

**Título: Data Mart de Gestión Económico - Financiera de la Empresa de Proyectos de Arquitectura e Ingeniería (EMPAI).**

**Autoras: Ing. Gisel González Hidalgo-Gato.  
MSc Emma Rizo Rizo.**

## **Introducción**

La EMPAI, como empresa que aplica el perfeccionamiento empresarial desde 1999, que se mantiene en constante desarrollo y cuenta con un grupo de herramientas y servicios informáticos ya implementados: sistemas de gestión de bases datos, intranet, Internet, correo electrónico, y otros, decidió comenzar a desarrollar e implementar herramientas como los Data Warehouse, que le permitirán conocer y evaluar la visión histórica de las variables y el análisis de los datos en el entorno, almacenando los datos a través de largos períodos de tiempo y realizando análisis multidimensionales de los mismos.

El presente trabajo tiene como objetivo el diseño del Data Mart para la gestión económico - financiera de la empresa. El cual servirá de apoyo en la toma de decisiones de la dirección de gestión económica y formará parte del Datawarehouse empresarial de la entidad.

## **Desarrollo**

### **Introducción al concepto de Data Warehouse (DW)**

El datawarehousing es el proceso continuo e iterativo de diseño, construcción, gestión y uso del Data Warehouse. El mismo soporta el procesamiento informático al proveer una plataforma sólida, a partir de los datos históricos para hacer el análisis. Facilita la integración de sistemas de aplicación no integrados. Organiza y almacena los datos que se necesitan para el procesamiento analítico-informático sobre una amplia perspectiva de tiempo. Y está dirigido a facilitar la disponibilidad oportuna de la información adecuada para que sea utilizada por el personal adecuado de la organización.

A principios de los años noventa, William H. Inmon, definió el término Data Warehouse (DW), cuya traducción es almacén de datos:

*“Un Data Warehouse o almacén de datos, es una colección de datos orientados a temas, integrados, variables en el tiempo y no volátiles, que se utilizan como apoyo al proceso de toma de decisiones”.[1]*

Se puede caracterizar un Data Warehouse haciendo un contraste de cómo los datos de un negocio almacenados en un DW, difieren de los datos operacionales usados por las aplicaciones de producción o gestión.

<b>Base de Datos Operacional</b>	<b>Data Warehouse</b>
Datos Operacionales	Datos del negocio para Información
Orientado a la aplicación	Orientado al sujeto
Actual	Actual + histórico
Detallada	Detallada + más resumida
Cambia continuamente	Estable

### **Características de los almacenes de datos.**

- **Orientados a temas:** Los datos se organizan y resumen según las actividades que son de interés para la entidad, por ejemplo: la contabilidad, las ventas, la producción, cada una de las cuales se desarrolla alrededor de sujetos tales como clientes, productos, compradores, etc.
- **Integrados:** Los datos se almacenan en un formato único y consistente, independientemente de la fuente de procedencia: bases de datos transaccionales, hojas de cálculo, documentos de texto, etc. Esto se muestra de muchas maneras: en convenciones de nombres consistentes, en la medida uniforme de variables, en fuentes múltiples y otros.
- **Variables en el tiempo:** Los datos relativos a un período de tiempo se incrementan periódicamente, ya que la toma de decisiones se apoya en diferentes modelos que necesitan información histórica.

- No volátiles: A diferencia de los datos de los sistemas operacionales que cambian momento a momento, los datos del DW no se modifican ni se actualizan sólo se añaden nuevos. Esto garantiza la disponibilidad de los datos históricos y permite la optimización del acceso a los mismos.

### **Estructura básica de la arquitectura del DW.**

Según [2] como se muestra en la figura No.1.1 la estructura básica del DW está compuesta por varios los elementos y procesos indispensables en el diseño del mismo.

El primer elemento son las fuentes de datos: que son las bases de datos transaccionales (BD) y las fuentes externas. Estos datos al extraerse pasan por un proceso de limpieza, transformación y carga hacia del DW (ETL, uno de los procesos básicos de un Data Warehouse) y el mismo consiste en:

**Extracción:** este es el primer paso para obtener la información hacia el ambiente del Data Warehouse.

**Transformación:** una vez que la información es extraída hacia el área de tráfico de datos, pasa por procesos de transformación y limpieza, seleccionándose únicamente los campos necesarios para el Data Warehouse. Estas transformaciones son necesarias para asegurar la calidad de los datos que serán almacenados en el DW e incluyen, entre otras, la corrección de errores, la eliminación de redundancia y la resolución de inconsistencias, problemas que se presentan comúnmente cuando diversas bases de datos convergen en un sistema único.

