

Universidad de las Ciencias Informáticas
Facultad 3



Título

*Construcción de un Data Mart para el Sistema
Integral de Gestión de Medicamentos.*

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

Clasificación: Implementación.

*Autor: Yuniesky Nueva Aguilar
Tutor: Ing. Dayron Jesús Barrueco Barreto
Ing. Yusniel Hidalgo Delgado*

*Ciudad de la Habana, Junio de 2011
"Año 53 de la Revolución"*



“El mérito verdadero es el que el hombre adquiere con su voluntad, con su esfuerzo, con su constancia”.

Fidel Castro Ruz

Síntesis del Tutor:

Dayron Jesús Barrueco Barreto graduado de Ing. en Ciencias Informáticas por la Universidad de las Ciencias Informáticas en el 2008. Actualmente se desempeña como arquitecto de Base de Datos del proyecto ERP en la UCID. No posee categoría docente.

Años de graduado: 2 años de graduado.

E-mail: djbarrueco@uci.cu

Síntesis del Co-Tutor:

Yusniel Hidalgo Delgado graduado de Ing. en Ciencias Informáticas por la Universidad de las Ciencias Informáticas en el 2010. Actualmente se desempeña como desarrollador del proyecto ERP en el Centro de Gestión de Entidades CEIGE y Jefe del grupo de Inteligencia Artificial de la facultad 3. No posee categoría docente.

E-mail: yhdelgado@uci.cu

- *En especial a mis padres Juana y Amado por estar siempre apoyándome y dándome aliento en todos los momentos difíciles de mi carrera.*
- *A mis queridos abuelos que aunque no todos pueden compartir conmigo este momento me educaron y me apoyaron siempre.*
- *A mis amigos y compañeros de clases los cuales siempre me apoyaron y ayudaron en los momentos difíciles.*
- *A mis tutores excelentes profesionales, Dayron Jesús Barrueco y Yusniel Hidalgo, por su incondicional apoyo y ayuda en todo momento.*
- *A todos los profesores que ayudaron en mi formación profesional y lograron con su amor lo que soy hoy en día.*

A todos ustedes muchas gracias.



A mis padres, porque creyeron en mí y no dejaron un día de alimentar mis ansias de superación profesional, en agradecimiento a su apoyo incondicional y su esfuerzo es que hoy puedo ver mi sueño y el de ellos también, hecho realidad. Fue siempre su dedicación hacia mí y su contante preocupación lo que me ayudo a superar los tiempos difíciles de mi carrera.

A la memoria de mis Abuelos fallecidos los cuales siempre se preocuparon por mí, demostrándome siempre su cariño y ternura.

A toda mi familia en general, a mi querida hermana Yannelis Nueva Aguilar y a mi adorada abuela paterna Dolores Mora Mojena y a mi prima Liudmila a la que le agradezco mucho sus consejos.

A una de las más grandes alegrías de mi vida, mi querida novia Janna Nimo la cual siempre estuvo a mi lado brindándome su amor cariño y respeto. A la cual estaré agradecido por el resto de la vida por estar a mi lado apoyándome y juntos ver logrado materializar el sueño de ambos. Gracias mi amor por aparecer en mi vida.



Declaración de Autoría

Declaro ser el único autor de este trabajo y autorizo a la Universidad de las Ciencias Informáticas para que haga el uso que estime pertinente del mismo. Para que así conste firmo la presente a los ____ días del mes de _____ del año _____.

Yuniesky Nueva Aguilar

Ing. Dayron Jesús Barrueco Barreto

Firma del Autor

Firma del Tutor

Ing. Yusniel Hidalgo Delgado

Firma del Tutor

El grupo empresarial de la industria químico farmacéutica de Cuba (QUIMEFA) cuenta con una infraestructura capaz de suministrar los medicamentos que necesita el país. Actualmente es una empresa líder en su campo que incluye una amplia gama de productos químicos de uso industrial, productos avalados y destacados por su calidad. Aparejado a la producción y comercialización, la empresa tiene la necesidad de manejar la información que se encuentra en grandes volúmenes de datos para tomar decisiones de importancia vital en la organización.

La empresa QUIMEFA como parte del proceso de desarrollo del Sistema Integral de Gestión CEDRUX, se traza como objetivo la construcción de un data mart que contenga información consolidada y apoye una correcta toma de decisiones.

Actualmente la información se encuentra almacenada en una base de datos relacional sobre el sistema gestor de bases de datos PostgreSQL lo que impide hacer un análisis exhaustivo de la información consolidada. La metodología utilizada en la construcción del data mart fue un híbrido entre la metodología Hefestos y Kimball, desarrollada en el UCID. Se utilizaron para el desarrollo de la aplicación herramientas de software libre.

Se obtuvo un producto de software que consiste en un data mart, el mismo brinda reportes y acceso eficiente a la información que necesitan los directivos de QUIMEFA. Los reportes generados se pueden analizar mediante gráficas y tablas con valores numéricos, garantizando una mejor comprensión y análisis a la hora de tomar decisiones.

Introducción	1
Capítulo 1 Fundamentación Teórica.....	6
1.1 Introducción.....	6
1.2 Almacén de datos	6
1.2.1 Data mart.....	6
1.2.2 Diferencias entre almacén de datos y una Base de Datos Operacional	8
1.3 Metodologías para el Diseño e implementación de un data mart.....	9
1.3.1 Metodología Kimball	10
1.3.2 Metodología Hefesto.....	10
1.3.3 Metodología Inmon	12
1.4 Procesamiento Analítico Transaccional en línea (OLTP).....	14
1.5 Modelado Multidimensional.....	14
1.5.1 Esquema Estrella.....	15
1.5.2 Esquema Copo de Nieve	16
1.5.3 Esquema de Constelación	18
1.6 Procesamiento Analítico en Línea (OLAP).....	18
1.6.1 ROLAP.....	19
1.6.2 MOLAP	20
1.6.3 HOLAP.....	21
1.7 Herramientas para la construcción del data mart.....	21
1.7.1 Sistemas Gestores de Bases de Datos	21
1.7.1.1 ORACLE.....	21
1.7.1.2. MYSQL.....	22
1.7.1.3. POSTGRESQL.....	22
1.7.2. Herramientas de Integración de Datos	24
1.7.2.1 Clovert ETL.....	24
1.7.2.2 Enhydrooctopus.....	24
1.7.2.3 Kettle	25
1.7.3 Herramientas para OLAP.....	26
1.7.3.1 Olap4J.....	26
1.7.3.2 Mondrian.....	26
1.7.4 Schema Workbench.....	26

1.8 Herramientas CASE	27
1.8.1 Rational Rose	27
1.8.2 Visual Paradigm.....	27
1.8.3 Enterprise Architect.....	27
1.9 Servidores de Aplicaciones	28
1.9.1 Jboss	28
1.9.2 Apache Tomcat.....	28
1.10 Propuesta de la plataforma a utilizar	28
1.10.1 Módulos de la plataforma Pentaho BI.....	29
1.11 Data Marts existentes.....	30
1.11.1 Ámbito Internacional	30
1.11.2 Ámbito Nacional.....	30
1.11.3 Universidad de las Ciencias Informáticas (UCI).....	31
1.12 Conclusiones Parciales	31
Capítulo 2 Análisis y Diseño	32
2.1 Introducción.....	32
2.2 Propuesta del Sistema	32
2.3 Descripción de la empresa.....	33
2.4 Descripción de las fases de la Metodología Híbrida	34
2.4.1 Fase 1: Análisis de los requerimientos	34
2.4.1.1 Planificación de la entrevista	34
2.4.1.2 Identificar las preguntas	35
2.4.1.3 Indicadores y perspectivas de análisis	36
2.4.1.4 Modelo conceptual	38
2.4.1.5 Aprobación del Modelo Conceptual por el cliente.....	39
2.4.2 Fase 2: Análisis de los OLTP	39
2.4.2.1 Definir estado general de los sistemas fuentes	39
2.4.2.2 Determinación de indicadores	39
2.4.2.3 Establecer correspondencias	40
2.4.2.4 Nivel de granularidad.....	41
2.4.2.5 Modelo conceptual ampliado	42
2.4.2.6 Identificar dimensiones.....	43

2.4.2.7 Identificar hechos	44
2.4.3 Fase 3: Modelo lógico del data mart	44
2.4.4 Fase 4: Proceso de Extracción Transformación y Carga	45
2.4.4.1 Mapeo de datos Fuente-A-Destino	45
2.5 Conclusiones Parciales	48
Capítulo 3 Proceso Analítico en Línea	50
3.1 Introducción	50
3.2 Jerarquía de las tablas de dimensiones	50
3.3 Medidas	51
3.4 Consultas	52
3.5 Resultados para los clientes de Mondrian	52
3.6 Validación	54
3.6.1 Objetivos de la validación	54
3.6.2 Validación de pruebas mediante pruebas de conceptos	54
3.6.3 Pruebas de Integración	56
3.6.4 Prueba de validación del cliente	57
3.6.5 Pruebas del sistema	58
3.7 Despliegue del data mart	60
3.8 Conclusiones Parciales	61
Conclusiones	62
Principales Aportes	63
Recomendaciones	64
Referencia Bibliografía	65
Anexos	67

Figura 1 Estructura de la Metodología Hefesto	11
Figura 2 Estructura de la Metodología Hibrida a utilizar	14
Figura 3 Estructura de un Cubo	15
Figura 4 Esquema Estrella	16
Figura 5 Esquema Copo de Nieve	17
Figura 6 Esquema de Constelación	18
Figura 7 Sistemas Gestores de Bases de Datos	21
Figura 8 Herramientas de Integración de Datos	24
Figura 9 Herramientas CASE	27
Figura 10 Servidores de Aplicaciones	28
Figura 11 Plataforma Pentaho.....	29
Figura 12 Modelo de análisis de procesos del data mart.....	33
Figura 13 Ejemplo de Indicadores y Perspectivas	37
Figura 14 Modelo Conceptual del data mart	38
Figura 15 Correspondencias con la base de datos origen.....	40
Figura 16 Modelo Conceptual Ampliado	42
Figura 17 Identificación de dimensiones	43
Figura 18 Relación hecho-indicador.....	44
Figura 19 Estructura del data mart para QUIMEFA	45
Figura 20 Base de datos origen (QUIMEFA).....	47
Figura 21 Selección de los datos de lotes vencidos	47
Figura 22 Carga de datos	48
Figura 23 Jerarquía de los atributos de la dimensión tiempo.....	50
Figura 24 Cubos multidimensionales con sus medidas y dimensiones asociadas.....	51
Figura 25 Consultas MDX.....	52
Figura 26 Cubo Vencimiento de Lotes con sus dimensiones y medidas según el filtrado	53
Figura 27 Informaciones gráficas	53
Figura 28 Otra vista de información gráfica.....	54
Figura 29 Reportes	57
Figura 30 Pruebas de tiempos de ejecución # 1	59

Figura 31 Pruebas de tiempos de ejecución # 2	59
Figura 32 Pruebas de tiempos de ejecución # 3	60
Figura 33 Diagrama de despliegue del data mart	61

Tabla 1 Comparación entre BD operacional y Almacenes de Datos	9
Tabla 2 Mapeo de datos	46
Tabla 3 Características de servidores	58

Introducción

Actualmente la sociedad se encuentra sumergida en la era de la informatización. Las nuevas necesidades de información por parte de directores y ejecutivos de las distintas organizaciones, exigen tomar decisiones basadas en la calidad, requiriendo de ellas características como **confiabilidad** (que los datos representen la realidad), **oportunidad** (que la información se encuentre disponible cuando se necesite), **legibilidad** (que se encuentre en un formato que facilite su entendimiento), y **globalidad** (que incluya todo lo que se necesite). [1]

En el mundo de las grandes empresas los data mart son centro de atención, ya que constituyen uno de los soportes fundamentales para el proceso de toma de decisiones gerenciales; de ahí la importancia de que la información guardada en ellos sea confiable y con calidad.

En estos tiempos la información ha pasado de ser el proceso de automatización de tareas rutinarias a convertirse en un proceso de verdadera informatización de empresas y entidades, por estas razones se convierte la información en un aliado estratégico de gran importancia para la supervivencia de las empresas. Este ha sido el principal motivo que ha impulsado la difusión de los almacenes de datos al contar las empresas con una gran cantidad de datos mal gestionados.

Para mantenerse competitiva una organización necesita una buena gestión de datos que minimice las duplicidades en su tratamiento y asegure la calidad de los mismos, de manera que puedan servir como fuente para la toma de decisiones estratégicas y tácticas.

Esto conlleva a afirmar, que los aspectos relativos a la explotación, el control y la auditoría de los data marts cobran gran importancia tanto para los especialistas informáticos como para los directivos de las empresas.

Atendiendo a lo planteado anteriormente las empresas necesitan sistematizar la información para la toma de decisiones, principalmente aquellas en las que se identifican grandes volúmenes de datos sobre clientes, productos y transacciones.

En el caso de las empresas cubanas la introducción de tecnologías de la información y las comunicaciones en el manejo del conocimiento y la toma de decisiones tiene como objetivo fundamental el uso más racional de los recursos, el logro de una mayor productividad y la obtención de los productos con una mayor calidad. [2]

Como parte del proceso de informatización de las entidades empresariales pertenecientes a la Administración Central del Estado Cubano, se desarrolla en la Universidad de las Ciencias Informáticas el Sistema de Gestión Empresarial CEDRUX. Dicho software posee un subsistema de control logístico el cual controla procesos de negocio tales como: inventario, facturación, activos fijos tangibles, entre otros. El subsistema desarrollado fue personalizado para el control de medicamentos del Grupo Empresarial Farmacéutico (QUIMEFA), el cual se denomina: Sistema de Gestión Integral de Medicamentos, en lo adelante SGIM. Sin embargo, el SGIM no posee una herramienta que apoye la toma de decisiones por la gerencia de la empresa a partir de la obtención de una visión global del comportamiento de los principales procesos de negocio a lo largo del tiempo. La empresa maneja muchos datos y tiene como consecuencia directa la necesidad por parte de la dirección empresarial de aprovechar estos valiosos activos. Además, la generación de reportes del sistema es muy compleja, no tienen la debida precisión ya que no incluye toda la información tanto cualitativa como cuantitativa, lo cual trae consigo que las estadísticas que se hacen referentes a los distintos procesos se elaboran basadas en estimaciones y por tanto no cumplen las expectativas. Los directivos de la empresa QUIMEFA necesitan obtener información rápida y precisa de los distintos procesos que lleva a cabo su entidad.

Con el análisis de la situación problemática expuesta se desarrollará este trabajo investigativo para dar solución al siguiente **Problema a resolver**: ¿Cómo contribuir a la toma de decisiones a partir del análisis del comportamiento de los procesos de negocio gestionados en el Sistema de Gestión Integral de Medicamentos?

El **Objeto de Estudio** de la investigación lo constituyen los Sistemas de Apoyo a la toma de Decisiones y el **Campo de Acción** se define como: Sistemas de apoyo a la toma de decisiones en la empresa QUIMEFA.

Para darle solución al problema descrito, se ha planteado el siguiente **Objetivo General**:

Obtener una herramienta de software que permita el análisis del comportamiento de los principales procesos de negocio gestionados por el Sistema de Gestión Integral de Medicamentos.

De la cual se desglosan los siguientes **Objetivos específicos**:

1. Obtener el estado del arte sobre los principales sistemas de soporte a la toma de decisiones y sistemas de software que realizan procesos similares.

2. Obtener el análisis de los principales requerimientos de información en QUIMEFA.
3. Obtener el diseño y la implementación de la solución informática.
4. Validar la solución propuesta.

Para dar cumplimiento a los objetivos planteados se han definido las siguientes **tareas de investigación**:

1. Revisión de la bibliografía existente sobre los principales temas dentro del objeto de estudio.
2. Caracterización de los modelos multidimensionales existentes en la implementación de un Almacén de datos.
3. Identificación de los principales requerimientos para la construcción del data mart.
4. Definición de la base tecnológica a utilizar en la implementación del data mart que satisfaga los principales requerimientos identificados.
5. Implementación del data mart.
6. Realización de pruebas al data mart para verificar su correcto funcionamiento.

Idea a Defender: Si se implementa un data mart para el Sistema de Gestión Integral de Medicamentos, entonces se contribuirá a la mejora de la toma de decisiones a partir del análisis del comportamiento de los procesos de negocio gestionados por dicho sistema.

Diseño Metodológico:

Métodos de investigación científica utilizada

El presente trabajo emplea como **Método Teórico**, el **Método Inductivo – Deductivo**, el **Histórico-Lógico** y el **Método Modelación** que plantea:

Método Inductivo – Deductivo

Inducción: se define como una forma de razonamiento por medio de la cual se pasa del conocimiento de casos particulares a un conocimiento más general que refleja lo que hay en común entre los fenómenos individuales.

Deducción: es una forma de razonamiento, mediante el cual se pasa de un conocimiento general a otro de menor nivel de generalidad.

Con este método, se parte de leyes y principios generales para explicar y solucionar problemas particulares.

- Este método fue utilizado para lograr un entendimiento detallado de cómo ocurre el proceso de negocio en la muestra seleccionada y así proponer una solución a la problemática planteada. Se utilizará también para lograr un entendimiento acerca de los conocimientos generales sobre los almacenes de datos.

Histórico - Lógico:

Histórico: analizan la trayectoria completa del fenómeno, su condicionamiento a los diferentes períodos de la historia, revela las etapas principales de su desenvolvimiento y las conexiones históricas fundamentales.

Lógico: investiga las leyes generales del funcionamiento y desarrollo de los fenómenos.

- Se evidenciará la utilización de este método en el estudio de las bases de datos desde que surgieron los modelos de datos relacionales hasta los multidimensionales.

Método Modelación:

La modelación es una reproducción simplificada de la realidad, es justamente el proceso mediante el cual creamos modelos con vistas a investigar la realidad.

- Su utilización permitirá crear abstracciones que explican la realidad, por ejemplo, todos los modelos y diagramas que se generan en la metodología a utilizar para desarrollar la propuesta de solución.

Además se aplicó como **Método Empírico** la **Entrevista** y **Observación** que plantea:

Método de Entrevista

La entrevista es una conversación planificada entre el investigador y el entrevistado para obtener información. Su uso constituye un medio para el conocimiento cualitativo de los fenómenos o sobre características personales del entrevistado y puede influir en determinados aspectos de la conducta humana por lo que es importante una buena comunicación.

- Este método se evidenció en las entrevistas realizadas con las partes involucradas en el fenómeno que se está estudiando, esperando obtener información valiosa para el entendimiento global de todo el proceso.

Método Observación

Es la percepción planificada dirigida a un fin y relativamente prolongada de un hecho o fenómeno. Es el instrumento universal del científico, se realiza de forma consciente y está orientada a un objetivo determinado.

- Este método es usado para estudiar la metodología que se va a utilizar.

Resultados de la Investigación

- Obtención del código fuente de la solución informática.
- Obtención del modelo de datos de la solución informática.
- Obtención de los artefactos del diseño de la solución propuesta.

Capítulo 1 Fundamentación Teórica

1.1 Introducción

En este capítulo se aborda una breve descripción de las tecnologías y herramientas seleccionadas para el desarrollo de la solución propuesta. Se expondrán conceptos asociados a los almacenes de datos y su repercusión a nivel mundial, nacional y en la UCI. Además se hace énfasis en las ventajas y desventajas de los data marts, así como los modelos multidimensionales utilizados actualmente para ofrecer un mejor desempeño del sistema.

