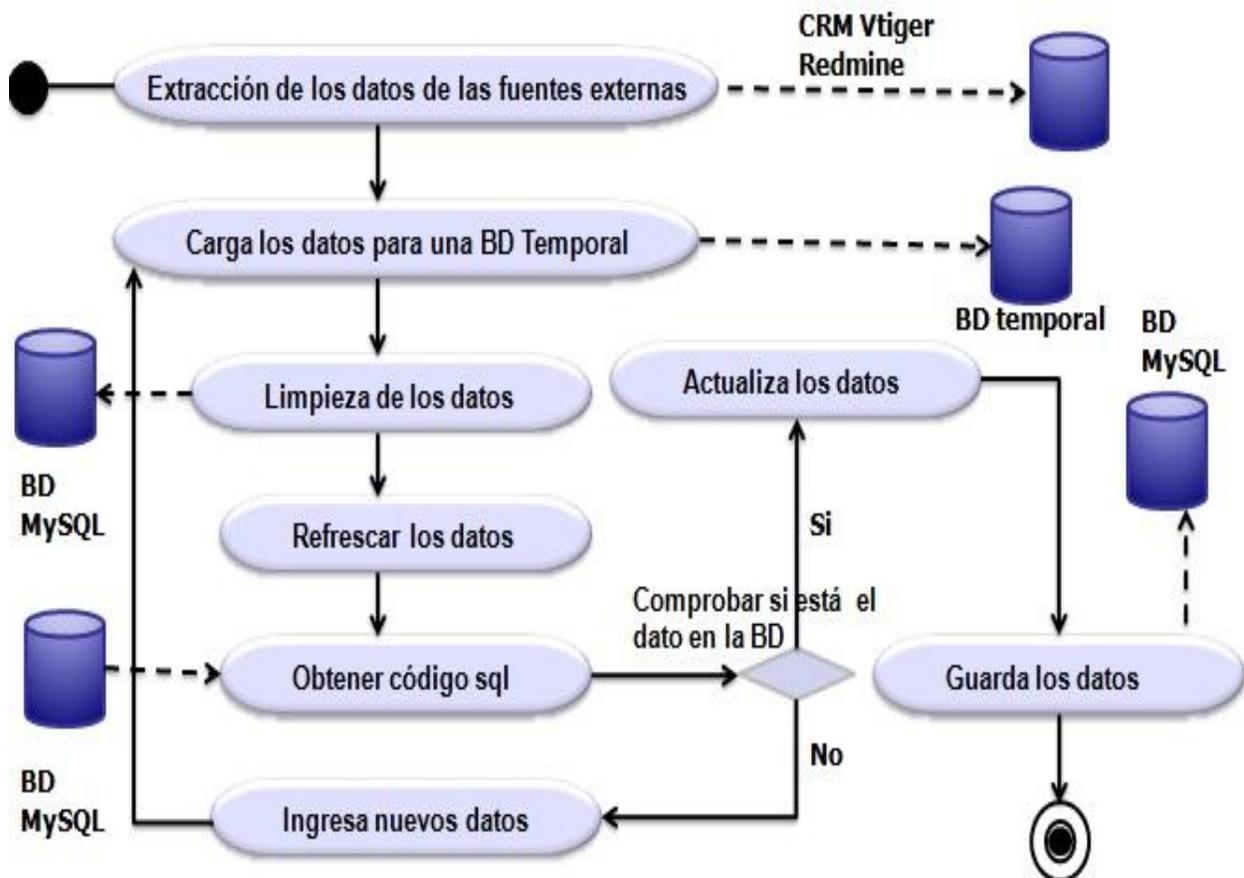


Figura 3.4 Proceso ETL. Diagrama de actividades.

3.2. Creación de Cubos Multidimensionales

El cubo multidimensional está basado en el modelo lógico diseñado en el capítulo anterior, donde se especifican las tablas de dimensiones y las tablas de hechos; estas últimas originan los cubos. La creación de un cubo tiene las siguientes finalidades:

Capítulo III: Diseño del Almacén de Datos

- Ejemplificar la creación de cubos multidimensionales.
- Propiciar la correcta distinción entre hechos de una tabla de hechos e indicadores de un cubo.
- Propiciar la correcta distinción entre campos de una tabla de dimensión y atributos de un cubo.

A partir del modelo lógico se crearon 4 cubos, cada uno de los cuales está orientado a un tipo de análisis en particular según los objetivos estratégicos a cumplir. Los cubos se crean en un mismo Schema y pueden coexistir sin ningún inconveniente.