**Carga:** al final del proceso de transformación, los datos están en forma para ser cargados. [3]

Otro elemento importante del DW lo constituyen **los metadatos**, que no son más que "datos sobre los datos". En la medida en que aumenta la variedad y cantidad

de datos se hace más necesario una descripción de las características de los mismos. Existen distintos tipos:

- Metadatos Administrativo (toda la información necesaria para el DW).
- Metadatos del negocio. (Información y términos del negocio, política de seguridad, etc.)
- Metadatos Operacional (Obtiene información recogida durante el proceso de almacenamiento de los datos).

Los **Data Marts** (DM), no son más que DW con funciones departamentales o regionales. Tienen las mismas características que los DW y brindan las mismas facilidades; pero se definen para satisfacer las necesidades de un departamento o sección dentro de una empresa [4]. Normalmente, en la práctica, suelen contener más cantidad de información agrupada que en detalle.

Existen dos tipos de DM [5]:

- Independientes: los datos son extraídos de los sistemas operacionales.
- Dependientes: los datos son extraídos del DW.

Para su construcción se puede seguir dos aproximaciones: [2]

(I) Definir primero el almacén de datos y, a partir de él, definir los data marts, ó

(II) Definir primero los data marts departamentales y, posteriormente integrarlos en un almacén de datos global para la organización.

Y por último encontramos las herramientas de Acceso al componente de almacenamiento físico del DW: son las que permiten a los usuarios el obtener la información. Estas pueden ser herramientas específicas de mercado para visualización de bases multidimensionales almacenadas en Data Warehouse como también aplicaciones desarrolladas dentro de la organización del tipo EIS/DSS. Existen tres categorías en las que se pueden clasificar estas herramientas: las consultas o informes simples, los cubos OLAP (procesamiento analítico en línea) y la minería de datos.

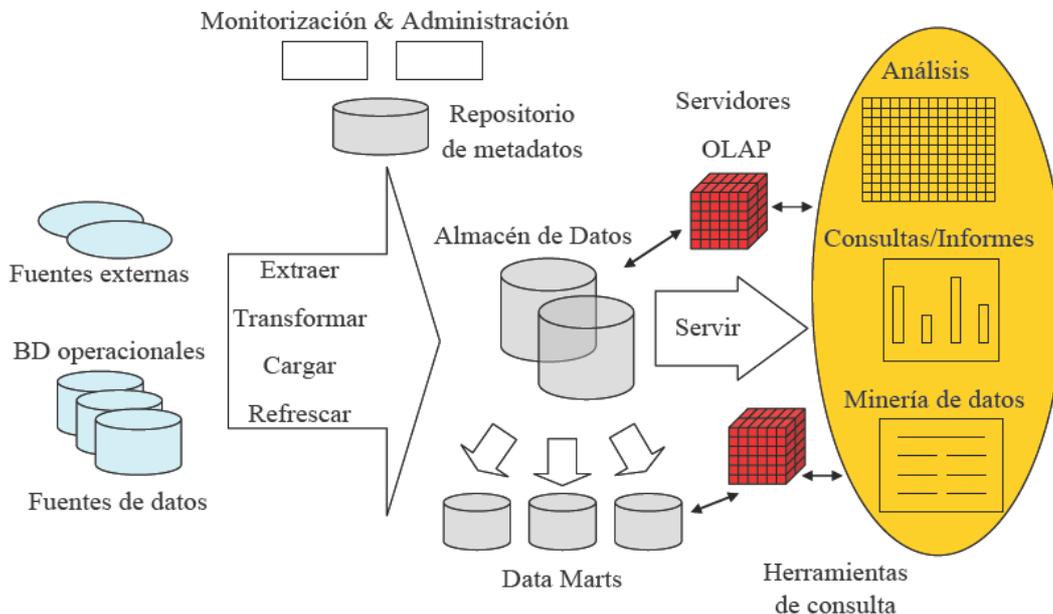


Figura No.1.1 Arquitectura básica del Data Warehouse.

### Diseñando el Data Mart.

Para realizar el diseño del DW corporativo de la EMPAI[6], se realizó un estudio del funcionamiento de la empresa, sus objetivos y líneas de desarrollo, estructura funcional y las áreas claves de la misma. A partir de ahí y teniendo en cuenta los criterios e información recopilada en las encuestas realizadas a los expertos y directivos se seleccionaron las áreas que podían aportar información útil para la toma de decisiones en la entidad y se definió el Esquema General del DW corporativo, el cual podemos observar en la figura No.1.2.

De los dos enfoques básicos que existen para la construcción de un DW: **Top-Down** y **Bottom-Up**. Se seleccionó el segundo para diseñar el DW Corporativo, construyendo una serie de DM incrementales de cada una de las direcciones señaladas en el Esquema General, cada uno de los cuales finalmente se integrarán en el DW Corporativo. Comenzándose el diseño de los DM en aquellas direcciones donde la información estaba organizada en bases de datos

transaccionales[6]. En el caso específico de nuestra investigación, trabajamos en el rediseño e implementación del Data Mart de Gestión Económica.