1.2 Almacén de datos

Muchos autores y especialistas del tema han emitido sus criterios acerca de una definición exacta de Almacén de datos y aunque el término ha sido analizado en diferentes puntos de vistas todos concluyen en esencia que el Almacén de datos posibilita la extracción de datos de sistemas operacionales y fuentes externas. Permite la integración y homogenización de los datos de toda la empresa, provee información que ha sido transformada y resumida, para ayudar en el proceso de toma de decisiones estratégicas y tácticas. A continuación se mostrarán los conceptos emitidos por los siguientes autores:

Según W. H. Inmon (considerado por muchos el padre del *Data Warehouse*), quien define: *“Un Almacén de datos es una colección de datos orientada al negocio, integrada, variante en el tiempo y no volátil para el soporte del proceso de toma de decisiones de la gerencia”*.

Según **Ralph Kimball** (considerado el principal promotor del enfoque dimensional para el diseño de almacenes de datos), *“Un Almacén de datos es una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis”*. [3]

1.2.1 Data mart

Un data mart es una versión especial de almacén de datos con el objetivo de responder a un determinado análisis, función o necesidad y con una población de usuarios específica, ayudando potencialmente a las empresas y entidades a la toma de decisiones. La principal diferencia entre un almacén de datos y un data mart es su alcance. El data mart está pensado para cubrir las necesidades

de un grupo de trabajo o de un determinado departamento dentro de la organización. Es el almacén natural para los datos departamentales. En cambio, el ámbito del almacén de datos es la organización en su conjunto. En síntesis los data mart son pequeños almacenes de datos centrados en un tema o un área de negocio específico dentro de una organización. Un data mart puede ser alimentado desde los datos de un almacén de datos, o integrar por sí mismo un compendio de distintas fuentes de información. Por tanto, para crear el data mart de un área funcional de la empresa es preciso encontrar la estructura óptima para el análisis de su información, estructura que puede estar desplegada sobre una base de datos *On Line Transaction Processing* (OLTP), como el propio Almacén de datos, o sobre una base de datos *On-Line Analytical Processing* (OLAP), estas estructuras son analizadas más adelante. 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 data marts. [4]

➤ Data mart OLAP

Se basan en los populares cubos OLAP los cuales junto con las dimensiones y las jerarquías son la esencia de la navegación multidimensional del OLAP. Los data marts OLAP 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 [4].

➤ Data mart OLTP

Pueden basarse en un simple extracto del almacén de datos, 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 [4].

Los data marts que están dotados con estas estructuras óptimas de análisis presentan las siguientes ventajas:

- Poco volumen de datos.
- Mayor rapidez de consulta.
- Consultas SQL y/o MDX sencillas.
- Validación directa de la información.
- Facilidad para el trabajo con el historial de datos.

Una vez mencionadas las características de los diferentes tipos de data mart, se seleccionó desarrollar para la empresa QUIMEFA un data mart OLAP, por presentar una estructura óptima que favorece a la entidad.

Desventaja de los data marts

No permite el manejo de grandes volúmenes de información comparado con un almacén de datos, por lo que muchas veces se debe recurrir a un conjunto de data marts para cubrir todas las necesidades de información de la empresa.

1.2.2 Diferencias entre almacén de datos y una Base de Datos Operacional

Los datos de un almacén de datos y los datos operacionales son accedidos por usuarios que los utilizan de formas diferentes. Los usuarios que acceden a los datos operacionales, comúnmente efectúan tareas predefinidas y actualizables que requieren acceso a una sola base de datos desde una aplicación y almacenan datos actuales. Por el contrario, los usuarios que acceden al almacén de datos, efectúan tareas complejas que requieren acceso a un conjunto de datos desde fuentes múltiples, frecuentemente no son predecibles y almacenan datos históricos. [5]

Características del almacén de datos:

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

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

La siguiente tabla muestra otros aspectos de interés que diferencian los almacenes de datos a las bases de datos operacionales. [3]

Tabla 1 Comparación entre BD operacional y Almacenes de Datos

Aspectos	BD operacional	Almacenes de Datos
Objetivo	De tipo operativo (operaciones del día a día)	Análisis y toma de decisiones
Proceso	De transacciones. Repetitivo y conocido.	De consultas masivas. Puntual y no conocido.
Actividad	Predomina la actualización.	Predomina la consulta
Rendimiento	Importancia del tiempo de respuesta de la transacción instantánea	Importancia de la respuesta masiva.
Organización	Organización -Estructura normalmente relacional	Visión multidimensional
Horizonte histórico	30 a 90 días	5 a 10 años
Volumen de datos	Pequeño/medio. Del orden del Mb a Gb.	Medio/grande. Del orden del Gb a Tb.

1.3 Metodologías para el Diseño e implementación de un data mart

Para los profesionales de Inteligencia de Negocios (BI), ya sean con experiencia o interesados en la materia, el tema de la metodología de BI es considerada parte medular de la formación profesional. Muchos profesionales se guían por metodologías basadas en las corrientes de Kimball y/o Inmon, así como muchos otros diseñan sus propias metodologías basadas en alguna de las corrientes mencionadas o que sean afines a la tecnología seleccionada. [6]

Es necesario conocer en qué consiste la Inteligencia de negocios ya que este término no incluye solamente las metodologías. Se define como inteligencia de negocios: al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización. [7]

El uso de la metodología orienta a definir correctamente los objetivos del negocio y precisar las metas del proyecto, pero no garantiza un proyecto exitoso debido a los factores directos e indirectos que influyen en este tema, como lo son las herramientas seleccionadas, la cultura organizacional, perfiles técnicos y de negocio de los consultores, presupuesto del proyecto, entre otros.

1.3.1 Metodología Kimball

Apunta a una solución completa que se puede implementar en poco tiempo. Esto cumple con lo que se necesita en una metodología de inteligencia de negocios, dar resultados rápidos para demostrar el valor de la solución al negocio y no perder la confianza de los usuarios. Se enfoca principalmente en el diseño de la base de datos que almacenará la información para la toma de decisiones.

El ciclo de vida de Kimball:

- Ilustra el flujo general de implementación de un almacén de datos (DW).
- Identifica secuencia de tareas ordenadas y actividades principales que deben suceder concurrentemente.
- Muchas necesidades deben ser acomodadas para lograr una única necesidad de la organización.
- No todos los detalles de las tareas del ciclo de vida deben ser ejecutados en todos los proyectos. [8]

1.3.2 Metodología Hefesto

Su idea está fundamentada en una amplia investigación, comparación de metodologías existentes y experiencias propias en procesos de confección de almacenes de datos. La metodología HEFESTO puede resumirse a través de la figura # 1 mostrada a continuación.



Figura 1 Estructura de la Metodología Hefesto

Para trabajar con la metodología Hefesto se comienza recolectando las necesidades de información de los usuarios y se obtienen las preguntas claves del negocio. Luego, se deben identificar los indicadores resultantes de los interrogativos y sus respectivas perspectivas de análisis, mediante las cuales se construirá el modelo conceptual de datos del almacén de datos (DW). Después, se analizarán los OLTP para señalar las correspondencias con los datos fuentes y seleccionar los campos de estudio de cada perspectiva. Una vez hecho esto, se pasará a la construcción del modelo lógico del depósito, precisándose las jerarquías que intervendrán. Por último, se definirán los procesos de carga, transformación, extracción y limpieza de los datos fuente. [9]

Esta metodología cuenta con las siguientes características:

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

- Es independiente de las estructuras físicas que contengan el almacén de datos ya sea esquemas de modelación de estrella, esquema de copo de nieve o constelación.

1.3.3 Metodología Inmon

Esta metodología la definió su autor Bill Inmon en el año 1992 en el libro “Building the Data Warehouse (DW)”. En él proponía los mecanismos necesarios para llevar a cabo la correcta realización de un almacén de datos. Para Bill Inmon, el diseño de un almacén de datos comienza ya con la mera introducción de datos en el mismo. El enfoque de Inmon de desarrollar una estrategia de almacén de datos e identificar las áreas principales desde el inicio del proyecto es necesario para asegurar una solución integral ya que esto ayuda a evitar la aparición de situaciones inesperadas que puedan poner en peligro el proyecto, debido a que se conoce con antelación y bastante exactitud la estructura que presentarán los principales núcleos del desarrollo, lo que permite enfocar los esfuerzos del desarrollo actual para ser compatible con los subsiguientes. De forma contraria a la de Kimball esta metodología puede tener una implementación mucho más tardada, y es recomendada cuando se hace demasiado difícil representar el modelo a través de dimensiones y la complejidad de la solución se hace demasiado grande, pero no es muy recomendable para proyectos sencillos ya que va de lo más general el DW a lo más específico data mart. [10]

La metodología que se utilizará como guía para la construcción del data mart es un híbrido conformado por combinación de la metodología de Hefestos y Kimball, adoptada por el centro de datos de la UCID, adaptándola a las necesidades propias de la institución. Se decidió realizar un híbrido entre estas dos metodologías y no adoptar una de las dos, porque se quiere aprovechar la facilidad que brinda Hefestos de una metodología sencilla y entendible con la de Kimball de generar artefactos para la documentación a la hora de realizar el data mart.

Es importante destacar que algunos de los artefactos que se adoptaron por el centro de datos fueron modificados, pues se determinó que no hacía falta documentar todo lo que este requería. Además, la metodología rectora fue la de Hefestos y no Kimball. Teniendo en cuenta esto, se puede decir que las características de la metodología híbrida son: [11]

- 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 de datos.
- 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 almacén de datos como para data mart.

Esta metodología, al igual que la de Hefestos, cuenta con las mismas cuatro fases de desarrollo, la diferencia radica en que se le han incorporado algunos pasos en cada una de estas fases y se han adoptado y creado artefactos para documentar lo que se realiza en cada fase, lo cual sirve para la reutilización del conocimiento en otros proyectos. [11]

A continuación en la figura # 2 se muestra las fases de la metodología híbrida.



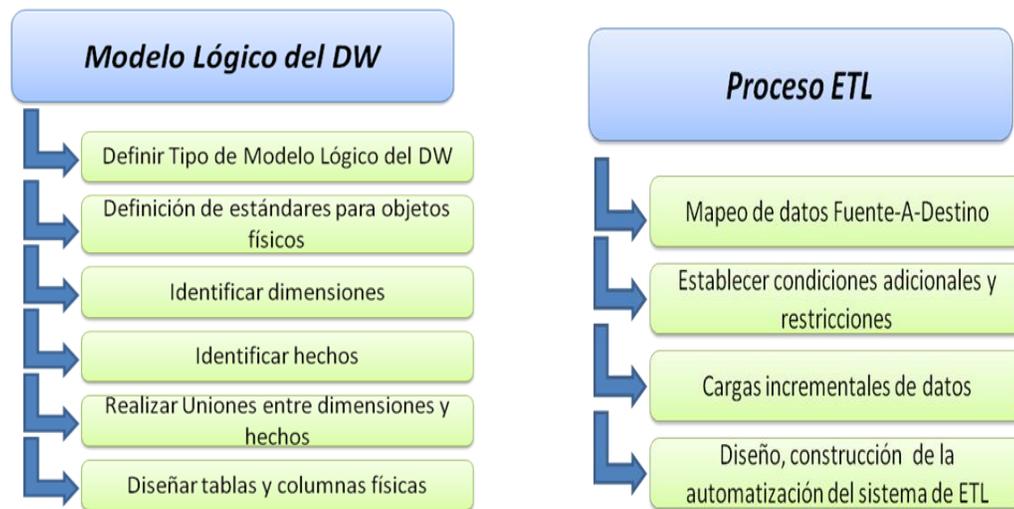


Figura 2 Estructura de la Metodología Híbrida a utilizar

1.4 Procesamiento Analítico Transaccional en línea (OLTP)

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

Entre los OLTP más habituales que pueden existir en cualquier organización se encuentran: [12]

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

1.5 Modelado Multidimensional

El análisis de los datos en un tiempo finito ha traído consigo estudios sobre la mejor forma de almacenar y representar estos datos para que puedan ser consultados de una forma más rápida. El uso del Modelo Multidimensional es una de las aproximaciones más acertadas y seguidas por los especialistas.

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

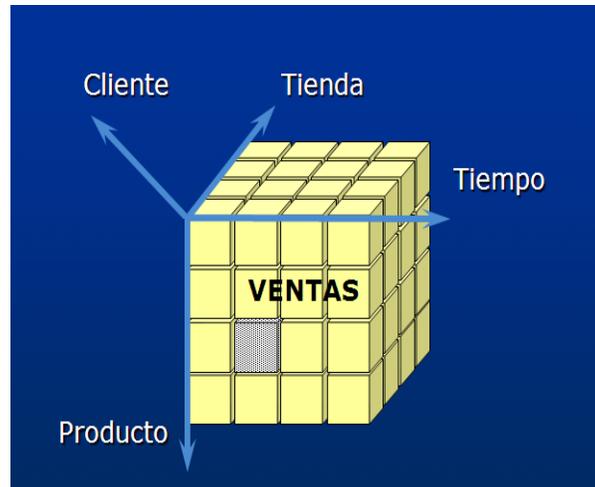


Figura 3 Estructura de un Cubo

A continuación se mencionarán los modelos multidimensionales más usados en la construcción de un data mart.

1.5.1 Esquema Estrella

Un esquema en estrella es un modelo de datos que tiene una tabla de hechos que contiene los datos para el análisis, rodeada de las tablas de dimensiones. Este aspecto, de tabla de hechos (o central) más grande rodeada de radios o tablas más pequeñas es lo que asemeja a una estrella, dándole nombre a este tipo de construcciones. Las tablas de dimensiones tendrán siempre una clave primaria simple, mientras que en la tabla de hechos, la clave principal estará compuesta por las claves principales de las tablas dimensionales. [13]

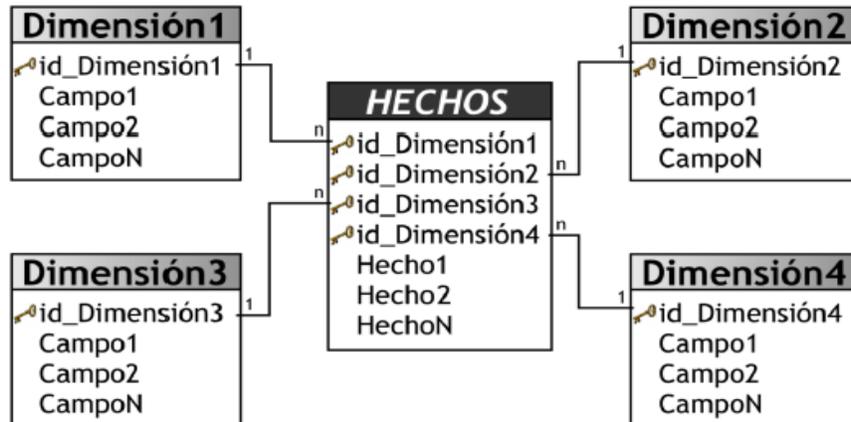


Figura 4 Esquema Estrella

A continuación se destacan algunas características de este modelo, que ayudarán a comprender mejor el por qué de sus ventajas:

- Posee los mejores tiempos de respuesta.
- Su diseño es fácilmente modificable.
- Existe paralelismo entre su diseño y la forma en que los usuarios visualizan y manipulan los datos.
- Simplifica el análisis.
- Facilita la interacción con herramientas de consulta y análisis.

1.5.2 Esquema Copo de Nieve

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

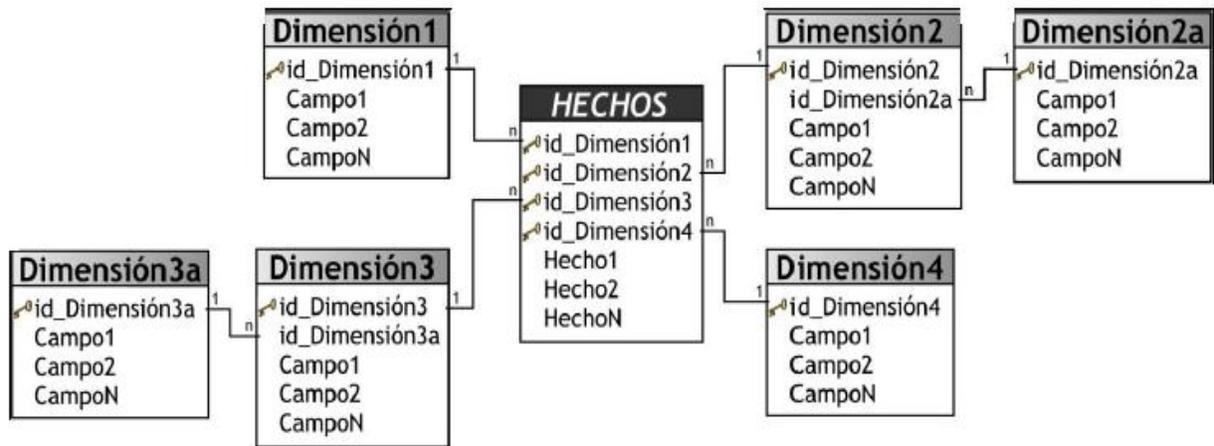


Figura 5 Esquema Copo de Nieve

Como se puede apreciar en la figura anterior, existe una tabla de hechos central que está relacionada con una o más tablas de dimensiones, quienes a su vez pueden estar relacionadas o no con una o más tablas de dimensiones. Este modelo es más cercano a un modelo de entidad relación, que al modelo en estrella, debido a que sus tablas de dimensiones están normalizadas. [13]

Una de los motivos principales de utilizar este tipo de modelo, es la posibilidad de segregar los datos de las tablas de dimensiones y proveer un esquema que sustente los requerimientos de diseño. Otra razón es que es muy flexible y puede implementarse después de que se haya desarrollado un esquema en estrella.

Se pueden definir las siguientes características de este tipo de modelo:

- Posee mayor complejidad en su estructura.
- Hace una mejor utilización del espacio.
- Es muy útil en tablas de dimensiones de muchas tuplas.
- Las tablas de dimensiones están normalizadas, por lo que requiere menos esfuerzo de diseño.
- Puede desarrollar clases de jerarquías fuera de las tablas de dimensiones, que permiten realizar análisis de lo general a lo detallado y viceversa.

A pesar de todas las características y ventajas que trae aparejada la implementación del esquema copo de nieve, existen dos grandes inconvenientes de ello:

- Si se poseen múltiples tablas de dimensiones, cada una de ellas con varias jerarquías, se creará un número de tablas bastante considerable, que pueden llegar al punto de ser inmanejables.

- Al existir muchas uniones y relaciones entre tablas, el desempeño puede verse reducido.

1.5.3 Esquema de Constelación

Su diseño y cualidades son muy similares a las del esquema en estrella, pero posee una serie de diferencias con el mismo, que son precisamente las que lo destacan y caracterizan. Entre ellas se pueden mencionar: [13]

- Permite tener más de una tabla de hechos, por lo cual se podrán analizar más aspectos claves del negocio con un mínimo esfuerzo adicional de diseño.
- Contribuye a la reutilización de las tablas de dimensiones, ya que una misma tabla de dimensión puede utilizarse para varias tablas de hechos.
- Posibilidad de navegar de un hecho hacia otro y la optimización del espacio gracias a la compartición de dimensiones, evitando así la redundancia de datos.