3.3.1. Publicación de los cubos en Pentaho.

Para crear los reportes se utilizan los objetivos estratégicos del CESIM, tomando la información procedente del Grupo de Mercadotecnia y los departamentos de producción. De esta forma se calculan los indicadores que intervienen en cada objetivo, mediante la herramienta Pentaho que permite que el usuario maneje los reportes según sus necesidades. (Ver anexos 4, 5, 6,7)

Grupo de Mercadotecnia:

- 1- Realizar estudios de mercado para el 80% de los productos, servicios y soluciones informáticas integrales identificadas.
- 2- Conformar el catálogo de productos, servicios y soluciones informáticas integrales.

Departamentos de producción:

- 1- Identificar al menos cinco nuevos productos de informática médica.
- 2- Lograr que el 50% de las soluciones aceptadas por los clientes sean desplegadas en el período en cuestión.

En este capítulo se realiza el diseño de los procesos ETL, los cuales dan seguimiento al análisis de los datos, y al diseño lógico realizado en el capítulo anterior. Este diseño permite tener una base para realizar la implementación ETL, que da paso a la implementación del Almacén de Datos del CESIM. Se crean los cubos según los objetivos estratégicos actuales del centro y se publican en la herramienta Pentaho para la utilidad de los interesados.

Conclusiones

Conclusiones

Al concluir la investigación se obtiene:

- Del estudio del arte la herramienta Pentaho, la cual es esencial para llevar a cabo la toma de decisiones en el CESIM.
- El diseño del Almacén de Datos del CESIM con la información específica del grupo de Mercadotecnia y de los departamentos de producción.
- De esta forma se logra dar seguimiento a los objetivos estratégicos y medir el desempeño del trabajo a partir de indicadores.
- Además, el uso del Almacén de Datos constituye un nuevo paso en lo referente a las nuevas tecnologías y el avance del mundo de la Inteligencia de Negocios.

Recomendaciones

Recomendaciones

Se recomienda validar los resultados en el Grupo de Mercadotecnia y los departamentos de producción, para poder integrar los objetivos estratégicos de todas las áreas del CESIM al almacén de datos.

Además, sería fundamental tener todas las bases de datos operacionales necesarias para el desarrollo del almacén, con todos los atributos que den respuesta a cada objetivo estratégico planteado.

Continuar con la investigación para llegar a realizar la implementación del almacén de datos para el CESIM.

Referencias Bibliográficas

Referencias Bibliográficas

1. **Briceño, Edgar Armando Vega.** Los Sistemas de Información y su importancia para las organizaciones y empresas. [En línea] 2008.
<http://www.gestiopolis.com/Canales4/mkt/simparalass.htm>.
2. Sistemas de Soporte a la Decisión (DSS). [En línea] 2007.
http://www.sinnexus.com/business_intelligence/sistemas_soporte_decisiones.aspx.
3. **Contreras, Laura.** BI: Inteligencia de Negocios. [En línea] 8 de Junio de 2009.
http://www.sinnexus.com/business_intelligence/sistemas_soporte_decisiones.aspx.
4. Data Warehouse. [En línea] 2007.
http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.
5. **Domínguez, Antonio Blasco.** Data Warehousing. [En línea] 14 de Abril de 2003.
http://www.programacion.com/articulo/data_warehousing_201.
6. **Bressán, Griselda E.** "Objetivos del Data Warehouse.". [En línea] 2003.
<http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MineriaDatosBressan.htm#Objetivos%20del%20Data%20Warehouse>.
7. Bill Inmon. [En línea] 26 de Mayo de 2009. <http://kle.sisorg.com.mx/articulo02.html>.
8. Data Warehouse. [En línea] 2007.
http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.
9. **Algorry, Aldo Marcelo.** Data Warehousing. [En línea] <http://www.sqlmax.com/dataw1.asp>.
10. Data Warehouse. Aspectos Teóricos. [En línea] 2007. <http://www.sqlmax.com/dataw1.asp>.
11. Datamart. [En línea] 2007. http://www.sinnexus.com/business_intelligence/datamart.aspx.
12. Características del datamart. [En línea] 2007.
http://www.sinnexus.com/business_intelligence/datamart.aspx.
13. Datamart OLAP. [En línea] 2007.
http://www.sinnexus.com/business_intelligence/datamart.aspx.