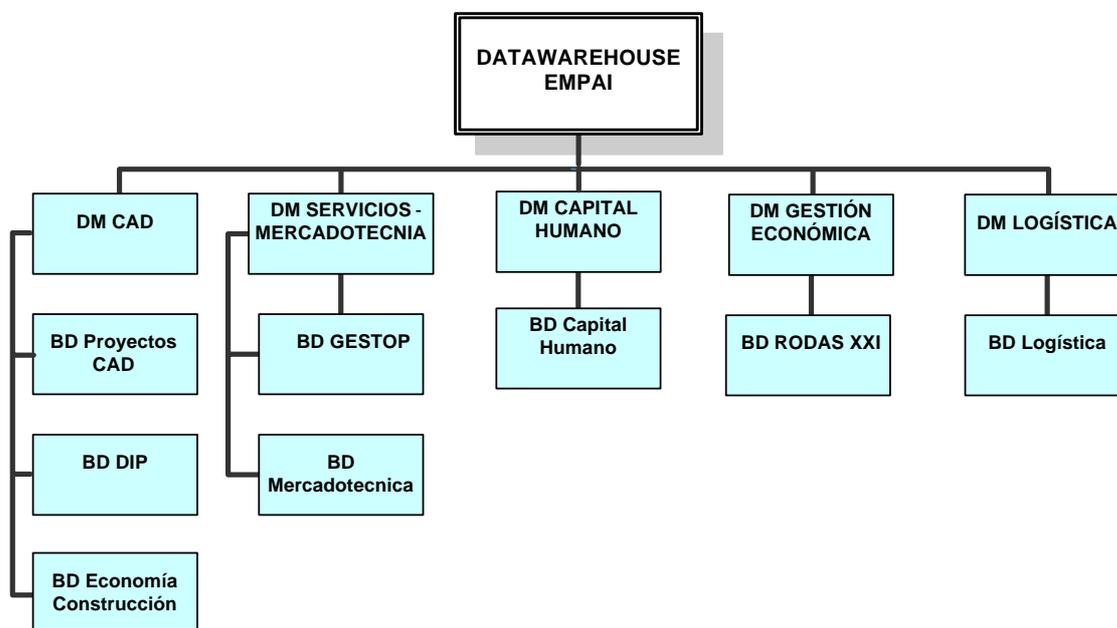


Figura No. 1.2: Esquema General del Datawarehouse Corporativo.

### Captura de Requisitos.

Según la metodología utilizada para el diseño e implementación del Data Mart una aproximación de ingeniería de requisitos debe facilitar la especificación de objetivos estratégicos, a partir de los cuales se obtengan las necesidades de información relacionados con medidas de interés para los procesos de negocio (contenidos en hechos) y el contexto para el análisis de estas medidas (dimensiones).[6]

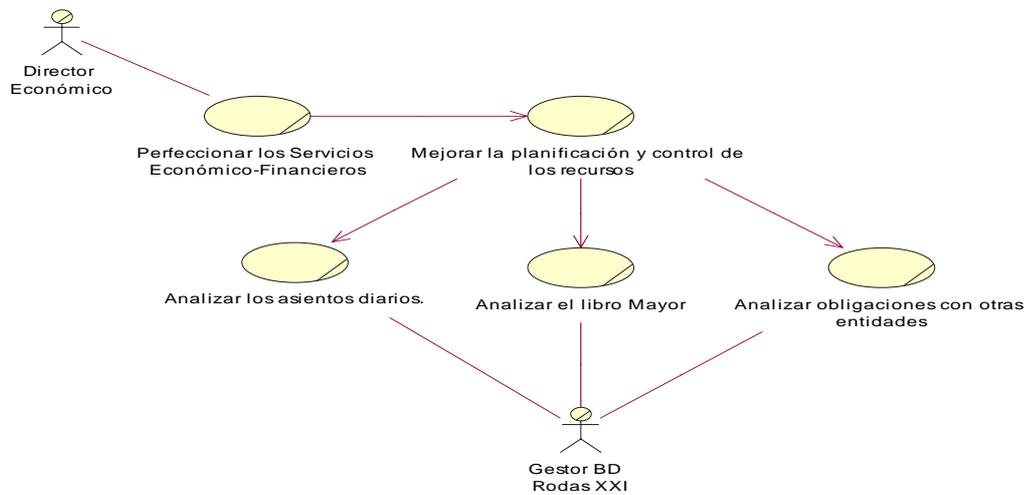
De ahí que se realizaran entrevistas a los usuarios finales para conocer los objetivos estratégicos, necesidades de información, frecuencia de actualización de los datos, etc. Una vez identificada toda la información se elabora el Diagrama de Casos de Uso del negocio (BUC), el cual define un proceso de negocio y representa visualmente los objetivos organizacionales, decisionales e informacionales del mismo, mejorando su comprensibilidad y permitiendo a los desarrolladores abstraer los objetivos de su realización.