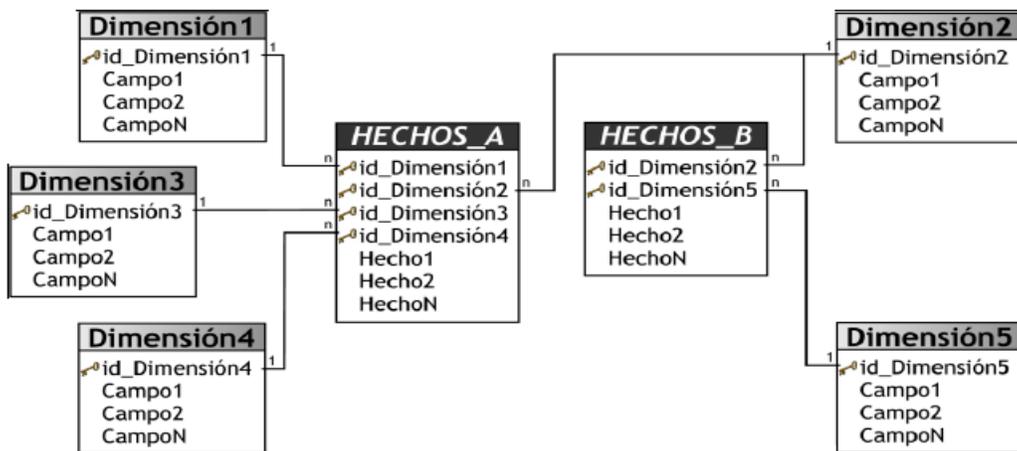


Figura 6 Esquema de Constelación

Teniendo como antecedente las características mencionadas de los distintos tipos de modelado multidimensional y adaptándolas a las necesidades específicas que presenta la base de datos a modelar. Se propone utilizar el modelado multidimensional constelación, ya que la base de datos perteneciente al data mart consta de varias tablas de hechos centrales y de este modo se garantizará la reutilización de las tablas de dimensiones.

1.6 Procesamiento Analítico en Línea (OLAP)

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

En este modelo los datos son vistos como cubos, los cuales consisten en categorías descriptivas (dimensiones) y valores cuantitativos (medidas). El modelo multidimensional de datos simplifica a los usuarios formular consultas complejas, arreglar datos de un reporte y cambiar de datos resumidos a detallados. [14]

Existen dos arquitecturas para los sistemas OLAP que son: OLAP multidimensional (MOLAP) y OLAP relacionales (ROLAP).

1.6.1 ROLAP

ROLAP (Relational On-line Analytical Process) arquitectura OLAP que crea vistas multidimensionales extrayendo los datos de bases de datos relacionales, simulan los datos multidimensionales usando sofisticadas técnicas de indexación.

Ventajas

- Puede manejar grandes cantidades de datos y no pone limitación en la cantidad existente de los mismos.
- Aprovecha las funciones inherentes de las base de datos relacionales, pues a menudo éstas traen consigo una serie de funcionalidades tecnológicas ROLAP.
- Puede escalar a un gran número de perspectivas de análisis de negocio (dimensiones).

Desventajas

- El rendimiento puede ser lento, debido a que cada informe ROLAP es esencialmente una consulta SQL (o varias consultas SQL) en la base de datos relacional, el tiempo de consulta puede ser largo si el tamaño de los datos subyacentes es grande.
- Es difícil realizar cálculos complejos con SQL aunque no imposible. [15]

1.6.2 MOLAP

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

Ventajas

- Factibilidad para realizar cálculos con un alto grado de complejidad.
- Posee un buen rendimiento.
- Los cubos MOLAP permiten una recuperación rápida de los datos.

Desventajas

- Limitado en la cantidad de datos que puede manejar: debido a que todos los cálculos se realizan cuando el cubo se construye, no es posible incluir una gran cantidad de datos en el propio cubo.
 - Requiere de inversiones adicionales ya que la tecnología de cubos son a menudo propietarias.
- [15]

Comparación entre ROLAP y MOLAP

Cuando se comparan las dos arquitecturas, se pueden realizar las siguientes observaciones:

- ROLAP delega la negociación entre tiempo de respuesta y el proceso *batch* al diseño del sistema. Mientras, el MOLAP, suele requerir que sus bases de datos se precompilen para conseguir un rendimiento aceptable en las consultas, incrementando por tanto los requerimientos *batch*.
- Los sistemas con alta volatilidad de los datos (aquellos en los que cambian las reglas de agregación y consolidación), requieren una arquitectura que pueda realizar esta consolidación *ad-hoc*. Los sistemas ROLAP soportan bien esta consolidación dinámica, mientras que los MOLAP están más orientados hacia consolidaciones *batch*.
- ROLAP puede crecer hasta un gran número de dimensiones, mientras que MOLAP generalmente es adecuado para diez o menos dimensiones.
- ROLAP soporta análisis OLAP para grandes volúmenes de datos elementales, mientras que MOLAP se comporta razonablemente en volúmenes más reducidos (menos de 5 GB).

En resumen, ROLAP es una arquitectura flexible y general, que crece para dar soporte a amplios requerimientos OLAP. Sin embargo, MOLAP es una solución particular, adecuada para soluciones departamentales con unos volúmenes de información y número de dimensiones más modestos. [15]

Una vez analizadas las ventajas y desventajas de cada una de las arquitecturas, se propone la utilización de la arquitectura ROLAP para la construcción del data mart de la empresa QUIMEFA, ya que los datos se extraen de una base de datos relacional soportada por un sistema gestor de base de datos PostgreSQL. Además esta arquitectura ofrecerá un sistema escalable para grandes volúmenes de datos que soportará en un futuro la empresa.

Es importante mencionar que existe una tercera arquitectura para OLAP la cual se analiza a continuación.

1.6.3 HOLAP

Arquitectura HOLAP (*Hybrid On-line Analytical Process*) permite un análisis híbrido de la información, es decir que une lo mejor de los dos tipos anteriores. El análisis HOLAP ayuda a reducir costos de hardware, ya que se necesitan menos disco que en las bases de datos relacionales. Además, la respuesta de las consultas sobre las bases de datos multidimensionales son más rápidas que sobre las relacionales. Como aspecto negativo, los datos multidimensionales deben ser cargados antes de ser consultados y refrescados cuando se actualizan los datos de la organización. [15]

1.7 Herramientas para la construcción del data mart

1.7.1 Sistemas Gestores de Bases de Datos



Figura 7 Sistemas Gestores de Bases de Datos

1.7.1.1 ORACLE

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

- Entorno cliente/servidor.
- Gestión de grandes bases de datos.
- Usuarios concurrentes.
- Alto rendimiento en transacciones.
- Sistemas de alta disponibilidad.
- Disponibilidad controlada de los datos de las aplicaciones.
- Adaptación a estándares de la industria, como SQL-92.
- Gestión de la seguridad.
- Autogestión de la integridad de los datos.
- Opción distribuida.
- Portabilidad.
- Compatibilidad.
- Conectabilidad.
- Replicación de entornos. [16]

1.7.1.2. MYSQL

MySQL es un sistema de gestión de bases de datos relacional, licenciado bajo la GPL (acrónimo en inglés de GNU) de la GNU (*General Public License*). Su diseño multihilo le permite soportar una gran carga de forma eficiente.

Las principales características de este gestor de bases de datos son las siguientes:

- Aprovecha la potencia de sistemas multiprocesador, gracias a su implementación multihilo.
- Soporta gran cantidad de tipos de datos para las columnas.
- Dispone de API's en gran cantidad de lenguajes (C, C++, Java, PHP, etc.).
- Gran portabilidad entre sistemas.
- Soporta hasta 32 índices por tabla.
- Gestión de usuarios y contraseñas, manteniendo un buen nivel de seguridad en los datos. [17]

1.7.1.3. POSTGRESQL

PostgreSQL es un sistema de gestión de bases de datos objeto-relacional ya que incluye características de la orientación a objetos, como puede ser la herencia, tipos de datos, funciones,

restricciones, disparadores, reglas e integridad transaccional. A pesar de esto, PostgreSQL no es un sistema de gestión de bases de datos puramente orientado a objetos.

A continuación se mencionan las principales características de este gestor de bases de datos:

- Implementación del estándar (lenguaje) SQL92/SQL99.
- Soporta distintos tipos de datos: además del soporte para los tipos base, también soporta datos de tipo fecha, monetarios, elementos gráficos, datos sobre redes (MAC, IP), cadenas de bits, etc. También permite la creación de tipos propios.
- Incorpora una estructura de datos arreglo.
- Permite la declaración de funciones propias, así como la definición de disparadores.
- Soporta el uso de índices, reglas y vistas.
- Incluye herencia entre tablas (aunque no entre objetos, ya que no existen), por lo que a este gestor de bases de datos se le incluye entre los gestores objeto-relacionales.
- Permite la gestión de diferentes usuarios, como también los permisos asignados a cada uno de ellos.
- Soporta casi toda la sintaxis SQL (incluyendo sub consultas, transacciones, tipos y funciones definidas por el usuario).
- Cuenta con un amplio conjunto de interfaces con lenguajes de programación como por ejemplo C, C++, Java, Python, PHP y otros.
- El progreso incremental del gestor de datos de código abierto PostgreSQL brinda a los consumidores la opción de instalar una base de datos no privativa. [17]

Después del análisis realizado sobre los distintos sistemas gestores de base de datos, se utilizará el PostgreSQL. Dicho gestor posee licencia GPL por lo que no tiene costo asociado a la licencia del software. Está disponible en múltiples plataformas, cuenta con varias herramientas gráficas de alta calidad para administrar las bases de datos (pgAdmin , pgAccess, phppgadmin) y para hacer diseño de bases de datos (Tora , Data Architect), contando también con un amplio conjunto de enlaces con lenguajes de programación como por ejemplo C, C++, Java, Python, PHP y otros. Mencionar que es el gestor de base de datos que está definido en el UCID como parte de la política de desarrollo, por lo que se tiene experiencia previa en su uso. Todas estas características lo convierten en la herramienta ideal para el desarrollo del data mart para QUIMEFA.

1.7.2. Herramientas de Integración de Datos



Figura 8 Herramientas de Integración de Datos

1.7.2.1 Clovert ETL

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

Características principales:

- Realiza transformaciones gráficas basadas en XML para la descripción de los metadatos de los registros.
- Se distribuye bajo la licencia LGPL.
- Clover ETL soporta 4 diferentes tipos de datos: string, numeric, date, bytes.
- Arquitectónicamente, está conceptualizado en unidades lógicas separadas llamadas unidades de transformación que engloban funcionalidades de la transformación e inteligencia, cada una de las cuales puede ser utilizada como componente de una aplicación independiente en otras aplicaciones y servicios. Cada componente se ejecuta como un hilo independiente, creando un ambiente más tolerante a fallos. [18]

1.7.2.2 Enhydrooctopus

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

Características principales:

- La característica principal de Octopus es el requerimiento de que para acceder a las fuentes de datos que lo utilicen debe existir un manejador JDBC disponible.
- Herramienta poderosa capaz de: normalizar datos, crear llaves artificiales, tablas y llaves primarias.
- Todos los trabajos que corren en Octopus son independientes de los proveedores de bases de datos. [18]

1.7.2.3 Kettle

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

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

Características principales:

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

Para el proceso de integración de datos del data mart para QUIMEFA se adopta la herramienta Kettle, pues la misma funciona en distintos sistemas operativos como Windows, Unix y Linux. Ofrece una licencia pública GPL y soporta Oracle, DB2, SQL Server y Sybase así como MySQL, PostgreSQL, Hypersonic, FireBird SQL e Ingres. Además existen experiencias previas de trabajo con la herramienta en la universidad. [18]

1.7.3 Herramientas para OLAP

1.7.3.1 Olap4J

OLAP4J (Online Analytical Processing for Java) es una interfaz de aplicación para el ambiente Java 2 Platform, Enterprise Edition, que soporta la creación, almacenamiento y administración de datos para una aplicación OLAP. [18]

1.7.3.2 Mondrian

Mondrian es un motor de ROLAP desarrollado en Java que permite analizar grandes conjuntos de datos almacenados en un almacén de datos. Posee un alto desempeño y análisis interactivo de volúmenes de información de cualquier tamaño, además explora dimensionalmente los datos y realiza cálculos avanzados utilizando el lenguaje MDX (multidimensional expressions o expresiones multidimensionales). [19] Para acceder a las funcionalidades que brinda Mondrian se requiere de un cliente, es el caso de JPivot, la cual es una librería de Java Server Pages (JSP) personalizados que presentan tablas OLAP y permite realizar operaciones tales como Drill Down (apreciar los datos en un mayor detalle, bajando por la jerarquía de una dimensión) y Drill Up (apreciar los datos en menor nivel de detalle, subiendo por la jerarquía de una dimensión [20] además de consultas OLAP por medio del lenguaje MDX. Está diseñado para trabajar con motores OLAP como el de Mondrian, por lo cual no utiliza las interfaces de programación de aplicaciones o (APIs) de Mondrian directamente sino que implementa su propio modelo OLAP.

Como parte del estudio realizado se escoge la herramienta Mondrian para realizar el proceso analítico, por presentar un alto desempeño en el análisis interactivo de grandes o pequeños volúmenes de información y exploración dimensional de los datos.

1.7.4 Schema Workbench

Es un entorno visual para el desarrollo y prueba de cubos OLAP. Esta herramienta se utiliza para la creación de los archivos XML que se usan para la construcción de los cubos. Además permite la ejecución de consultas MDX contra el esquema y la base de datos. Es una herramienta muy orientada al desarrollador conocedor de la estructura de un esquema de Mondrian. Tiene dos áreas: la zona en que se muestra la jerarquía del esquema OLAP y la zona de edición de las propiedades de cada elemento. Su desarrollo fue implementado en Java multiplataforma. [21]

1.8 Herramientas CASE

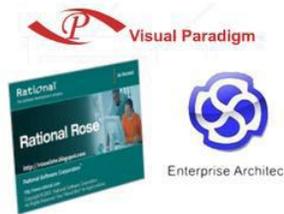


Figura 9 Herramientas CASE

1.8.1 Rational Rose

Rational Rose es la herramienta CASE que comercializan los desarrolladores de UML y que soporta de forma completa la especificación del UML. Propone la utilización de cuatro tipos de modelo para realizar un diseño del sistema, utilizando una vista estática y otra dinámica de los modelos del sistema, uno lógico y otro físico. Permite crear y refinar estas vistas creando de esta forma un modelo completo que representa el dominio del problema y el sistema de software. [22]

1.8.2 Visual Paradigm

Visual Paradigm para UML es una herramienta UML profesional que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. El software de modelado UML ayuda a una rápida construcción de aplicaciones de calidad mejores y a un menor costo. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. La herramienta también proporciona abundantes tutoriales, demostraciones interactivas y proyectos UML. [22]

1.8.3 Enterprise Architect

Enterprise Architect es una herramienta de construcción y modelado de software de alto rendimiento con una trazabilidad completa desde los requisitos iniciales hasta las decisiones de diseño de software, Enterprise Architect provee el tipo de visualización, colaboración eficiente y robusta requerida en los entornos de desarrollo de software que actualmente son altamente demandantes. Como una solución de modelado verdaderamente ágil, provee una sobrecarga de instalación baja, un rendimiento brillante y una interfaz intuitiva. Es una herramienta de análisis de negocio y UML orientada a objetos, provee el límite competitivo para el desarrollo de software, administración de proyecto, administración de requerimientos y análisis de negocio. Es una herramienta flexible para la plataforma de Windows y Linux. [23]

Entre las herramientas CASE estudiadas se escoge para el modelado del data mart el Visual Paradigm, pues provee beneficios significativos para ayudar a construir modelos de sistemas de software rigurosos donde es posible mantener la trazabilidad de manera consistente y soportar este proceso, en un ambiente fácil de usar, rápido y flexible.

1.9 Servidores de Aplicaciones



Figura 10 Servidores de Aplicaciones

1.9.1 Jboss

JBoss Application Server es el servidor de aplicaciones de código abierto más ampliamente desarrollado del mercado. Combinando una arquitectura orientada a servicios revolucionaria, con una licencia de código abierto. JBoss puede ser descargado, utilizado, incrustado y distribuido sin restricciones por la licencia.[24]

1.9.2 Apache Tomcat

Apache Tomcat es un servidor Web con soporte de servlets y JSPs. Incluye el compilador Jasper, que compila JSPs convirtiéndolas en servlets. El motor de servlets de Apache Tomcat a menudo se presenta en combinación con el servidor Web Apache. Tomcat puede funcionar como servidor Web por sí mismo, que puede ser usado como servidor Web autónomo en entornos con alto nivel de tráfico y alta disponibilidad. Dado que Tomcat fue escrito en Java, funciona en cualquier sistema operativo que disponga de la máquina virtual. [25]

Se selecciona como servidor de aplicaciones para el despliegue del data mart el Tomcat ya que es un servidor de aplicaciones de código abierto implementado en Java. Al estar basado en Java, puede ser utilizado en cualquier sistema operativo que lo soporte.

1.10 Propuesta de la plataforma a utilizar



Figura 11 Plataforma Pentaho

La plataforma Open Source Pentaho Business Intelligence cubre las amplias necesidades de análisis de los datos y de los informes empresariales. Las soluciones de Pentaho están escritas en Java y tienen un ambiente de implementación también basado en Java. Eso hace que Pentaho sea una solución muy flexible para cubrir una amplia gama de necesidades empresariales, tanto las típicas como las sofisticadas y específicas al negocio. [26]

1.10.1 Módulos de la plataforma Pentaho BI

- **Reporting** es el módulo de los informes y ofrece la solución adecuada a las necesidades de los usuarios. Pentaho Reporting permite generar informes de gran capacidad. Pentaho Reporting permite la distribución de los resultados del análisis en múltiples formatos, todos los informes incluyen la opción de imprimir o exportar a formato PDF, XLS, HTML y texto. Los reportes Pentaho permiten también programación de tareas y ejecución automática de informes con una determinada periodicidad.
- **Análisis** suministra a los usuarios un sistema avanzado de análisis de información. Con uso de las tablas dinámicas (*pivot tables, crosstabs*), generadas por Mondrian, el usuario puede navegar por los datos, ajustando la visión de los datos, los filtros de visualización, añadiendo o quitando los campos de agregación. Además, con el *Microsoft Excel Analysis Services*, se puede analizar los datos dinámicos en Microsoft Excel (usando la conexión a OLAP server Mondrian).
- **Integración de Datos** se realiza con una herramienta Kettle ETL (*Pentaho Data Integration*) que permite implementar los procesos ETL. Últimamente Pentaho lanzó una nueva versión, PDI 3.0 que marcó un gran paso adelante en OSBI ETL y que hizo que Pentaho Data Integration sea una alternativa interesante para las herramientas comerciales. [26]

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

1.11 Data Marts existentes

1.11.1 Ámbito Internacional

El actual mundo empresarial ha seguido muy de cerca el desarrollo de las nuevas tecnologías, dada la necesidad constante de consumir información y hacer un mejor uso de los datos ante un mercado más competitivo, es por eso que hoy en día la mayoría de las empresas líderes en sus distintas categorías cuentan con un sistema informático de apoyo a la toma de decisiones. Una de estas empresas es Petroecuador que cuenta con un data mart que les proporciona a los directivos información de los volúmenes y precios de embarques de crudo y productos para la exportación e importación, almacenándose históricamente para permitir el análisis de tendencias y proyecciones brindando una mejor toma de decisiones al momento de diseñar las estrategias de producción. Este data mart fue diseñado mediante el esquema de estrella cuenta de dos tablas de hechos y cinco dimensiones de las cual se vale para filtrar la información.