Referencias Bibliográficas

14. **Datamart** OLTP. [En línea] 2007.
http://www.sinnexus.com/business_intelligence/datamart.aspx.
15. **Estructura** Multidimensional. [En línea] 2007.
http://www.sinnexus.com/business_intelligence/datamart.aspx.
16. **Orol, Alfredo Martínez.** OLAP y el diseño de cubos. [En línea] Marzo de 2007.
<http://www.gestiopolis.com/canales8/ger/olap-online-analytic-processing.htm>.
17. **DAEDALUS.** Minería de Datos. [En línea] 2007. <http://www.daedalus.es/mineria-de-datos/>.
18. **Velasco, Roberto Hernando.** Estructura de un almacén de datos. [En línea]
<http://www.rhernando.net/modules/tutorials/doc/bd/dw.html>.
19. **Domínguez, Antonio Blasco.** Data Warehousing. [En línea] 14 de Abril de 2003.
http://www.programacion.com/articulo/data_warehousing_201/6.
20. **Curto, Josep.** Arquitectura de un datawarehouse. [En línea] 3 de Junio de 2007.
<http://informationmanagement.wordpress.com/2007/06/03/arquitectura-de-un-datawarehouse/>.
21. **Victoria.** Manual de MySQL. [En línea] 25 de Febrero de 2009.
<http://www.definicionabc.com/tecnologia/mysql.php>.
22. **Cárdenas, Luis Alejandro Esteban.** SQL Server. [En línea] 2008.
<http://alejandroesteban.wordpress.com/category/sql-server-2005/>.
23. **Patiño, José Juan Ramírez.** Desventajas de SQL. [En línea] 2010.
<http://www.slideshare.net/JJRP80/sql-server-3390264>.
24. **Pecos, Daniel.** Postgresql. [En línea] 2007.
http://danielpecos.com/docs/mysql_postgres/index.html.
25. **PostgreSQL** , motor de base de datos. [En línea] 15 de Agosto de 2007.
<http://buenmaster.com/?a=545>.
26. **Carcaterísticas de Postgre.** [En línea] 15 de Agosto de 2007. <http://buenmaster.com/?a=545>.

Referencias Bibliográficas

27. Visual paradigm. [En línea] 5 de Marzo de 2007. http://www.freedownloadmanager.org/es/downloads/Visual_Paradigm_for_UML_%28LE%29_%5BWindows%5D_14728_p/.
28. Características principales de visual Paradigm. [En línea] 5 de Marzo de 2007. http://www.freedownloadmanager.org/es/downloads/Visual_Paradigm_for_UML_%28LE%29_%5BWindows%5D_14728_p/.
29. **Herrera, Cristhian.** Gestión de relación con el cliente. [En línea] 2006. <http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=crm4>.
30. **Navarro, Eduardo.** CRM. [En línea] Junio de 2002. <http://www.gestiopolis.com/canales/demarketing/articulos/37/crmcliente.htm>.
31. **Complusoft.** Sugar CRM, gestión de clientes. [En línea] 2008. <http://www.complusoft.es/soluciones/sugar-crm>.
32. —. Sugar CRM. [En línea] 2008. <http://www.complusoft.es/soluciones/sugar-crm>.
33. **Dueñas, Joel Barrios.** Cómo instalar y configurar vtiger CRM. [En línea] 5 de Diciembre de 2007. <http://www.alcancelibre.org/staticpages/index.php/como-vtigercrm>.
34. Características del CRM. [En línea] 5 de Diciembre de 2007. <http://www.alcancelibre.org/staticpages/index.php/como-vtigercrm>.
35. La plataforma Pentaho Open Source Business Intelligence. [En línea] 2008. <http://pentaho.almacen-datos.com/>.
36. **Salinas, Alejandro.** Mondrian. [En línea] 12 de Marzo de 2008. <http://www.gravitar.biz/index.php/bi/introduccion-pentaho-parte-1/>.
37. —. Pentaho Reporting. [En línea] 12 de Marzo de 2008. <http://www.gravitar.biz/index.php/bi/introduccion-pentaho-parte-1/>.
38. —. Kettle. [En línea] 12 de Marzo de 2008. <http://www.gravitar.biz/index.php/bi/introduccion-pentaho-parte-1/>.

Referencias Bibliográficas

- 39.—.Weka: Minería de datos. [En línea] 12 de Marzo de 2008.
<http://www.gravitar.biz/index.php/bi/introduccion-pentaho-parte-1/>.
- 40.**Freira, I.** Manual de instalación de Redmine. [En línea] 29 de Abril de 2009.
http://www.mancomun.org/es/no_cache/actualidade/detalledenova/nova/ousli-libera-un-manual-de-instalacion-de-redmine/.
- 41.**Iglesias, Ridosbey Milián.** Redmine. [En línea] 26 de Febrero de 2010.
<http://www.gestiopolis.com/administracion-estrategia/herramientas-gestion-proyectos-software.htm>.
- 42.**Dario, Ing. Bernabeu Ricardo.** Metodología Hefesto. [En línea] 17 de Enero de 2009.
<http://www.dataprix.com/es/data-warehousing-hefesto>.
- 43.**Carlos.** Características de Hefesto. [En línea] 28 de Diciembre de 2005.
<http://www.dataprix.com/es/hefesto-metodologia-propia-para-la-construccion-un-data-warehouse>.
44. **Bernabeu, Dario.** *Metodología Hefesto.*

Bibliografía

Bibliografía

Algorry, Aldo Marcelo. Data Warehousing. [En línea] <http://www.sqlmax.com/dataw1.asp>.