Existen otras empresas que se han dedicado a crear proyectos de inteligencia de negocios que incluyen los data marts y ejemplo de ello es data mart S.A, empresa creada en el año 2000, radica actualmente en Santiago de Chile y ha experimentado un crecimiento notable en el mercado latinoamericano.

1.11.2 Ámbito Nacional

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

Una empresa como CIMEX, cuya labor incluye redistribuir productos a entidades para su venta a la población, con gran volumen de información a consultar de los productos comercializados en la corporación, cuenta con un data mart encargado de calcular las propuestas de compras permitiendo

una gestión de compra–venta eficiente, con una finalidad fundamental: “Disminuir los costos, sin afectar al cliente, permitiendo prestaciones eficientes y con la calidad requerida, aumentando las ganancias o utilidades de las empresas”

1.11.3 Universidad de las Ciencias Informáticas (UCI)

En la universidad hasta la fecha se han realizado tres trabajos de tesis relacionados con el tema “data marts”. Una de ellas “Data mart para la Unidad Central de Cooperación Médica (UCCM)” el mismo brinda reportes sobre las colaboraciones médicas en el exterior. El data mar de la UCCM fue nutrido por los datos de una base de datos relacional sobre el sistema gestor de base de datos MySQL, consta actualmente de un cubo nombrado colaboración medica relacionado con seis dimensiones. La modelación de la base de datos está diseñada mediante el modelo multidimensional estrella. Actualmente constituye una fuente de apoyo a la toma de decisiones de dicha entidad.

1.12 Conclusiones Parciales

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

Capítulo 2 Análisis y Diseño

2.1 Introducción

En este capítulo se describen todos los pasos a seguir en la metodología utilizada, con el objetivo de lograr una correcta organización y planificación para construir el data mart. Se definen los requerimientos de software, las estructuras de datos y los artefactos necesarios para la implementación de la solución propuesta.

2.2 Propuesta del Sistema

Se propone el desarrollo del sistema basado en la metodología Híbrida entre (Efestos-Kimball) a la cual se dio seguimiento durante todo el proceso de construcción. Para un mayor entendimiento y visión del trabajo realizado se construyó mediante la herramienta Visual Paradigm un flujo de procesos que representa toda una serie de pasos o iteraciones que se deben seguir durante el desarrollo del data mart (vale aclarar que este modelo no pertenece a ninguno de los pasos de la metodología Híbrida).

En la figura # 12 se muestra la secuencia de procesos a seguir para la construcción del data mart. El flujo de trabajo comienza cuando el usuario, en este caso uno de los directivos de QUIMEFA solicita información para la toma de decisiones.

Lo primero a analizar fue el proceso de extracción, transformación y carga, el cual tiene como entrada fundamental los recursos que se consumieron, estos recursos son los sistemas operacionales involucrados (BD QUIMEFA). Luego se realiza el procesamiento analítico en línea, que tiene como entrada el propio data mart con sus respectivos cubos y dimensiones. La salida es la información solicitada, que se representa en forma de gráfica o tabla para su mejor comprensión o estudio a la hora de tomar decisiones.

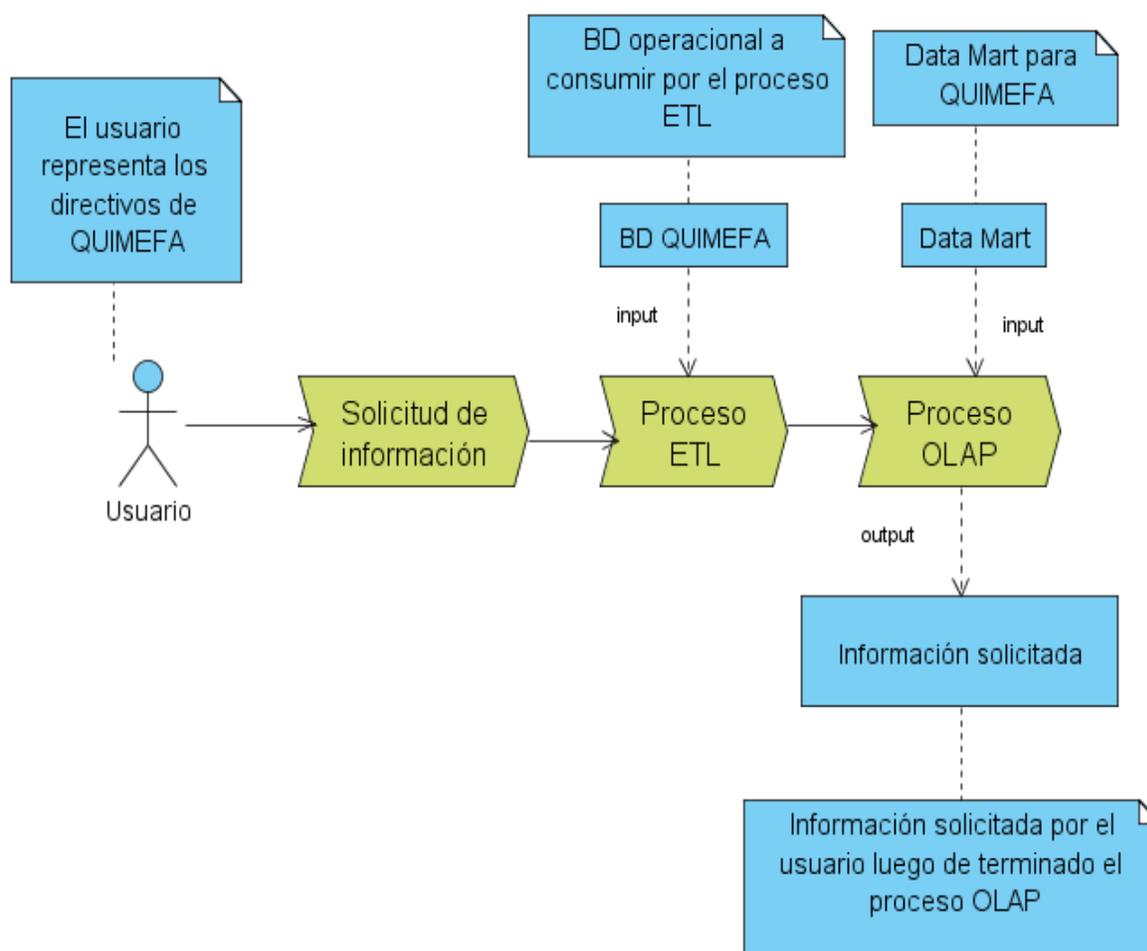


Figura 12 Modelo de análisis de procesos del data mart

2.3 Descripción de la empresa

La empresa analizada, desarrolla actividades tanto de producción como de comercialización de productos farmacéuticos, en un ambiente geográfico de alcance nacional e internacional. De acuerdo con su volumen de operaciones, se puede considerar de gran tamaño.

Presenta una estructura formalizada, que es capaz de suministrar los medicamentos que necesita el país. Su sostenido trabajo de investigación y desarrollo incorpora de forma continua, nuevos fármacos al arsenal terapéutico de la salud cubana.

Objetivo

Su objetivo principal es maximizar sus ganancias y su producción. Pero también, se puede adicionar el objetivo de expandirse a un nuevo nivel de mercado, con el fin de conseguir una mayor cantidad de clientes y reafirmarse competitivamente.

Procesos de interés por la empresa

Proceso de facturación, inventario y activos fijos tangibles.

Relación de las metas de la organización con las del data mart

El data mart coincide con la metas de la empresa, ya que esta necesita mejorar su eficiencia en la toma de decisiones y contar con información detallada. Esto es vital para procurar una mayor ventaja competitiva y conocer cuáles son los factores que inciden directamente sobre su rentabilidad.

El data mart aportar valor a la empresa. Entre las principales ventajas e inconvenientes que soluciona, se pueden mencionar los siguientes:

- Permite a los usuarios tener una visión general del negocio.
- Transforma datos operativos en información analítica enfocada a la toma de decisiones.
- Generar reportes dinámicos.
- Soporta la estrategia de negocio de la empresa.
- Aporta a la mejora continua de la estructura de la empresa.

2.4 Descripción de las fases de la Metodología Híbrida

2.4.1 Fase 1: Análisis de los requerimientos

2.4.1.1 Planificación de la entrevista

Se realizó una planificación de las entrevistas que se hicieron al cliente, en las que se detallaron datos sobre el encuentro como la hora, participantes, lugar de la entrevista y aseguramiento. El objetivo de este paso consistió en organizar cada una de las entrevistas y asegurar la calidad requerida, así como tener las ideas claras de lo allí expuesto.

En esta fase se realizaron cuatro entrevistas con el objetivo de obtener los requerimientos para el desarrollo del almacén de datos. A continuación se exponen algunos detalles de cada una de ellas.

Entrevista 1

Esta entrevista se realizó con el objetivo de conocer cómo funciona el negocio en los subsistemas de inventario, facturación y activos fijos tangibles en la empresa. La misma se realizó a los especialistas de dichos subsistemas. Al concluir, el equipo de desarrollo obtuvo la información del negocio que le sirvió de base para el posterior desarrollo del sistema, por lo que se cumplieron los objetivos propuestos.

Entrevista 2

En la entrevista estuvieron presentes, el equipo de desarrollo y los especialistas de los subsistemas a los cuales va dirigido el data mart. Su objetivo principal consistió, en identificar las perspectivas y los indicadores. En la misma se realizaron una serie de preguntas a los especialistas, en busca de las necesidades de información, los procesos sobre los cuales les interesaba realizar los análisis y que perspectivas e indicadores se utilizarían en los mismos. Los objetivos de la entrevista se cumplieron, pues se identificaron los indicadores y perspectivas a utilizar en la confección del modelo conceptual.

Entrevista 3

Esta entrevista se desarrolló con el objetivo de definir el nivel de granularidad de las perspectivas identificadas anteriormente. En la misma participaron, el equipo de desarrollo y los especialistas de los subsistemas de inventario, facturación y activos fijos tangibles. Una vez concluida la entrevista, quedó definido el nivel de granularidad para cada una de las perspectivas.

Entrevista 4

La entrevista fue realizada con el objetivo de precisar las formas de cálculo de los indicadores obtenidos en la segunda entrevista. Estuvieron presentes los especialistas de los subsistemas facturación, inventario y activos fijos tangibles, los cuales mostraron las fórmulas pertinentes para cada indicador y aceptaron algunas formas de cálculos propuestas por el grupo de desarrollo. Como resultado de la misma se le dio cumplimiento al objetivo propuesto ya que se definieron todas las fórmulas de cálculo a utilizar.

2.4.1.2 Identificar las preguntas

El análisis de los requerimientos de los diferentes usuarios, es el punto de partida de esta metodología, ya que ellos son los que deben en cierto modo, guiar la investigación hacia un desarrollo que refleje claramente lo que se espera del depósito de datos, en relación a sus funciones y

cualidades. El objetivo principal de este paso, fue obtener e identificar las necesidades de informaciones claves de alto nivel, que es esencial para llevar a cabo las metas y estrategias de la entidad.

Resultados de las entrevistas

- Se desea tener control sobre la cantidad de productos existentes en los laboratorios del país, por lo que se necesita saber de los productos: el precio, importe total, existencia y la fecha de emisión organizada por años.
- Se desea tener control sobre la cantidad de productos existentes en las droguerías del país, por lo que se necesita saber de los productos: el precio, importe total, existencia y la fecha de emisión organizada por años.
- Se desea tener el control de los movimientos generales de medicamentos en las droguerías del país, por lo que se necesita conocer el producto que está en movimiento, lote al que pertenece, cliente involucrado en dicho movimiento, documento generado por el proceso, la cantidad, el precio, el importe total del producto que se encuentra en movimiento y la fecha organizada por años.
- Se desea tener el control de los movimientos generales de medicamentos en los laboratorios del país, por lo que se necesita conocer el producto que está en movimiento, lote al que pertenece, cliente involucrado en dicho movimiento, documento generado por el proceso, la cantidad, el precio y el importe total del producto que se encuentra en movimiento entre las distintas entidades.
- Se desea tener control de los movimientos de entrada y salida de medicamentos, por lo que se necesita saber la cantidad del producto, precio, importe, entrada o salida, y fecha del movimiento organizada por años.
- Se desea tener control sobre los lotes de medicamentos vencidos por cada una de las entidades laboratorios/droguerías y se necesita saber de los mismos el tipo de producto, la cantidad vencida de producto por lotes y la fecha organizada por años.
- El sistema debe ser capaz de integrar los reportes, visualizándolos en formatos de hojas electrónicas de cálculo en Excel y visores en formato PDF.

2.4.1.3 Indicadores y perspectivas de análisis

Indicadores:

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

Perspectivas:

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

A continuación en la figura # 13 se muestra la identificación de estas características mediante un ejemplo del resultado de la propia entrevista.

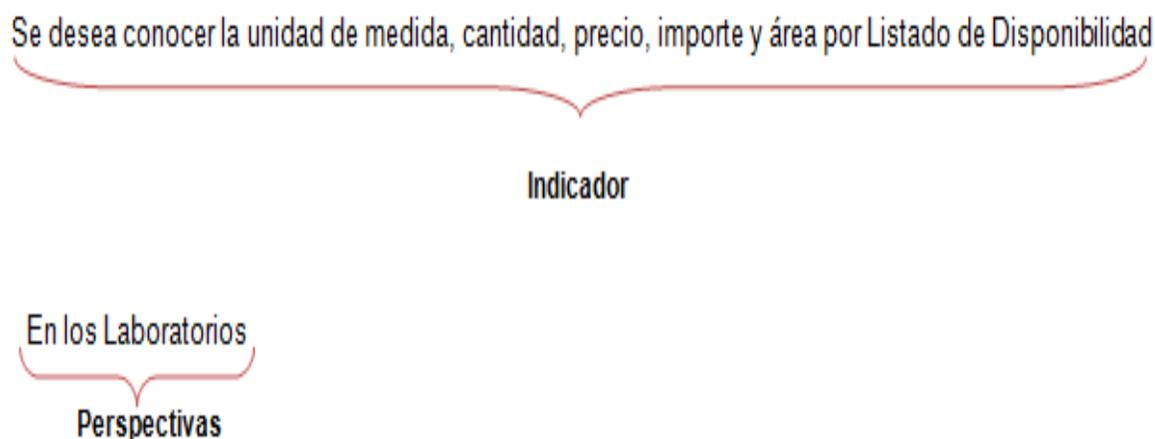


Figura 13 Ejemplo de Indicadores y Perspectivas

2.4.1.4 Modelo conceptual

Una vez identificados los indicadores y perspectivas se confeccionó el modelo conceptual de los datos del data mart, en el que se reflejaron las perspectivas a la izquierda y los indicadores a la derecha, como se muestra en la figura # 14, proporcionando una idea del alcance del data mart.

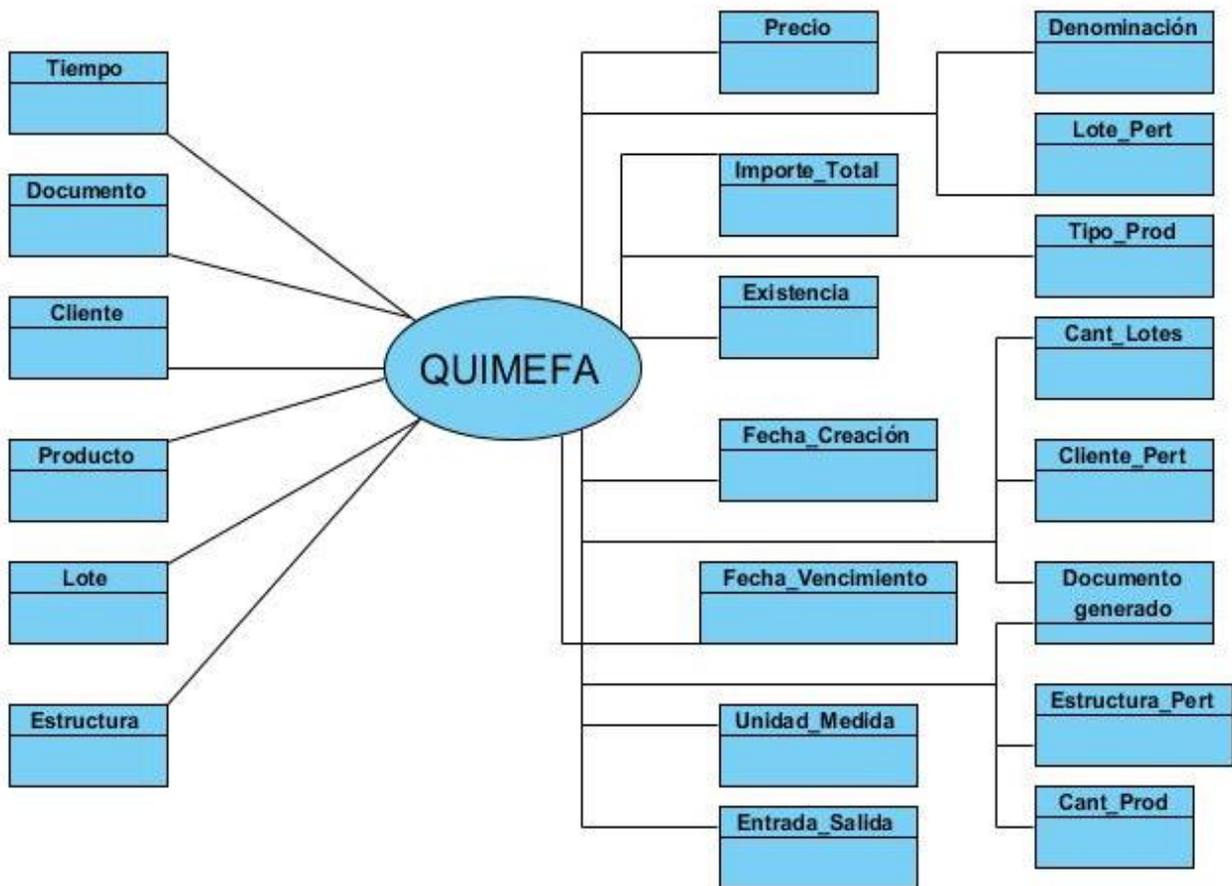


Figura 14 Modelo Conceptual del data mart

A través de este modelo, se observó cuales son los alcances del proyecto para luego poder trabajar sobre ellos. Además al poseer un alto nivel de definición de los datos, permitió ser presentado ante los usuarios, explicándolo con facilidad.

2.4.1.5 Aprobación del Modelo Conceptual por el cliente

Para asegurar que el modelo conceptual se construyó correctamente y que reflejó lo que el cliente necesitaba, se aprobó por el mismo en una reunión donde se expuso el Modelo Conceptual, utilizando siempre un lenguaje entendible para ambas partes. Luego se dio paso a la firma del acta de validación, donde quedó reflejado el mutuo acuerdo con los resultados por ambas partes.

2.4.2 Fase 2: Análisis de los OLTP

2.4.2.1 Definir estado general de los sistemas fuentes

Con el fin de realizar un análisis de los OLTP se comenzó a definir el estado general de los sistemas fuentes, en cuanto a la disponibilidad, accesibilidad, información referente a cada una de las fuentes de datos, características técnicas como: el tipo de Sistema de Administración de Base de Datos con que se administran los sistemas fuentes y el tamaño de la base de datos (en GB) y además se tuvo en cuenta otros aspectos de orden general como: departamento a que pertenece esta base de datos o quién es el administrador de la misma.

Los datos de origen se obtuvieron de la base de datos que tiene por nombre "quimefa_integrada", dicha base de datos se encuentra soportada por el gestor de bases de datos PostgreSQL y pertenece al departamento de San Antonio de la empresa QUIMEFA.

2.4.2.2 Determinación de indicadores

En este paso se mostró de forma explícita como se calculan los indicadores, definiendo los siguientes conceptos para cada uno de ellos:

- El precio de los productos se obtiene, sumando los precios individuales de cada uno de ellos según el lote al que pertenecen.
- El importe total de los productos es la sumatoria de cada uno de los precios de cada producto sin importar el lote al que pertenece.
- La existencia de un producto es la cantidad disponible de dicho producto en una determinada entidad.
- La cantidad de lotes vencidos se filtra por producto para luego sumarlos.

Funciones utilizadas: SUM y COUNT.

2.4.2.3 Establecer correspondencias

El objetivo de este paso, es el de examinar los OLTP o base de datos relacional disponible y comprobar que contengan la información requerida para poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos. La idea es, que todos los elementos del modelo conceptual (perspectivas e indicadores) estén presentes en los OLTP. A continuación en la figura # 15 se muestra una parte de las correspondencias realizadas teniendo en cuenta las perspectivas Cliente y Estructura además del indicador Cant_lotes, incluyendo también las tablas de la base de datos relacional cliente, estructura y detalles_venta.

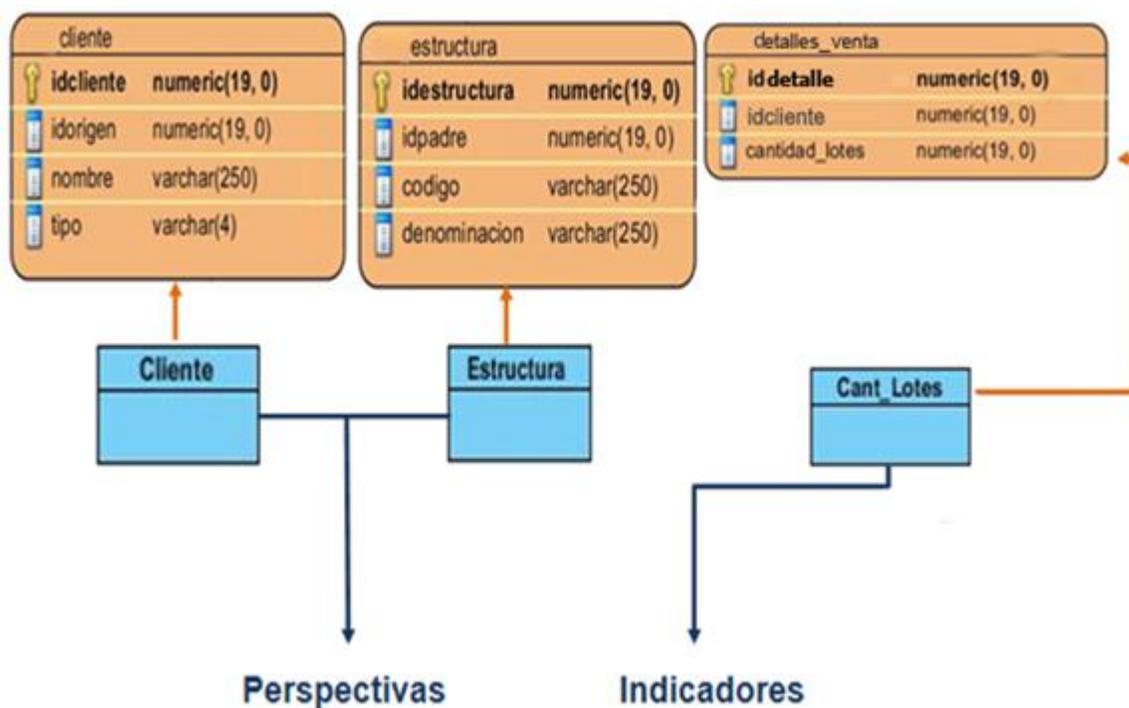


Figura 15 Correspondencias con la base de datos origen

Las relaciones identificadas fueron las siguientes:

- La tabla “Cliente” se relaciona con la perspectiva “cliente”.
- La tabla “Estructura” se relaciona con la perspectiva “estructura”.
- El campo “cantidad_lotes” de la tabla “detalles_venta” con el indicador “Cant_Lotes”.

Es importante aclarar que algunos de los valores correspondientes a los indicadores no se encontraron de forma explícita dentro de una tabla, pues fueron valores que se obtuvieron mediante consultas realizadas a varias de estas.

2.4.2.4 Nivel de granularidad

Una vez que se establecieron las relaciones con los OLTP, se examinaron y se seleccionaron los campos contenidos en cada perspectiva, ya que a través de estos se manipulan y se filtran los indicadores.

Para ello, basándose en las correspondencias establecidas en el paso anterior, se presentó al usuario los datos de análisis disponibles para cada perspectiva. Es importante conocer en detalle qué significa cada campo y/o valor de los datos encontrados en los OLTP, por lo cual, fue conveniente que el encargado del sistema, proporcionara al equipo de desarrollo como entrada para este paso, un diccionario de datos.

Luego de exponer frente al usuario, los datos existentes explicando su significado, valores posibles y características, este decidió cuáles son los que consideraba relevantes para consultar los indicadores y cuáles no.

Se seleccionaron los campos que integran cada perspectiva garantizando la granularidad de la información encontrada en el data mart.

De acuerdo a las correspondencias establecidas, el encargado de la BD proporcionó un documento con un análisis de los campos residentes en cada tabla a la que se hizo referencia. Explicando en cada caso el significado de los mismos. A continuación se muestra un ejemplo de algunos datos referentes a la perspectiva cliente.

- Id_Cliente: es la clave primaria de la tabla “Cliente”, y representa unívocamente a un cliente en particular.
- Código: representa el código del cliente. Este campo es calculado de acuerdo a una combinación de las iniciales del nombre del cliente, el grupo al que pertenece y un número incremental.
- Nombre: representa el nombre del cliente involucrado en una determinada operación.

2.4.2.5 Modelo conceptual ampliado

En este paso, y con el fin de graficar los resultados obtenidos en los pasos anteriores, se amplió el modelo conceptual, colocando bajo cada perspectiva los campos elegidos y bajo cada indicador su respectiva fórmula de cálculo, quedando este debidamente documentado en la planilla “Modelo Conceptual Ampliado”. A continuación en la figura # 16 se muestra la gráfica del modelo conceptual ampliado.

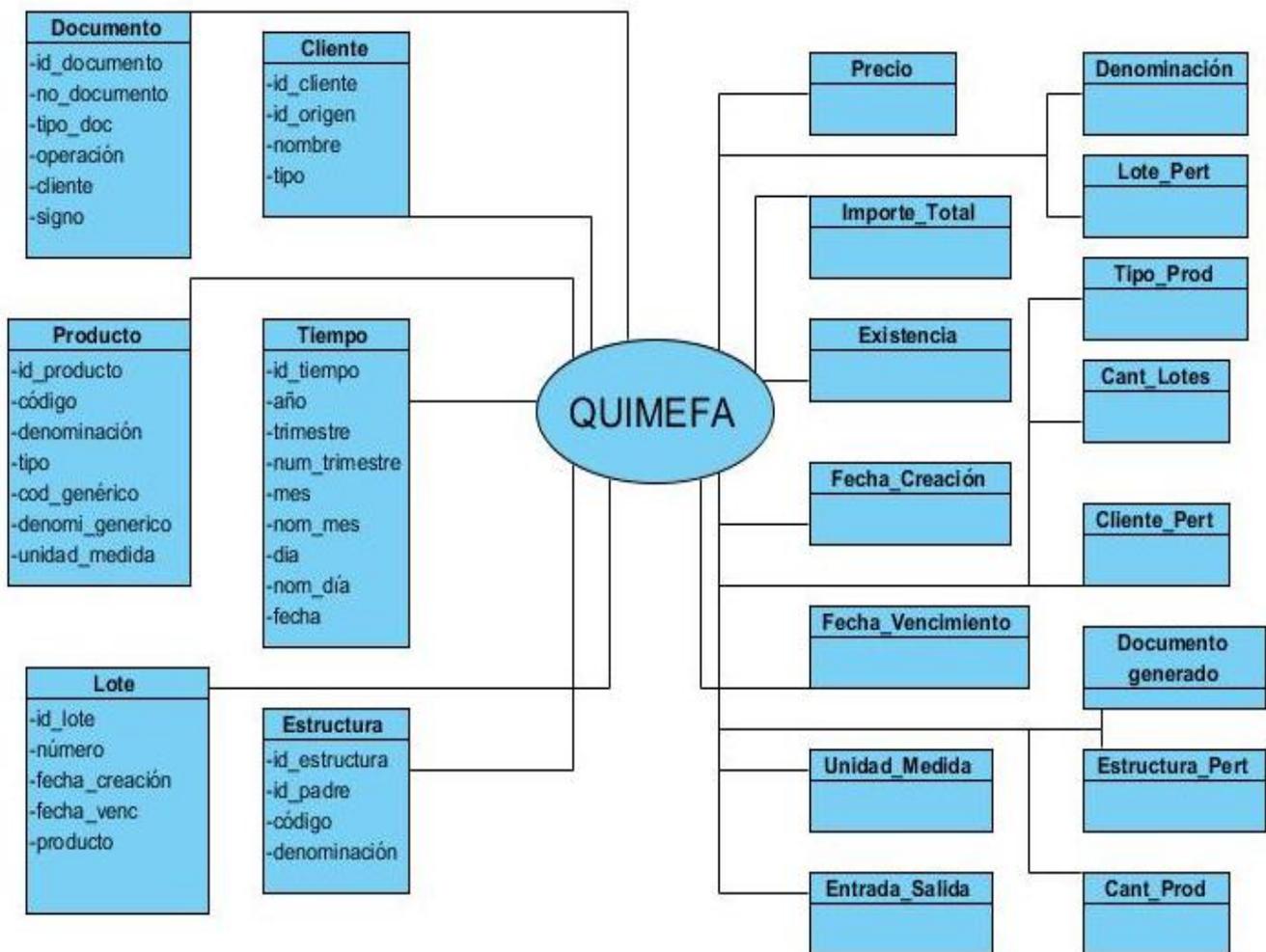


Figura 16 Modelo Conceptual Ampliado

2.4.2.6 Identificar dimensiones

Teniendo como apoyo el modelo conceptual ampliado se crean las dimensiones, donde cada perspectiva que se definió constituye una tabla de dimensión.

Para ello se tomó cada perspectiva con sus campos relacionados y se realizó el siguiente proceso:

- Se eligió el nombre que identifica la tabla de dimensión.
- Se añadió un campo que representa su clave principal.
- Se redefinieron los nombres de los campos que no eran suficientemente intuitivos.

A continuación en la figura # 17 se muestra como quedaron confeccionadas las tablas de dimensiones a partir de las perspectivas.

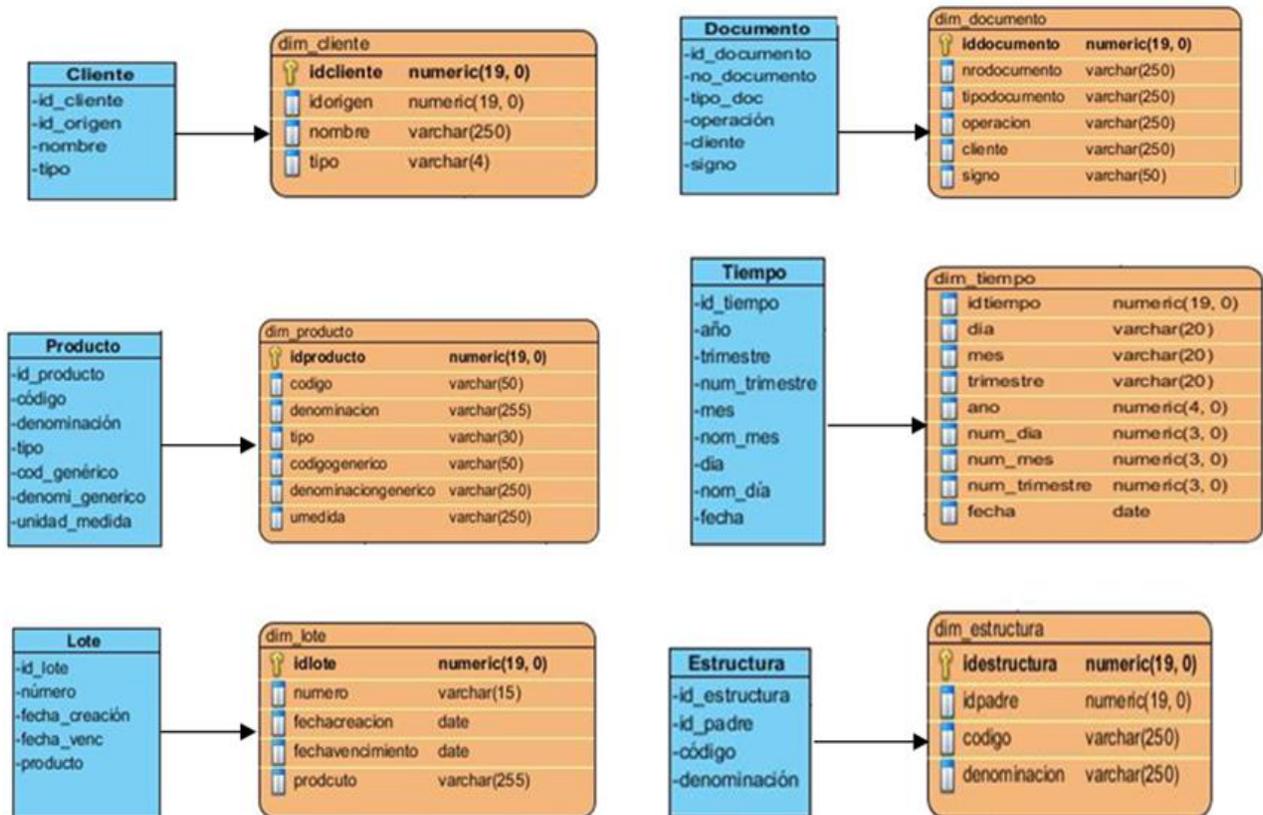


Figura 17 Identificación de dimensiones

2.4.2.7 Identificar hechos

A continuación se muestra como se confeccionaron las tablas de hechos o cubos, utilizando para ello el siguiente ejemplo: tabla de hecho Cubo_Vencimiento.

- La tabla de hechos tiene el nombre “cub_Vencimiento”.
- Su clave principal es la combinación de las claves principales de las tablas de dimensiones antes definidas que tienen relación con la tabla de hechos: “id_estructura”, “id_lote”, “id_producto”, “id_tiempo”.

La figura # 18 muestra la relación que existe entre los indicadores seleccionados con la tabla de hechos.

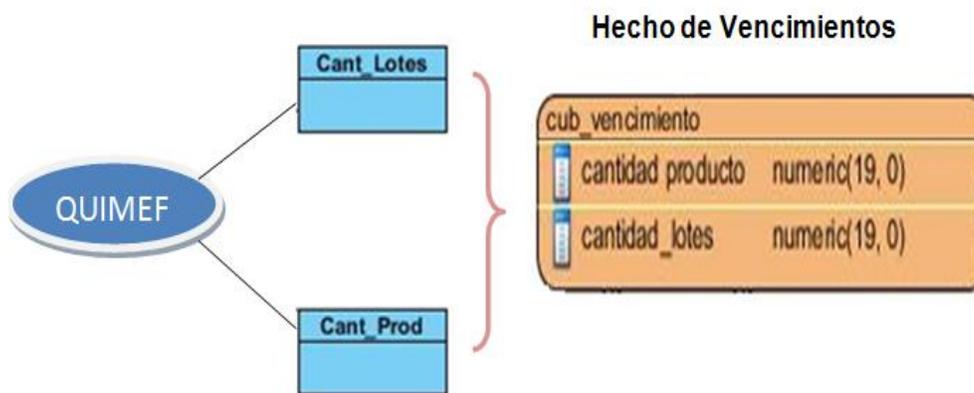


Figura 18 Relación hecho-indicador

2.4.3 Fase 3: Modelo lógico del data mart

A continuación, se confeccionó el modelo lógico de la estructura del data mart, teniendo como base el modelo conceptual creado. Para ello primero se definió el tipo de modelo que se utilizó, eligiéndose constelación por las ventajas y características mencionadas en el capítulo anterior. Posteriormente se diseñaron las tablas y columnas físicas del data mart y finalmente se establecieron las relaciones pertinentes entre ellas. En la figura # 19 se muestra una parte de la estructura lógica del data mart.