Bernabeu, Dario. Metodología Hefesto.

Bill Inmon. [En línea] 26 de Mayo de 2009. <http://kle.sisorg.com.mx/articulo02.html>.

Bressán, Griselda E. "Objetivos del Data Warehouse.". [En línea] 2003. <http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MineriaDatosBressan.htm#Objetivos%20del%20Data%20Warehouse>.

Briceño, Edgar Armando Vega. Los Sistemas de Información y su importancia para las organizaciones y empresas. [En línea] 2008. <http://www.gestiopolis.com/Canales4/mkt/simparalas.htm>.

Características del CRM. [En línea] 5 de Diciembre de 2007. <http://www.alcancelibre.org/staticpages/index.php/como-vtigercrm>.

Características del datamart. [En línea] 2007. http://www.sinnexus.com/business_intelligence/datamart.aspx.

Características principales de visual Paradigm. [En línea] 5 de Marzo de 2007. http://www.freedownloadmanager.org/es/downloads/Visual_Paradigm_for_UML_%28LE%29_%5BWindows%5D_14728_p/.

Características de Postgre. [En línea] 15 de Agosto de 2007. <http://buenmaster.com/?a=545>.

Cárdenas, Luis Alejandro Esteban. SQL Server. [En línea] 2008. <http://alejandroesteban.wordpress.com/category/sql-server-2005/>.

Carlos. Características de Hefesto. [En línea] 28 de Diciembre de 2005. <http://www.dataprix.com/es/hefesto-metodologia-propia-para-la-construccion-un-data-warehouse>.

Complusoft. Sugar CRM, gestión de clientes. [En línea] 2008. <http://www.complusoft.es/soluciones/sugar-crm>.

Bibliografía

Contreras, Laura. BI: Inteligencia de Negocios. [En línea] 8 de Junio de 2009. http://www.sinnexus.com/business_intelligence/sistemas_soporte_decisiones.aspx.

Curto, Josep. Arquitectura de un datawarehouse. [En línea] 3 de Junio de 2007. <http://informationmanagement.wordpress.com/2007/06/03/arquitectura-de-un-datawarehouse/>.

DAEDALUS. Minería de Datos. [En línea] 2007. <http://www.daedalus.es/mineria-de-datos/>.

Dario, Ing. Bernabeu Ricardo. Metodología Hefesto. [En línea] 17 de Enero de 2009. <http://www.dataprix.com/es/data-warehousing-hefesto>.

Data Warehouse. [En línea] 2007. http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.

Data Warehouse. [En línea] 2007. http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.

Data Warehouse. Aspectos Teóricos. [En línea] 2007. <http://www.sqlmax.com/dataw1.asp>.

Datamart OLAP. [En línea] 2007. http://www.sinnexus.com/business_intelligence/datamart.aspx.

Datamart OLTP. [En línea] 2007. http://www.sinnexus.com/business_intelligence/datamart.aspx.

Datamart. [En línea] 2007. http://www.sinnexus.com/business_intelligence/datamart.aspx.

Domínguez, Antonio Blasco. Data Warehousing. [En línea] 14 de Abril de 2003. http://www.programacion.com/articulo/data_warehousing_201.

Domínguez, Antonio Blasco. Data Warehousing. [En línea] 14 de Abril de 2003. http://www.programacion.com/articulo/data_warehousing_201/6.

Dueñas, Joel Barrios. Cómo instalar y configurar vtiger CRM. [En línea] 5 de Diciembre de 2007. <http://www.alcancelibre.org/staticpages/index.php/como-vtigercrm>.

Bibliografía

Estructura Multidimensional. [En línea] 2007.
http://www.sinnexus.com/business_intelligence/datamart.aspx.

Freira, I. Manual de instalación de Redmine. [En línea] 29 de Abril de 2009.
http://www.mancomun.org/es/no_cache/actualidade/detalledenova/nova/ousli-libera-un-manual-de-instalacion-de-redmine/.

Herrera, Cristhian. Gestión de relación con el cliente. [En línea] 2006.
<http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=crm4>.

Iglesias, Ridosbey Milián. Redmine. [En línea] 26 de Febrero de 2010.
<http://www.gestiopolis.com/administracion-estrategia/herramientas-gestion-proyectos-software.htm>.

Kettle. [En línea] 12 de Marzo de 2008. <http://www.gravitar.biz/index.php/bi/introduccion-pentaho-parte-1/>.

La plataforma Pentaho Open Source Business Intelligence. [En línea] 2008.
<http://pentaho.almacen-datos.com/>.