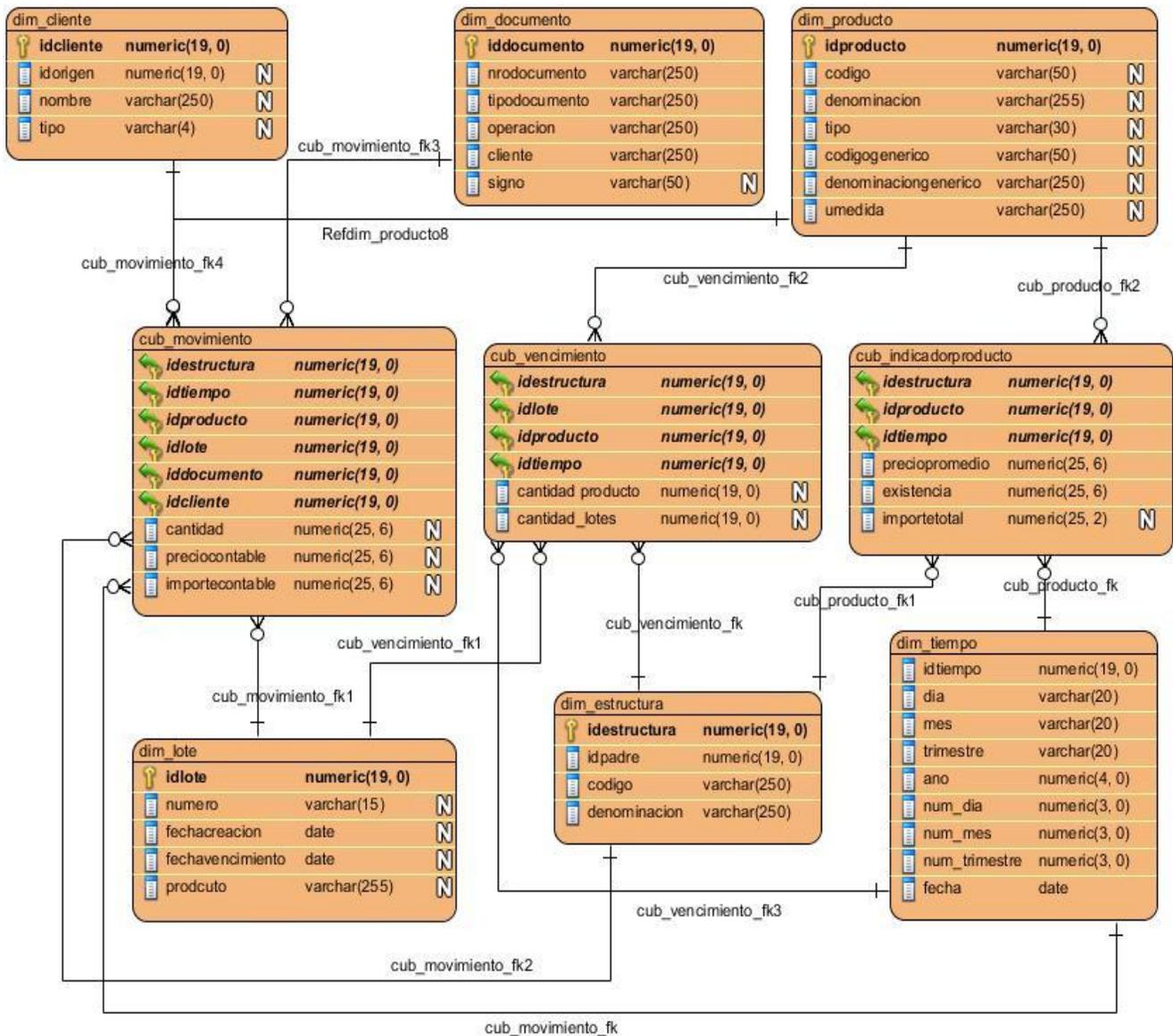


Figura 19 Estructura del data mart para QUIMEFA

2.4.4 Fase 4: Proceso de Extracción Transformación y Carga

2.4.4.1 Mapeo de datos Fuente-A-Destino

Se estableció un mapeo de los datos desde la fuente hacia el destino de los mismos con el objetivo de evitar pérdidas de datos en el proceso de Extracción, Transformación y Carga (ETL). Esta documentación quedó guardada en la planilla “Mapa Lógico de Datos”.

A continuación en la tabla # 2 se muestra un fragmento de este proceso.

Tabla 2 Mapeo de datos

Mapa Lógico de Datos					
Nombre BD Fuente: quimefa_integrada			Nombre BD Destino: BD data mart		
Nombre de la tabla	Nombre del campo	Tipo de dato	Nombre de la tabla	Nombre del campo	Tipo de dato
Producto	Id producto	NUMERIC (19,0)	dim_producto	id producto	NUMERIC (19,0)
Producto	Unidad de medida	VARCHAR (25)	dim_producto	UM	VARCHAR(25)
Cliente	Id cliente	NUMERIC (19,0)	dim_cliente	idcliente	NUMERIC (19,0)
Cliente	Id origen	NUMERIC (19,0)	dim_cliente	Id_origen	NUMERIC (19,0)

ETL son las siglas en Inglés de Extraer, Transformar y Cargar (Extract, Transforms and Load). Este proceso permite la transferencia de datos desde múltiples fuentes, este caso en particular fue de la base de datos perteneciente al sistema integral de medicamentos QUIMEFA, donde posteriormente los datos fueron reformateados y limpiados para ser cargados en la base de datos del data mart.

En este proceso se llevaron a cabo una serie de pasos los cuales tuvieron como objetivo fundamental, el traspaso de datos desde las aplicaciones de producción de QUIMEFA al sistema para la toma de decisiones. La aplicación ETL Pentaho Data Integration fue la encargada de leer los datos primarios, realizar transformaciones, hacer un proceso cualitativo y una correcta filtración para en un final tener dichos datos disponibles.

- 1- Primer paso:** extracción de datos de la base de datos origen, la cual tiene como característica ser una base de datos relacional, esta extracción convirtió los datos en un formato preparado para iniciar el proceso de transformación. Una parte intrínseca que se tuvo en cuenta, fue analizar los datos extraídos. Este chequeo permitió verificar si los datos cumplían la pauta o estructura que se esperaba y de no ser así los datos se rechazaron. Un requerimiento importante que se asumió es que la extracción causara un impacto mínimo en el sistema origen, ya que si los datos a extraer eran muchos, el sistema de origen podría ralentizar e incluso colapsar. Particularmente en este caso no se identificaron fallas de este tipo. A continuación se muestra en la figura # 20 la configuración de la conexión a la base de datos origen.

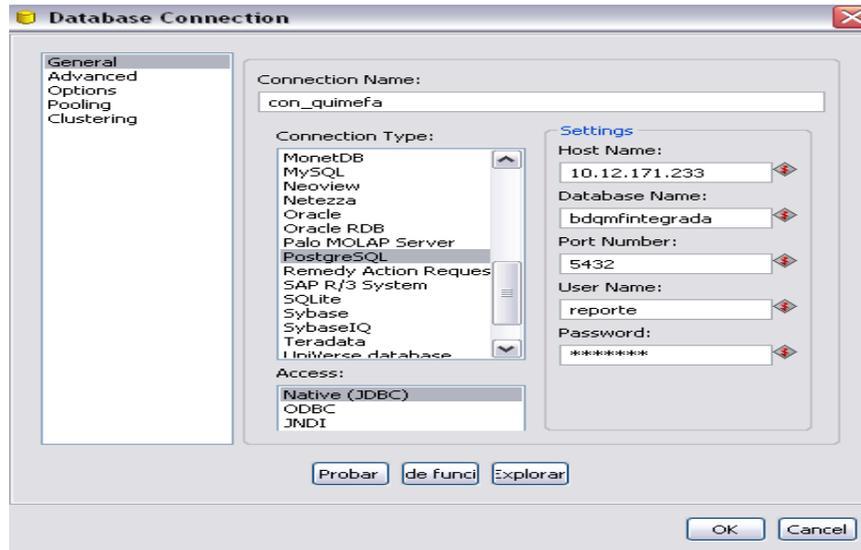


Figura 20 Base de datos origen (QUIMEFA)

2- **Segundo paso:** en la fase de transformación se aplicaron una serie de reglas de negocio o funciones sobre los datos extraídos de la BD QUIMEFA para convertirlos en datos cargados. Algunas fuentes requerían de una pequeña manipulación de datos, ejemplo figura # 21.

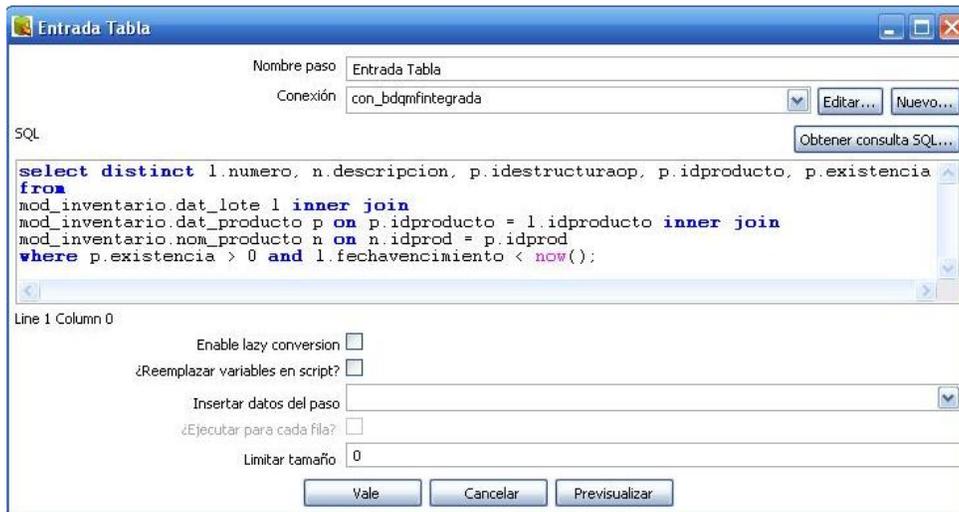


Figura 21 Selección de los datos de lotes vencidos

A continuación se mencionan algunas de las siguientes transformaciones:

- Seleccionar sólo ciertas columnas para su carga (por ejemplo, que las columnas con valores nulos no se carguen).
- Codificar valores libres (por ejemplo, convertir “Hombre” en “H” o “Sr” en “1”).
- Obtener nuevos valores calculados (por ejemplo, total_venta = cantidad * precio).
- Calcular totales de múltiples filas de datos (por ejemplo, precio total de cada región).
- Generación de campos clave en el destino.
- Datos erróneos: ejecutar políticas de tratamiento de excepciones (por ejemplo, rechazar el registro completo, dar al campo erróneo un valor nulo o un valor centinela).

3- Tercer paso Fase de carga: los datos de la fase anterior **transformación** fueron cargados en el sistema destino, es decir la BD del data mart, ejemplo figura # 22. En los anexos 2.1, 2.2 y 2.3 se muestran las demás cargas correspondientes a dicho proceso.

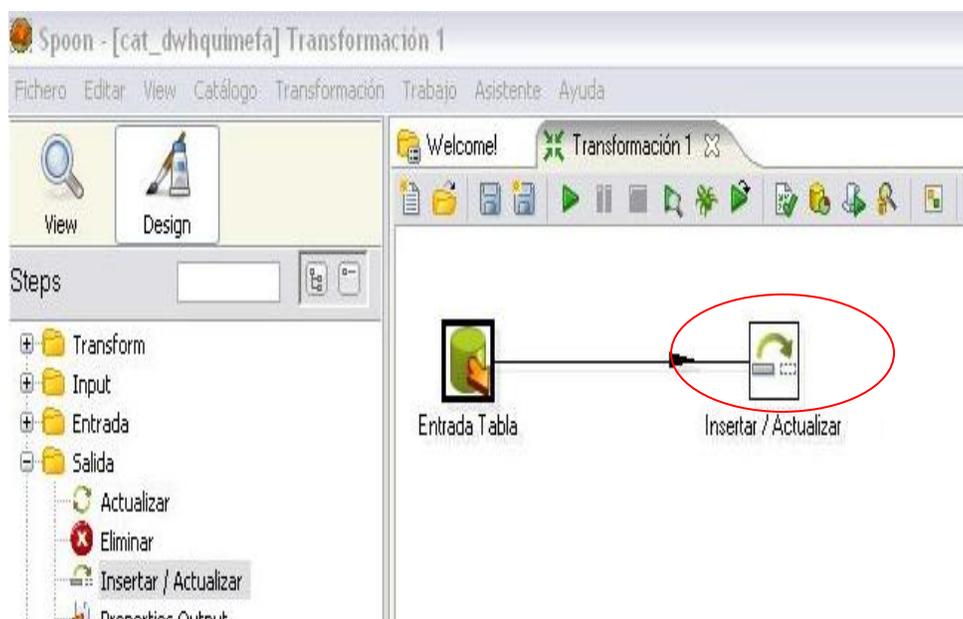


Figura 22 Carga de datos

2.5 Conclusiones Parciales

En el capítulo se desarrollaron cada uno de los pasos de la metodología Híbrida. Se obtuvieron las preguntas claves del negocio, identificándose los indicadores y perspectivas del análisis. Se elaboró el modelo conceptual de datos proporcionando una idea clara del alcance del data mart. Además, se

confeccionó el modelo lógico de la estructura del data mart para luego realizar el proceso de extracción, transformación y carga de los datos en el almacén.

Capítulo 3 Proceso Analítico en Línea

3.1 Introducción

En este capítulo se muestra la construcción de los cubos de datos, donde se definen las jerarquías y atributos de las dimensiones a las que pertenecen y la implementación de los indicadores o medidas. Se muestran las herramientas disponibles para que los usuarios puedan consultar la información almacenada, la cual puede ser analizada mediante tablas de datos y esquemas con un alto grado de definición. Además se muestran los resultados obtenidos en la construcción del data mart.

3.2 Jerarquía de las tablas de dimensiones

Determinar la jerarquía de las dimensiones resulta de gran importancia para organizar los niveles dentro de una dimensión, permitiendo un orden lógico de la información que se analiza ya sea de un nivel general a uno específico o viceversa. Las jerarquías que estuvieron presentes fueron: Tiempo y Estructura.

A continuación se muestra en la figura # 23 la jerarquía de alguno de los atributos de la dimensión tiempo.

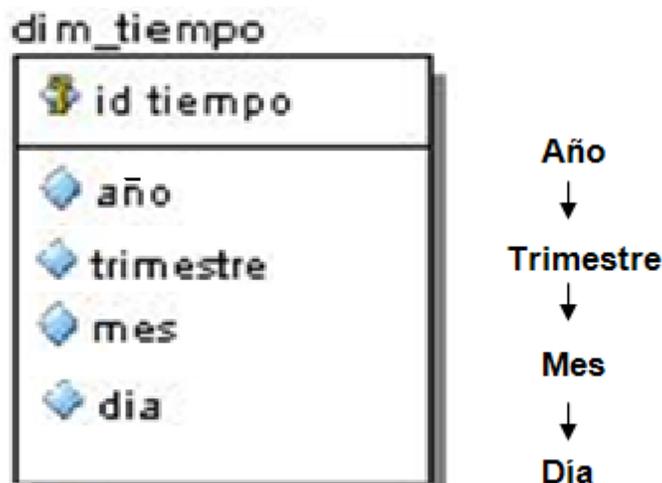


Figura 23 Jerarquía de los atributos de la dimensión tiempo

3.3 Medidas

Una vez establecidas las jerarquías de cada una de las dimensiones, se elaboraron las medidas o indicadores en busca de las respuestas de lo que se deseaba conocer por los directivos de QUIMEFA. Para ello se utilizaron consultas MDX (multidimensional expressions o expresiones multidimensionales) que se ejecutan sobre el data mart para obtener los indicadores que contribuyen a la toma de decisiones.

Desde la base de datos del data mart se cargaron las dimensiones y las tablas de hechos para calcular las medidas. Posteriormente se utilizó la herramienta Workbench para diseñar los cubos de datos, los cuales fueron nombrados de la siguiente manera: Existencia de producto, Movimientos generales, Movimientos de entrada y salida y por último el cubo Vencimiento de lotes. Una vez definidos los cubos se agregaron las medidas y se formularon las consultas MDX, para obtener indicadores como: el precio, el importe total y la existencia de un producto entre otros. En la figura # 24 se muestran dos de los cubos con sus respectivas medidas y dimensiones.

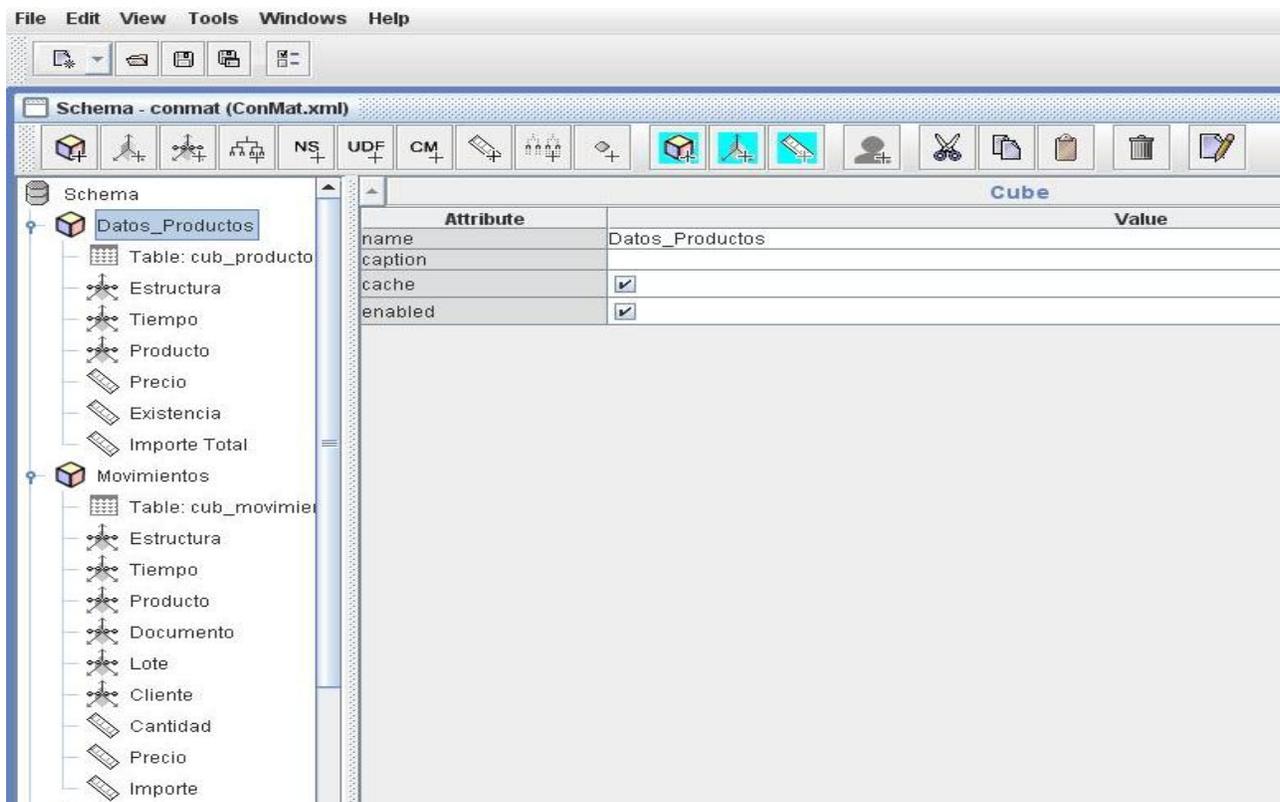
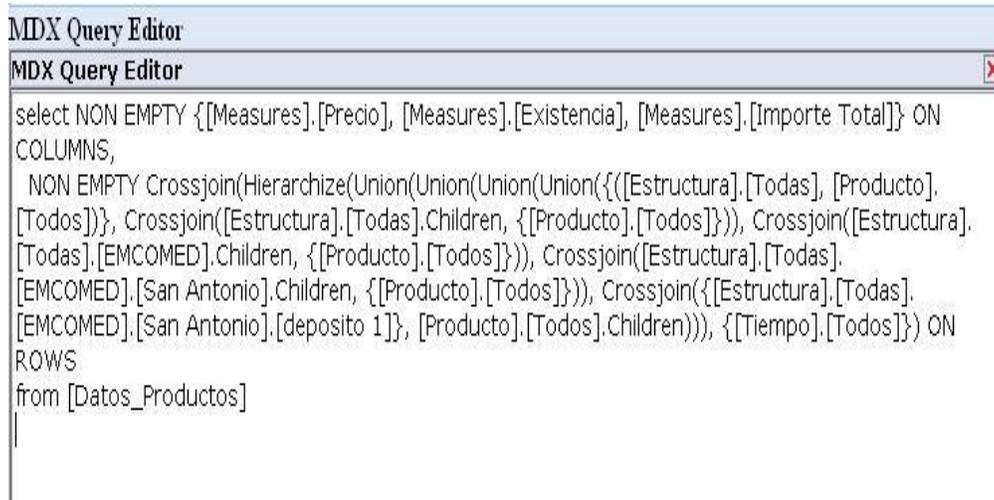


Figura 24 Cubos multidimensionales con sus medidas y dimensiones asociadas

3.4 Consultas

Una de las ventajas de las expresiones multidimensionales es que permiten consultar objetos multidimensionales y devolver un conjunto de celdas que contenga los datos del cubo. Dichas consultas calcularon los datos numéricos visualizados en los resultados.



```
MDX Query Editor
MDX Query Editor
select NON EMPTY {[Measures].[Precio], [Measures].[Existencia], [Measures].[Importe Total]} ON
COLUMNS,
NON EMPTY Crossjoin(Hierarchize(Union(Union(Union(Union({([Estructura].[Todas], [Producto].
[Todos])), Crossjoin([Estructura].[Todas].Children, {[Producto].[Todos]))}, Crossjoin([Estructura].
[Todas].[EMCOMED].Children, {[Producto].[Todos]))}, Crossjoin([Estructura].[Todas].
[EMCOMED].[San Antonio].Children, {[Producto].[Todos]))}, Crossjoin({[Estructura].[Todas].
[EMCOMED].[San Antonio].[deposito 1], [Producto].[Todos].Children))), {[Tiempo].[Todos]}) ON
ROWS
from [Datos_Productos]
```

Figura 25 Consultas MDX

3.5 Resultados para los clientes de Mondrian

Una vez creados los cubos, sus componentes fueron almacenados en un archivo XML, para luego cargarlos en la herramienta Mondrian: herramienta escogida por su alto desempeño para realizar el procesamiento analítico en línea.

Con el cliente Jpivot disponible a través de la web, se interpretaron los datos contenidos en el fichero XML para visualizarlos. Una vez visualizado los datos se permite mostrar la información desde diferentes perspectivas según las opciones deseadas por el usuario, es decir filtrar los campos que se desea analizar. Los resultados de los informes son mostrados mediante tablas con datos ordenados y gráficas de comportamientos, como se muestran en las figuras # 26, 27 y 28.

Aplicaciones << Mondrian

Cubos OLAP en Mondrian

Buscar [] Buscar

Catalogos de Cubos OLAP

- Guimefa
 - Existencia Productos
 - Movimientos Generales
 - Movimientos E / S
 - Vencimiento Lotes

Resultado de Mondrian para Almacen de Datos: Vencimiento Lotes

Estructura			Medidas	
Estructura	Lote	Producto	cantidad producto	cantidad lote
-Todas	+Todos	+Todos	6.751	26
-EMCOMED	+Todos	+Todos	6.751	26
-San Antonio	+Todos	+Todos	6.751	26
deposito 1	-Todos	+Todos	6.751	26
	11006	-Todos	820	1
		-337	820	1
		33796320500000	820	1
	11008	+Todos	820	1
	123	+Todos	820	1
	450665	+Todos	7	1
	lote AA	+Todos	7	1
	8565	+Todos	300	1
	11004	+Todos	560	1
	54	+Todos	560	1
	11001	+Todos	129	1
	12	+Todos	129	1
	3	+Todos	129	1

Figura 26 Cubo Vencimiento de Lotes con sus dimensiones y medidas según el filtrado

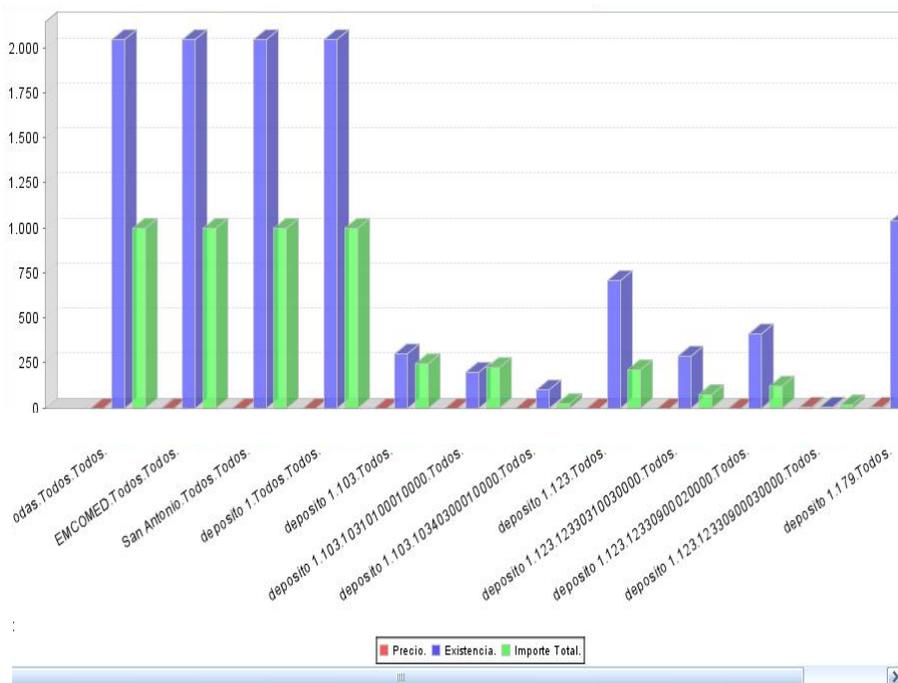


Figura 27 Informaciones gráficas



Figura 28 Otra vista de información gráfica

3.6 Validación

3.6.1 Objetivos de la validación

La validación se hace con el objetivo de encontrar defectos en el data mart y comprobar que cumple con los requisitos funcionales y no funcionales que se identificaron en las entrevistas realizadas a los directivos de la empresa QUIMEFA. Dentro de los aspectos generales que se evalúan se encuentran: visualización correcta de los datos, rendimiento de la aplicación en la transferencia de datos y la integración de la aplicación con otras herramientas de software.

Para dar cumplimiento a dicho objetivo se realizaron las siguientes pruebas de validación:

- Pruebas de conceptos.
- Pruebas de integración.
- Pruebas de validación con el cliente.
- Pruebas de sistema.

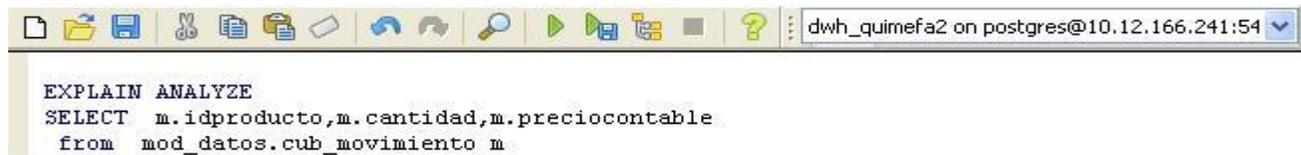
3.6.2 Validación de pruebas mediante pruebas de conceptos

Una de las estrategias utilizadas para la evaluación del diseño de la base de datos del data mart fueron las pruebas de conceptos, las cuales dentro de sus objetivos principales tienen la descripción de todas las pruebas que se realizan a nivel de diseño, así como validar el negocio implementado en la base de datos a partir de los escenarios. En este caso el escenario escogido fue, los reportes definidos por los clientes.

La visualización de los indicadores: movimientos de productos del almacén y control de lotes de medicamentos es realizada a través de reportes, los cuales fueron seleccionados para realizar la validación de las pruebas.

Reporte del Ciclo de movimientos generales de productos

Para realizar la prueba a dicho reporte se insertaron en las tablas un total de 541998 tuplas de datos, para ejecutar luego la siguiente consulta:



```
EXPLAIN ANALYZE
SELECT m.idproducto,m.cantidad,m.preciocontable
from mod_datos.cub_movimiento m
```

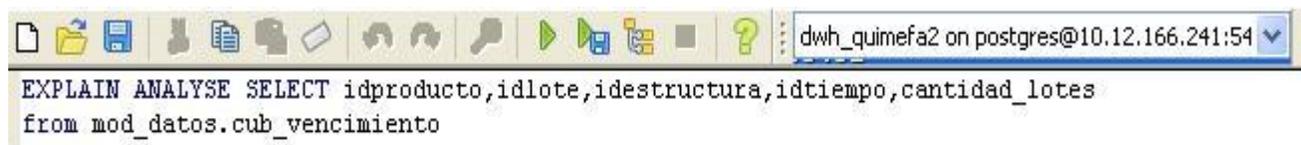
Esta consulta obtiene datos asociados a los movimientos generales de productos. Luego de ejecutada la consulta se obtuvo los siguientes resultados:

WindowAgg (cost=176697.32..186160.13 rows=54073 width=35) (actual time=38097.735..40461.934 rows=541998 loops=1) Total runtime: 50679.457 ms

La base de datos para esta función, cumple con un tiempo de respuesta aceptable, pues devuelve 6 filas en 50,688 segundos. De esta manera queda demostrado que el reporte asociado a los movimientos generales de producto se ejecuta correctamente.

Reporte Lotes vencidos

Para realizar la prueba a dicho reporte se insertaron en las tablas un total de 541998 tuplas de datos, para ejecutar luego la siguiente consulta:



```
EXPLAIN ANALYZE SELECT idproducto,idlote,idestructura,idtiempo,cantidad_lotes
from mod_datos.cub_vencimiento
```

Esta consulta devuelve la cantidad de lotes vencidos en un tiempo determinado.

Luego de ejecutada la consulta se obtuvo los siguientes resultados:

WindowAgg (cost=187785.32..213470.09 rows=540732 width=63) (actual time=44478.087..54786.774 rows=541998 loops=1)

Al igual que en reporte anterior, la base de datos para esta función cumple con un tiempo de respuesta aceptable, pues devuelve 8 filas en 55,016 segundos.

De manera general el comportamiento de ambos reportes no muestran dificultades.

Fecha de realización: 25 /5/2011

Responsable: Yuniesky Nueva Aguilar

Porcentaje de aprobación: 100%

3.6.3 Pruebas de Integración

Las pruebas de integración son aquellas en las que se examina la solución data mart y se verifica su integración a otras soluciones o aplicaciones a través del uso de un diseño documentado y de estándares de programación.

Para comprobar si la herramienta integra sus reportes a otras aplicaciones se seleccionaron algunos datos del cubo de productos, para posteriormente transformarlos en formatos de hojas electrónicas de cálculo y formato PDF. A continuación se presenta la muestra seleccionada en la figura # 29.

	A	B	C	D	E	F
1				Medidas		
2	Estructura	Producto	Tiempo	Precio	Existencia	Importe Total
3	Todas	Todos	Todos	2	5185906,	7.288.934,79
4	EMCOMED	Todos	Todos	2	5185906,	7.288.934,79
5	San Antonio	Todos	Todos	2	5185906,	7.288.934,79
6	deposito 1	Todos	Todos	2	5185906,	7.288.934,79
7		#null	Todos	3	2074,	6412,
8		103	Todos	1	300	246
9		1,03101E+13	Todos	1	200	223
10		1,03403E+13	Todos	0	100	23
11		123	Todos	1	707	214
12		179	Todos	2	5159292,	7.239.288,95
13		337	Todos	6	2010,	23.755,6
14		338	Todos	2	19125,	17.998,54
15		341	Todos	0	2115,	887
16		750	Todos	3	4	12
17		754	Todos	0	100	0
18	756	Todos	1	129	103	
19	774	Todos	3	50	18	

Estructura	Producto	Tiempo	Medidas	Existencia	Importe Total
Todas	Todos	Todos	Precio 2,195	5.185.906	7.288.934,79
EMCOMED	Todos	Todos	2,195	5.185.906	7.288.934,79
San Antonio	Todos	Todos	2,195	5.185.906	7.288.934,79
deposito 1	Todos	Todos	2,195	5.185.906	7.288.934,79
	#null	Todos	2,912	2.074	6.412
	103	Todos	0,672	300	246
	10310100010000	Todos	1,115	200	223
	10340300010000	Todos	0,23	100	23
	123	Todos	1,012	707	213,5
	179	Todos	2,126	5.159.292	7.239.288,95
	337	Todos	6,248	2.010	23.755,6
	338	Todos	2,053	19.125	17.998,54
	341	Todos	0,439	2.115	887,3
	750	Todos	3	4	12
	754	Todos	0	100	0
	756	Todos	0,798	129	102,9
	774	Todos	3	50	18

Figura 29 Reportes

Fecha de realización: 25 /5/2011

Responsable: Yuniesky Nueva Aguilar

Porcentaje de aprobación: 100%

3.6.4 Prueba de validación del cliente

Las pruebas de validación se enfocan para saber si la solución propuesta data mart es la herramienta que el cliente desea, es decir si se ha cumplido con todos los requerimientos que se acordaron con el cliente.

Para esta prueba fue necesario valerse de la documentación que contienen los requisitos, de forma tal que se pueda ver el cumplimiento sin omitir detalles. En esta fase se tuvo un intercambio con el cliente, donde se mostró al mismo las funcionalidades del sistema. Luego de la revisión el cliente constató que el data mart cumplió con las normas establecidas y dio paso a crear el acta de validación. El acta se muestra en el anexo # 3.

Fecha de realización: 25 /5/2011

Responsable: Yuniesky Nueva Aguilar

Cliente: Jefe de proyecto de la empresa QUIMEFA, Ing. Adail Sariego Mazpule.

Porcentaje de aprobación: 100%

3.6.5 Pruebas del sistema

Las pruebas de sistemas se encargan de saber si todos los componentes que permiten el funcionamiento de la solución data mart trabajan correctamente y se examinan las interfaces que utilizan los usuarios para visualizar la información apropiada.

A continuación se muestran las pruebas realizadas para validar el correcto funcionamiento del data mart.

Prueba de sistema # 1

La primera prueba de sistema consistió en desplegar el data mart hacia otro servidor con características diferentes en cuanto a: sistema operativo, capacidad de almacenamiento y procesamiento de los datos.

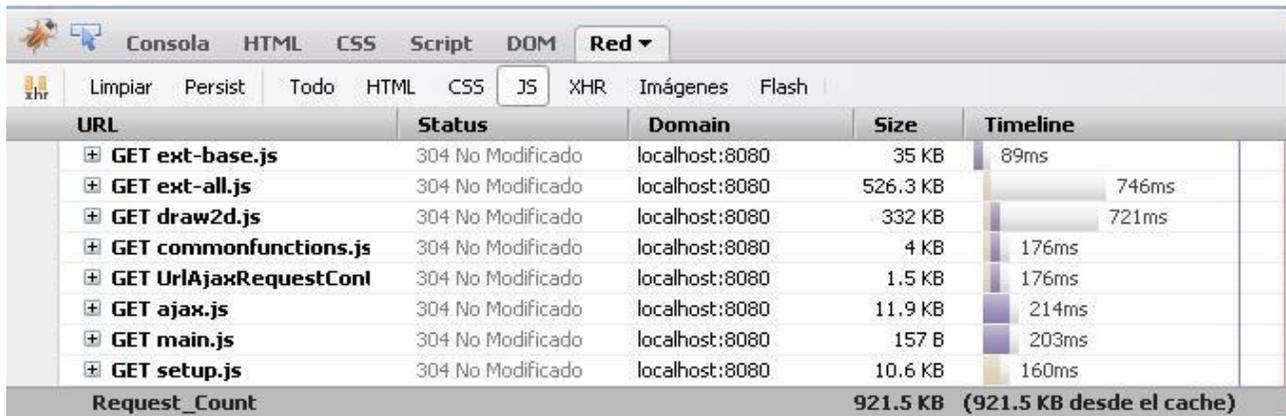
Tabla 3 Características de servidores

Aspectos	Servidor Origen	Servidor a desplegar
Sistema operativo	Window	Debian GNU Linux 6.0
Memoria RAM	1GHz	512 MB
Microprocesador	3.0	3.0
Capacidad de disco duro	120GB	500 GB

Luego de desplegado el data mart, se observó que la aplicación funcionó correctamente aunque tuvo un pequeño descenso de procesamiento de datos. Esta prueba se utilizó como medida a la hora de proponer el despliegue de la aplicación.

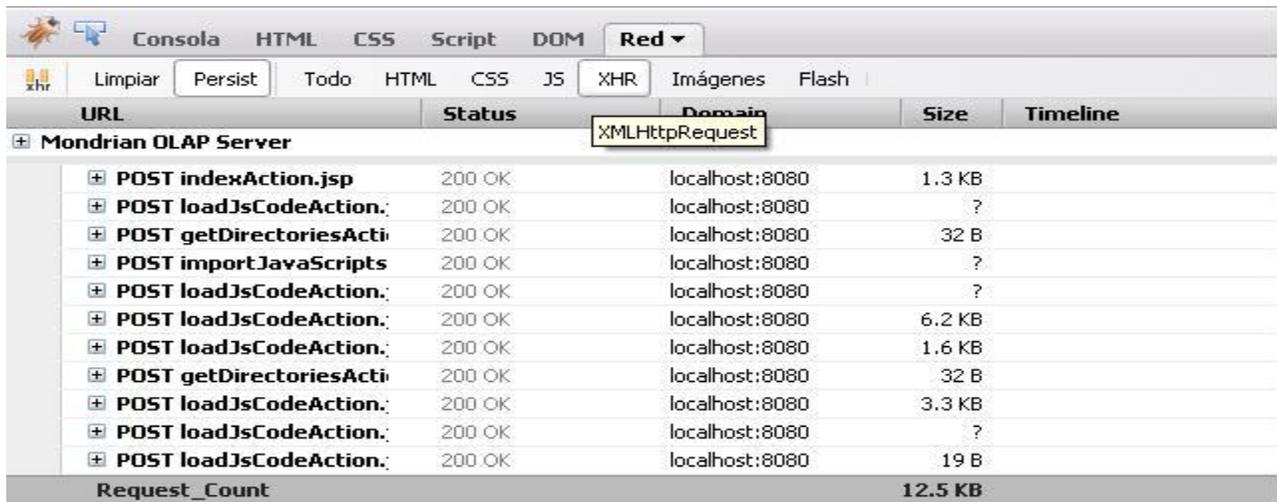
Prueba de sistema # 2

Esta prueba consiste en medir el tiempo de ejecución de los distintos procesos pertenecientes a la aplicación, utilizando el plugin Firebug del navegador web Mozilla Firefox. A continuación en las figuras # 30, 31 y 32 se muestran dichas pruebas.