Navarro, Eduardo. CRM. [En línea] Junio de 2002.
<http://www.gestiopolis.com/canales/demarketing/articulos/37/crmcliente.htm>.

Orol, Alfredo Martínez. OLAP y el diseño de cubos. [En línea] Marzo de 2007.
<http://www.gestiopolis.com/canales8/ger/olap-online-analytic-processing.htm>.

Patiño, José Juan Ramírez. Desventajas de SQL. [En línea] 2010.
<http://www.slideshare.net/JJRP80/sql-server-3390264>.

Pecos, Daniel. Postgresql. [En línea] 2007.
http://danielpecos.com/docs/mysql_postgres/index.html.

Pentaho Reporting. [En línea] 12 de Marzo de 2008.
<http://www.gravitar.biz/index.php/bi/introduccion-pentaho-parte-1/>.

Bibliografía

PostgreSQL , motor de base de datos. [En línea] 15 de Agosto de 2007.
<http://buenmaster.com/?a=545>.

Salinas, Alexandro. Mondrian. [En línea] 12 de Marzo de 2008.
<http://www.gravitar.biz/index.php/bi/introduccion-pentaho-parte-1/>.

Sistemas de Soporte a la Decisión (DSS). [En línea] 2007.
http://www.sinnexus.com/business_intelligence/sistemas_soporte_decisiones.aspx.

Sugar CRM. [En línea] 2008. <http://www.complusoft.es/soluciones/sugar-crm>.

Velasco, Roberto Hernando. Estructura de un almacén de datos. [En línea]
<http://www.rhernando.net/modules/tutorials/doc/bd/dw.html>.

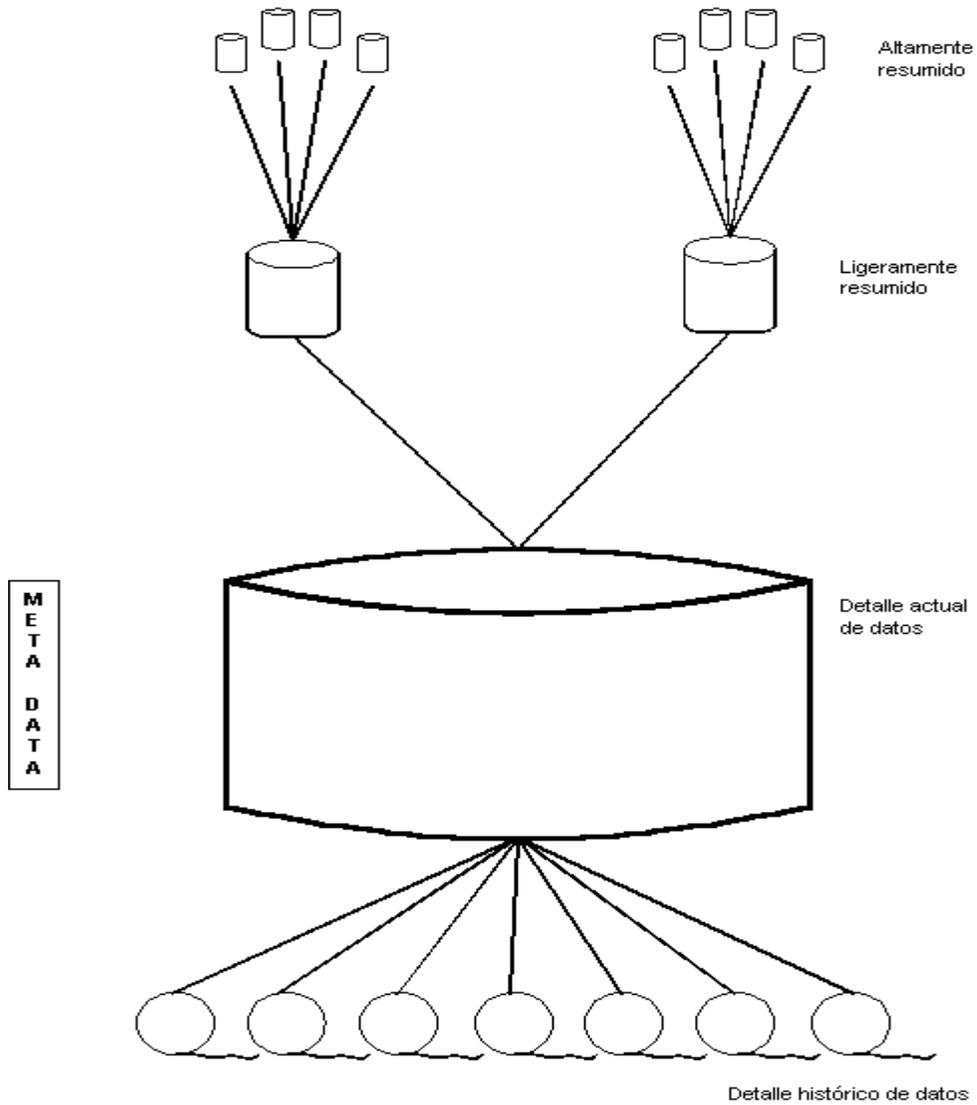
Victoria. Manual de MySQL. [En línea] 25 de Febrero de 2009.
<http://www.definicionabc.com/tecnologia/mysql.php>.

Visual paradigm. [En línea] 5 de Marzo de 2007.
http://www.freedownloadmanager.org/es/downloads/Visual_Paradigm_for_UML_%28LE%29_%5BWindows%5D_14728_p/.

Weka: Minería de datos. [En línea] 12 de Marzo de 2008.
<http://www.gravitar.biz/index.php/bi/introduccion-pentaho-parte-1/>.

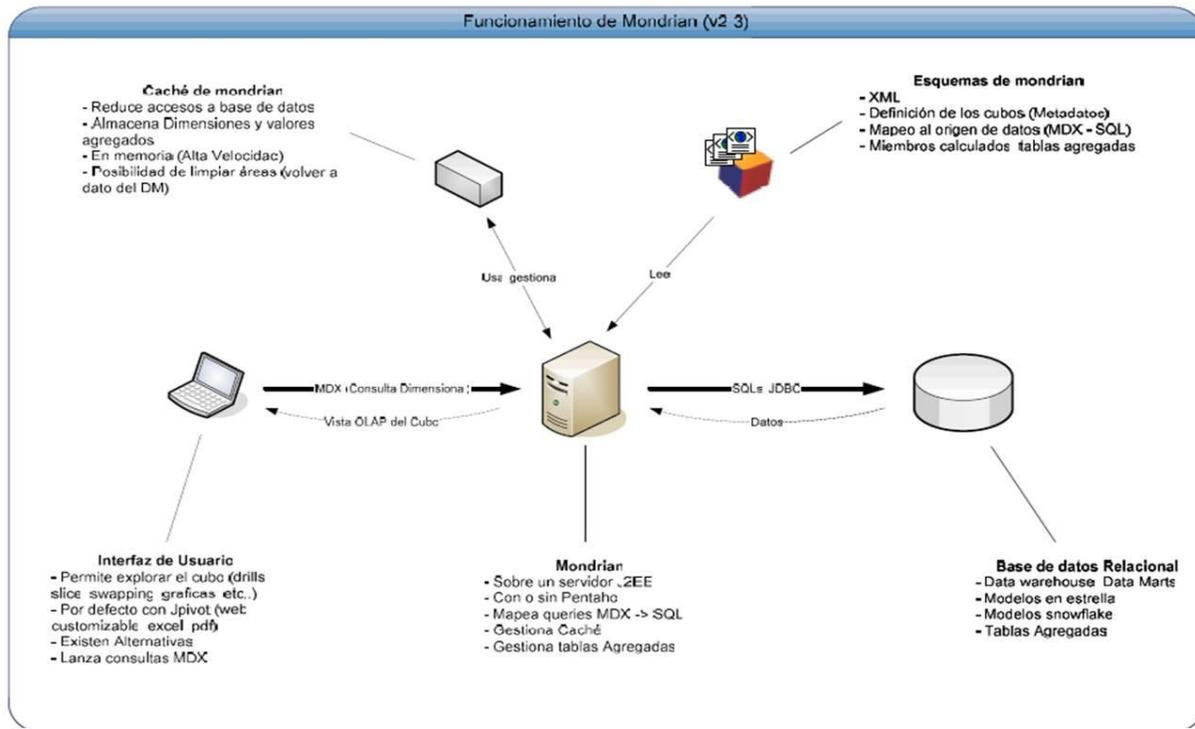
Anexos

- Anexo 1. Estructura de los datos en un almacén de datos.