URL	Status	Domain	Size	Timeline
GET ext-base.js	304 No Modificado	localhost:8080	35 KB	89ms
GET ext-all.js	304 No Modificado	localhost:8080	526.3 KB	746ms
GET draw2d.js	304 No Modificado	localhost:8080	332 KB	721ms
GET commonfunctions.js	304 No Modificado	localhost:8080	4 KB	176ms
GET UrlAjaxRequestConl	304 No Modificado	localhost:8080	1.5 KB	176ms
GET ajax.js	304 No Modificado	localhost:8080	11.9 KB	214ms
GET main.js	304 No Modificado	localhost:8080	157 B	203ms
GET setup.js	304 No Modificado	localhost:8080	10.6 KB	160ms
Request_Count			921.5 KB	(921.5 KB desde el cache)

Figura 30 Pruebas de tiempos de ejecución # 1



URL	Status	Domain	Size	Timeline
Mondrian OLAP Server				
POST indexAction.jsp	200 OK	localhost:8080	1.3 KB	
POST loadJsCodeAction.;	200 OK	localhost:8080	?	
POST getDirectoriesActi	200 OK	localhost:8080	32 B	
POST importJavaScripts	200 OK	localhost:8080	?	
POST loadJsCodeAction.;	200 OK	localhost:8080	?	
POST loadJsCodeAction.;	200 OK	localhost:8080	6.2 KB	
POST loadJsCodeAction.;	200 OK	localhost:8080	1.6 KB	
POST getDirectoriesActi	200 OK	localhost:8080	32 B	
POST loadJsCodeAction.;	200 OK	localhost:8080	3.3 KB	
POST loadJsCodeAction.;	200 OK	localhost:8080	?	
POST loadJsCodeAction.;	200 OK	localhost:8080	19 B	
Request_Count			12.5 KB	

Figura 31 Pruebas de tiempos de ejecución # 2

URL	Status	Domain	Size	Timeline
GET ext-base.js	304 No Modificado	localhost:8080	35 KB	89ms
GET ext-all.js	304 No Modificado	localhost:8080	526.3 KB	746ms
GET draw2d.js	304 No Modificado	localhost:8080	332 KB	721ms
GET commonfunctions.js	304 No Modificado	localhost:8080	4 KB	176ms
GET UrlAjaxRequestContl	304 No Modificado	localhost:8080	1.5 KB	176ms
GET ajax.js	304 No Modificado	localhost:8080	11.9 KB	214ms
GET main.js	304 No Modificado	localhost:8080	157 B	203ms
GET setup.js	304 No Modificado	localhost:8080	10.6 KB	160ms
Request_Count			921.5 KB	(921.5 KB desde el cache)

Figura 32 Pruebas de tiempos de ejecución # 3

Una vez realizadas las pruebas se observó que los tiempos de carga de los proceso de red están acordes y permiten una visualización rápida de la información.

3.7 Despliegue del data mart

Para realizar el despliegue del data mart para QUIMEFA se necesitan un total de 3 servidores, uno para la base de datos de QUIMEFA implementado sobre el gestor de base de datos PostgreSQL 8.3, otro para el data mart sobre el gestor de base de datos PostgreSQL 8.3 y el servidor de aplicación Web donde se ejecuta el Pentaho Mondrian sobre el Apache Tomcat.

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

A continuación en la Figura # 33 se muestra la conformación del diagrama de despliegue.

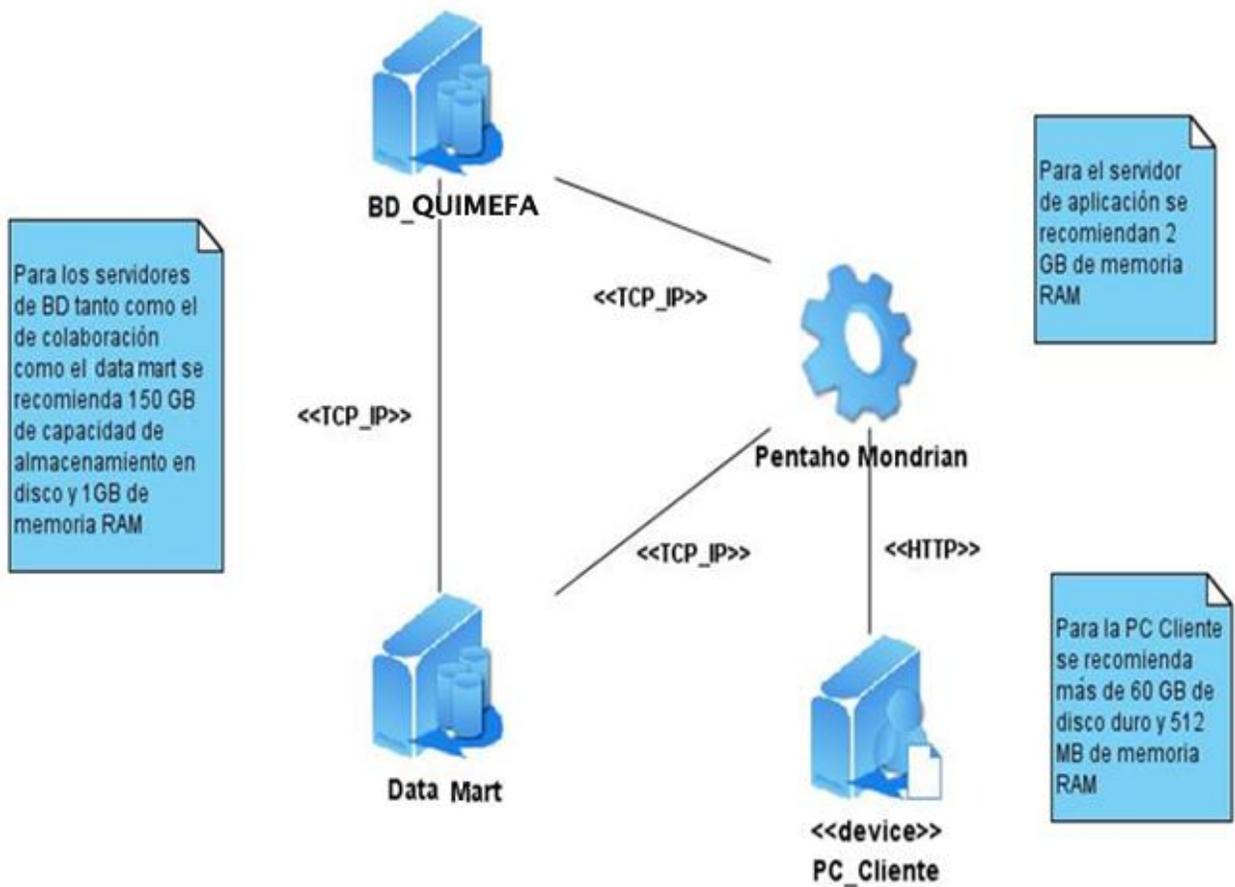


Figura 33 Diagrama de despliegue del data mart

3.8 Conclusiones Parciales

En el capítulo se realizó el procesamiento analítico en línea para agilizar las consultas de la información en el almacén, mostrando los datos a los directivos desde diferentes perspectivas. Además se expuso la forma en que debe realizarse el despliegue de la aplicación en QUIMEFA.

Conclusiones

En el presente trabajo se ha desarrollado un producto de software que consiste en un data mart para la empresa QUIMEFA, que permite el almacenamiento de grandes volúmenes de información relevante, transformando los datos operativos en información analítica enfocada a la toma de decisiones. Las principales contribuciones de este trabajo de diploma son:

- Se realizó un estudio del estado del arte de los principales sistemas de soporte a la toma de decisiones haciendo énfasis en los almacenes de datos y sistemas de software que realizan procesos similares, lo que permitió conocer las tendencias actuales en cuanto a la construcción y diseño de este tipo de herramientas .
- Se obtuvo un análisis de los principales requerimientos de información en QUIMEFA permitiendo al equipo de desarrollo tener una visión global del negocio.
- Se realizó el diseño y la implementación de la solución informática utilizando como guía la metodología Híbrida que cubrió todas las etapas del desarrollo, evitando pasar por alto detalles que causen una mala implementación de la solución data mart.
- Se logró mediante pruebas de Software validar la propuesta de solución, quedando comprobada la satisfacción del cliente.

Principales Aportes

Con el data mart los directivos de QUIMEFA disponen de una herramienta que les permite:

- Tener el control de los productos y el movimiento de los mismos en las diferentes entidades.
- Obtener una representación de los resultados en forma de tabla, gráfica, PDF y Excel.
- Enviar información detallada y consolidada a los niveles superiores.
- Almacenar y analizar la información histórica del sistema integral de medicamentos útil para el apoyo a la toma de decisiones.

Recomendaciones

Con el objetivo de mejorar el uso del data mart para QUIMEFA se recomienda:

- Aplicar técnicas de minería de datos sobre el data mart que les permita a los directivos de QUIMEFA analizar factores de influencia en determinados procesos, predecir, estimar variables o comportamientos futuros, además de obtener secuencias de eventos que provoquen comportamientos específicos que le ayuden a una mejor toma de decisiones.
- Integrar la solución al sistema ERP.
- Realizar nuevas entrevistas al cliente con el objetivo de incluir mejoras al sistema.

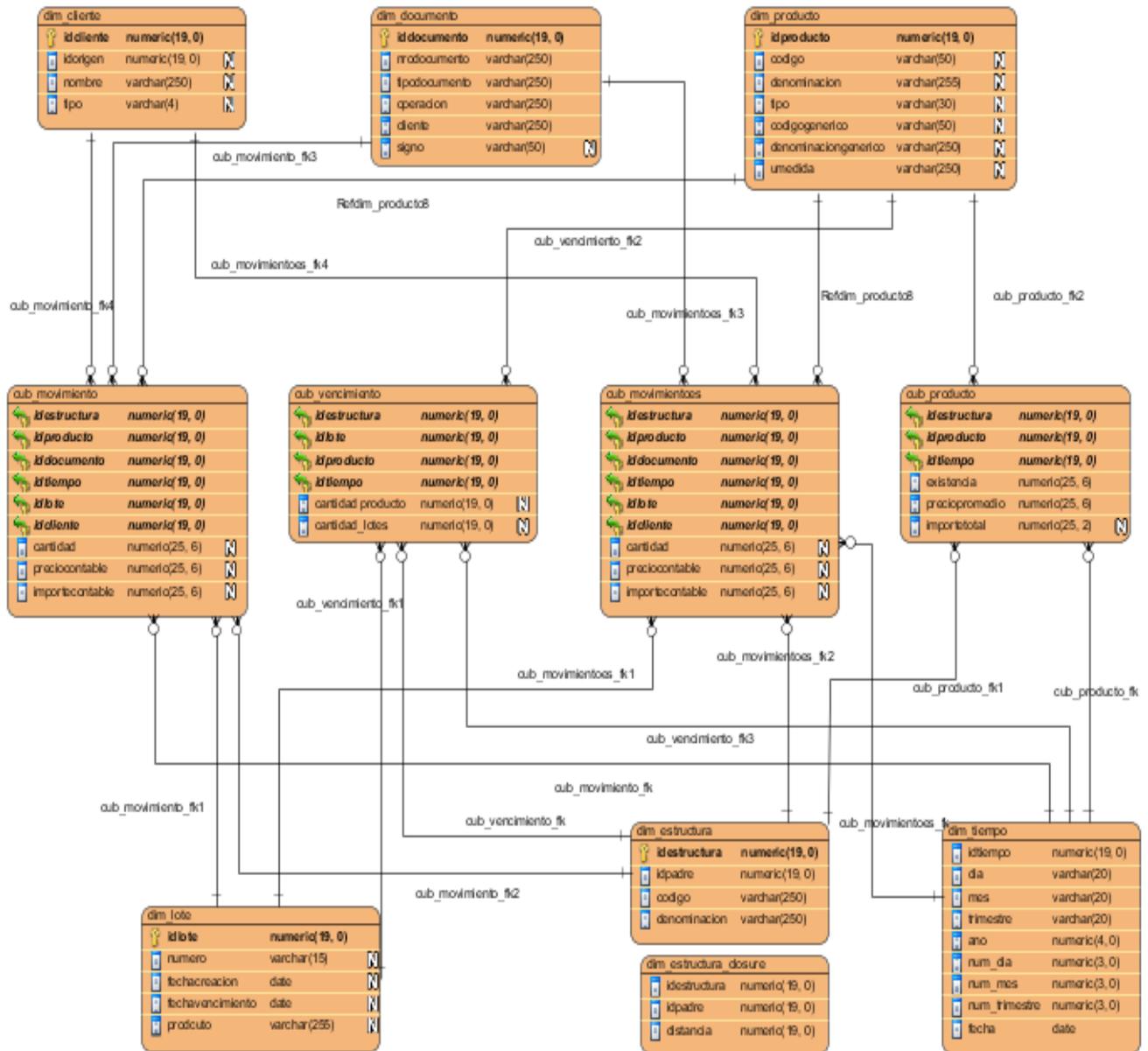
Referencia Bibliografía

1. José Ignacio Domínguez. Gestipolis.com. [En línea] [Citado el: 5 de 1 de 2011.] <http://www.gestipolis.com/Canales4/mkt/erroresinuno.htm>.
2. Arencibia, Mario Gonzáles. *Globalizacion y Mercado:Estrategia y alternativa*. La Habana: s.n., 2007.
3. Ing. Bernabeu Ricardo Dario. *Metodología propia para la Construcción de un Data Warehouse*. Córdoba, Argentina : s.n., 2009.
4. Grupo de Autores Sinnexus. Sinnexus. [En línea] [Citado el: 11 de 1 de 2011.] [http://www.sinnexus.com/business_intelligence/datamart.aspx /](http://www.sinnexus.com/business_intelligence/datamart.aspx/).
5. Roberto Hernando Velasco. Almacenes de datos (Datawarehouse) . [En línea] [Citado el: 7 de 1 de 2011.] <http://www.rhernando.net>.
6. límites, Colectivo de autores de GravatarInformación sin. GravatarInformación sin límites. [En línea] [Citado el: 20 de 10 de 2010.] <http://www.gravatar.biz/>.
7. Caralt, Jordi Conesa. *Introducción a la Business Intelligence*. Primera edición en lengua castellana. Barcelona : UOC1, 2010. 978-84-9788-886-8.
8. Mg. Oporto Díaz, Samuel. Metodologías para el Data Warehousing.
9. Dario, Bernabeu Ricardo. *Hefesto Data Warehousing:Investigación y sistematización de conceptos*. Córdoba : s.n., 2010. V2.
10. Sanz, Miguel Rodríguez. *Análisis y diseño de un data mart para el seguimiento académico de alumnos en un entorno universitario*. Madrid : s.n., 2010.
11. UCID, Colectivo de autores del. *Metodología Híbrida para la construcción de un Data Warehouse*. La Habana : s.n., 2010.
12. Ing. Bernabeu, Ricardo Darío. ARQUITECTURA DEL DATA WAREHOUSING.
13. ETL-Tools.Info, Colectivo de autores de. ETL-Tools.Info. [En línea] [Citado el: 22 de 2 de 2011.] http://etl-tools.info/es/bi/almacenedatos_esquema-estrella.htm.
14. **Slideshare, Colectivo de autores de**. Slideshare. [En línea] [Citado el: 1 de 3 de 2011.] http://www.slideshare.net/aza_12/unidad-3-tsbd-olap.

15. **López, Francisco Javier Molier.** *Grupo A y B Informatica :Bloque específico.* Valencia : s.n., 2007. Vol. V2. 84-665-1829-0.
16. **Velasco, Roberto Hernando.** *El SGBDR Oracle .*
17. **Pecos, Daniel.** *PostGreSQL vs. MySQL..*
18. Herramientas y suites de IE. [En línea] [Citado el: 1 de 2 de 2011.]
http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/gonzalez_m_ac/capitulo3.pdf.
19. González Morales. Inteligencia empresarial para la toma de decisiones en la PYME enfocada en la administración de la relación con el cliente (CRM) utilizando análisis de la canasta de compra (MBA). Escuela de Ingeniería y Ciencias, Universidad de las Américas Puebla. A. C. 2007. [En línea] [Citado el: 7 de 1 de 2011.]
20. Dario, B. R. (2007). *Investigacion y sitematizacion de conceptos .* Córdoba.
21. **Caralt, Jordi Conesa.** *Introducción a la Business Intelligence.* Primera edición en lengua castellana. Barcelona : UOC1, 2010. 978-84-9788-886-8.
22. Herramientas Case. [En línea] [Citado el: 27 de 2 de 2011.] <http://www.scribd.com/doc/3062020/Capitulo-I-HERRAMIENTAS-CASE>.
23. Enterprise Architect- Herramienta de diseño UML.[En línea] 01-2010
<http://www.sparxsystems.com.ar/products/ea.html>.
24. autores, C. d. (2007). *JBoss Enterprise Middleware.*
25. Tomcat: Introducción al Servidor de Aplicaciones. [En línea] 01-2010
<http://www.consultec.es/formacion/temarios/TEMCON-307.pdf>.
26. Gutiérrez Kafati, Elizabeth. La plataforma Pentaho Open Source Business Intelligence. [En línea] 01-2010.
http://egkafati.bligoo.com/content/view/219538/La_plataforma_Pentaho_Open_Source_Business_Intelligence.html.

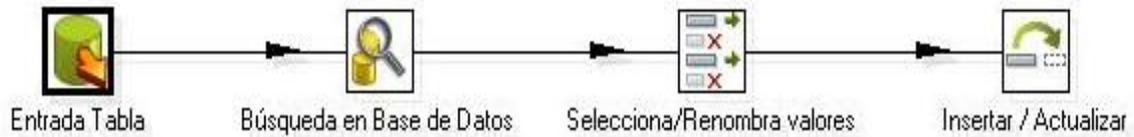
Anexos

Anexo 1 Base de datos del data mart, estructura Lógica

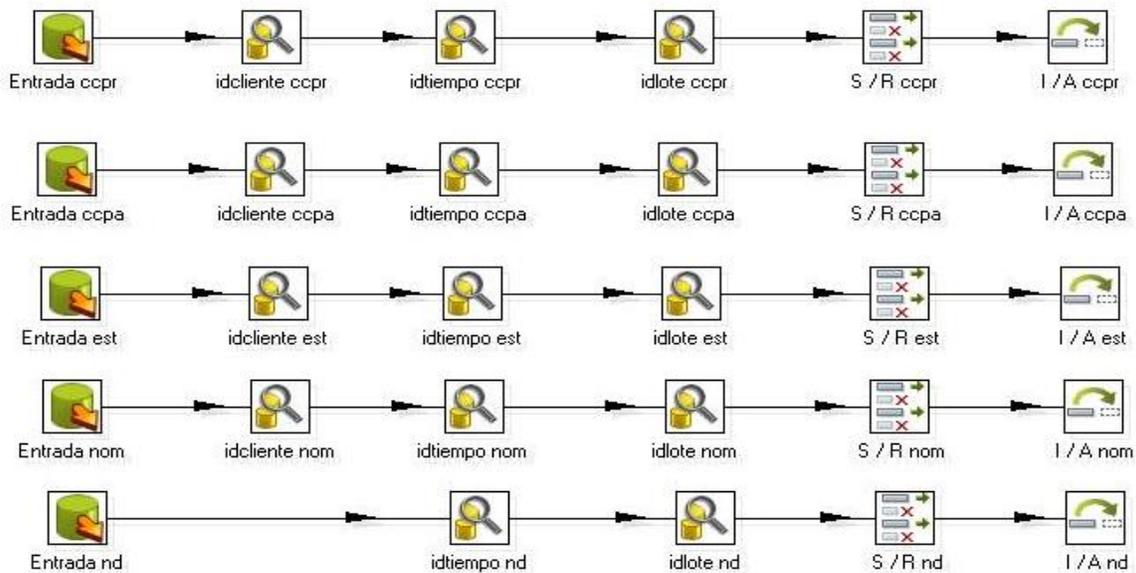


Anexo 2 Carga de los cubos del data mart

Anexo 2.1 Cubo de vencimiento de lotes



Anexo 2.2 Cubo movimientos



Anexo 2.3 Cubo de productos



Anexo 3 Carta de validación al data mart emitida por QUIMEFA