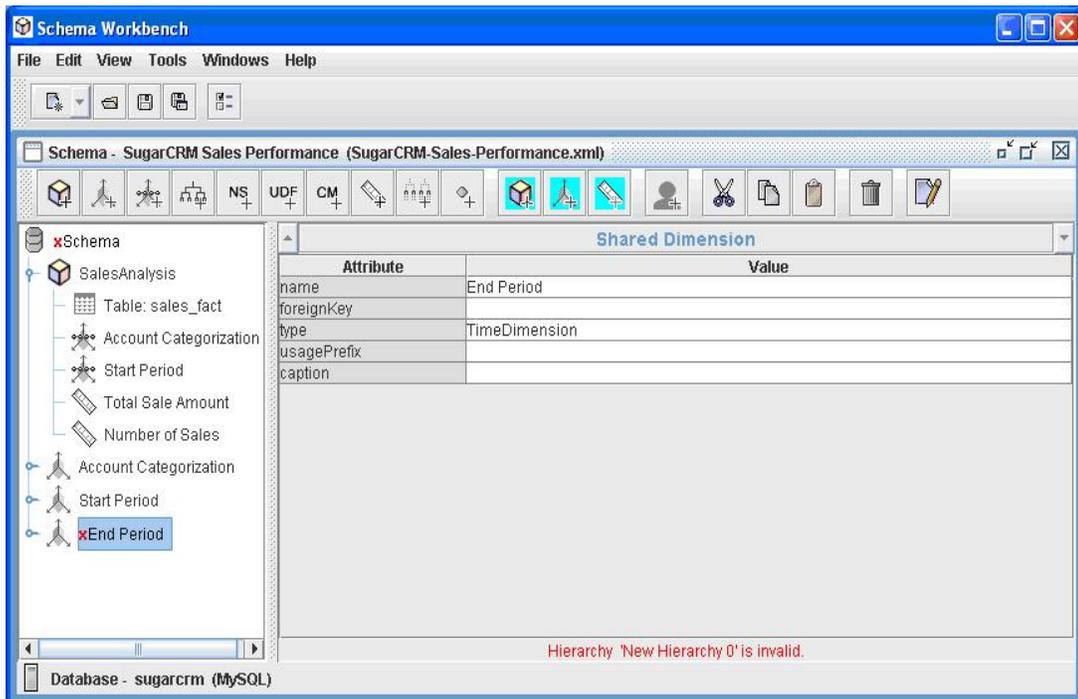


Anexos

• Anexo 2. Funcionamiento de Mondrian (Uno de los módulos de la herramienta Pentaho).

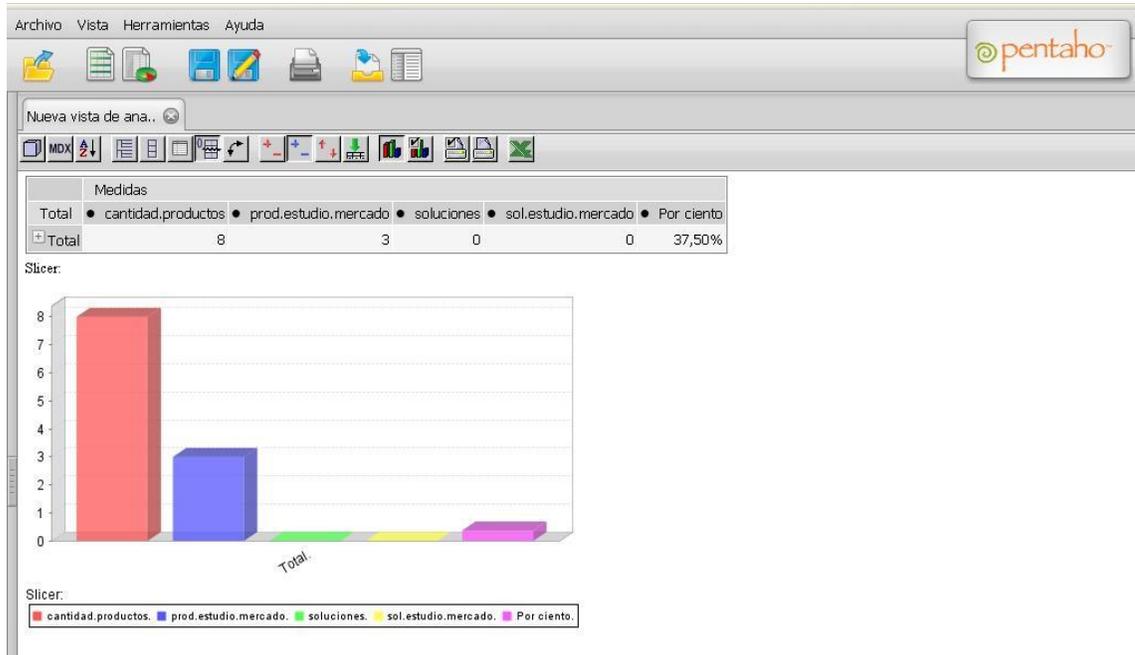


• Anexo 3. Mondrian Schema Workbench.

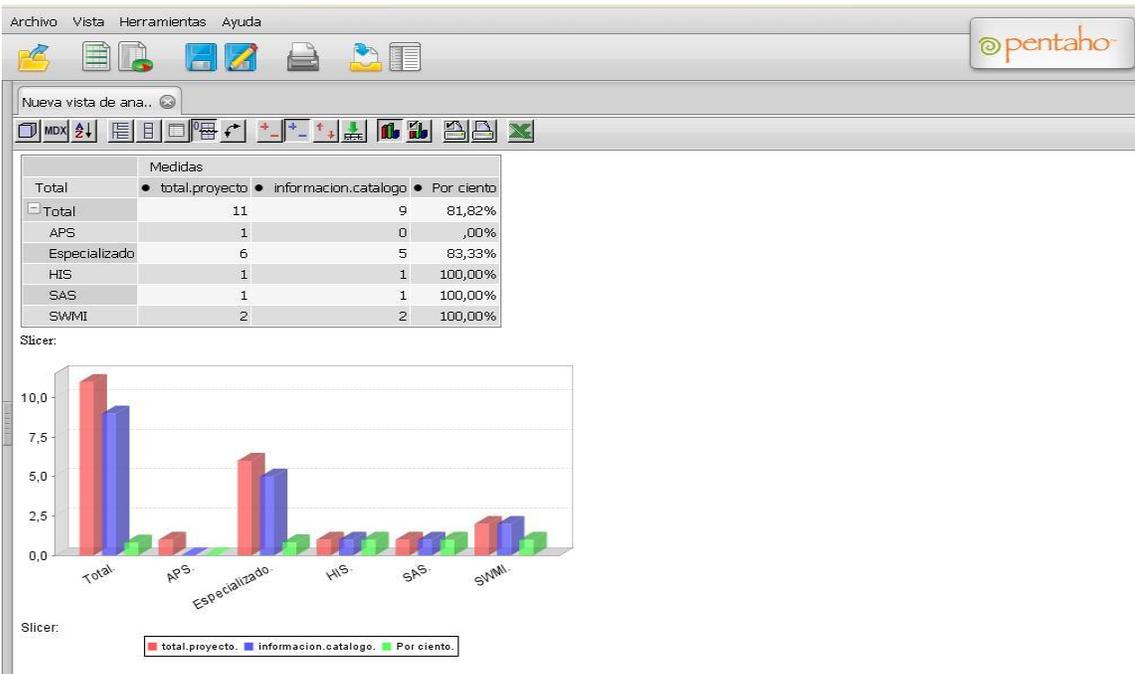


Anexos

- **Anexo 4.** Cubo “Estudios de mercado” que cumple con el primer objetivo estratégico del Grupo de Mercadotecnia.

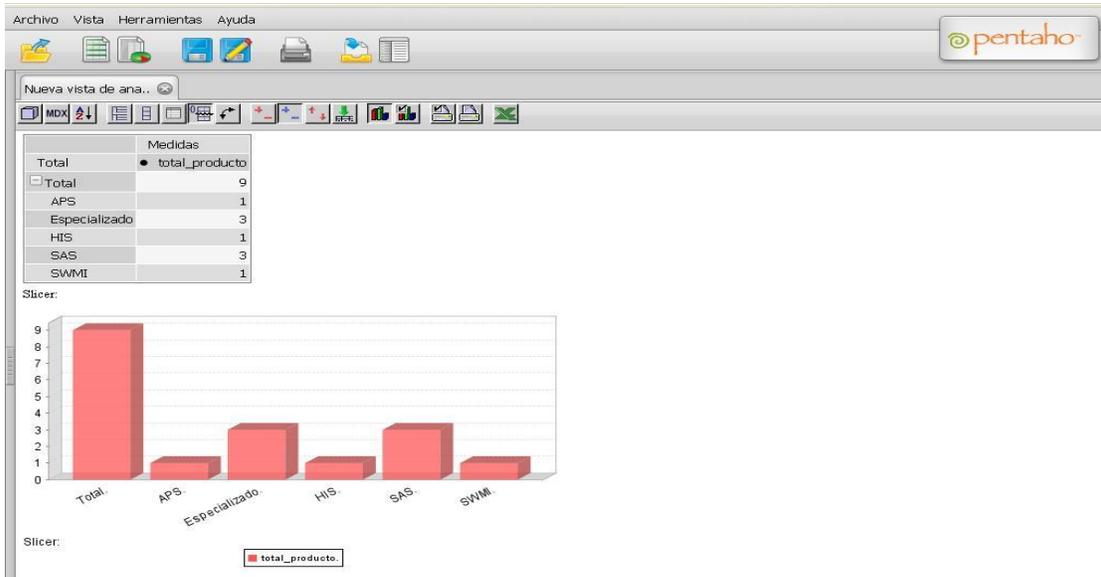


- **Anexo 5.** Cubo “Catálogo de productos” que cumple con el segundo objetivo estratégico del Grupo de Mercadotecnia.



Anexos

- **Anexo 6.** Cubo “Identificar 5 nuevos productos “que cumple con el primer objetivo estratégico de los departamentos de producción.



- **Anexo 7.** Cubo “Soluciones aceptadas” que cumple con el segundo objetivo estratégico de los departamentos de producción.