Para el diseño del Diagrama de Casos de Uso del negocio (BUC) se identificaron los actores involucrados en la toma de decisiones del proceso, los cuales son: el “Director Económico” y el “Gestor de BD RODAS XXI”.

Por entrevistas realizadas a los altos ejecutivos se definió que el Objetivo Organizacional (OO) es “Perfeccionar los servicios Económico – Financieros”. Refinando el OO: ¿cómo perfeccionar los servicios Económico – Financieros?, se obtuvo el Objetivo Decisional (OD): “Mejorar la planificación y control de los recursos económicos y financieros”. Y luego refinando el OD: ¿qué hacer para mejorar la planificación y control de los recursos económicos y financieros?, se obtiene que el Objetivo Informativo (OI) es “Analizar los asientos, el Libro Mayor General y las obligaciones con otras entidades”. Asociando cada actor con sus objetivos estratégicos se obtiene el Diagrama de Casos de Uso del Negocio. (BUC). Ver figura No. 1.3.

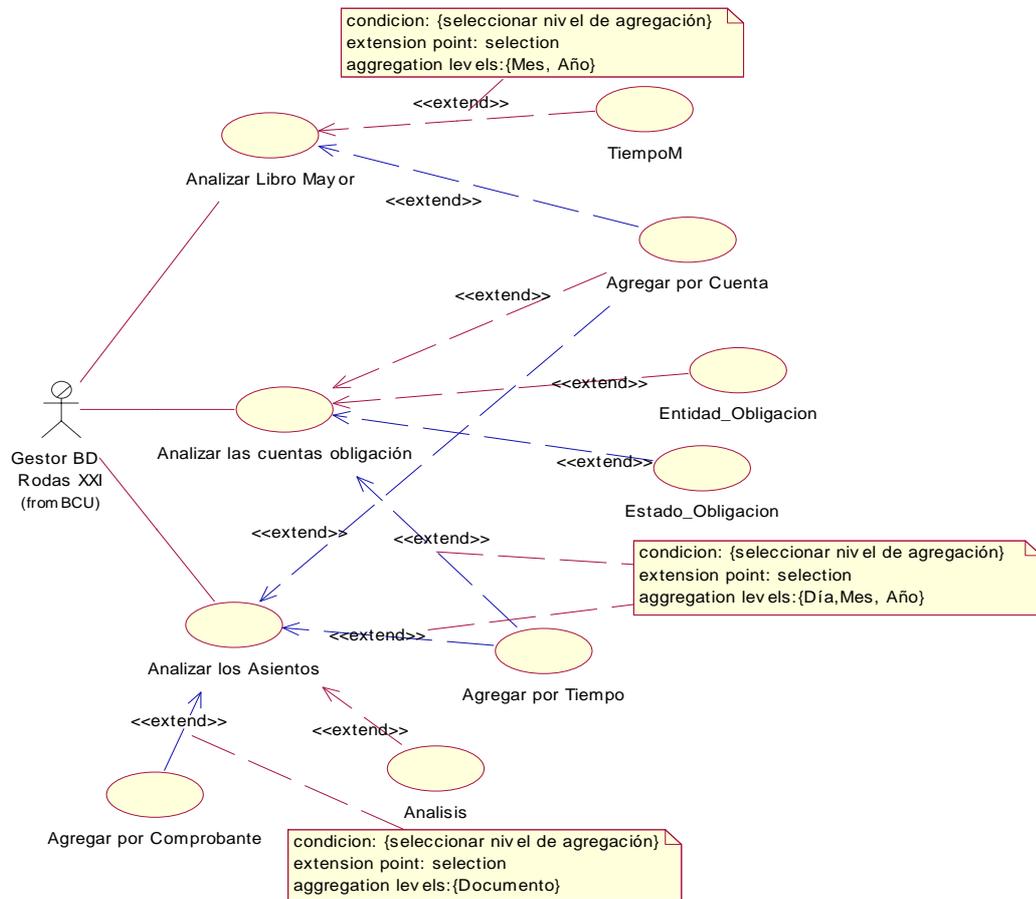
Finalmente, partiendo de todo este análisis se obtienen como requisitos de información que: Para mejorar la planificación y control de los recursos económicos y financieros se deben analizar los asientos, el libro mayor y las obligaciones en cada período contable y en los contextos de las cuentas, los análisis, los estados de las obligaciones, las entidades que contraen las obligaciones y los comprobantes.

Una vez definidos y representados los objetivos organizacionales en el Diagrama de BUC e identificados los requisitos de información necesarios para lograr los objetivos, estos últimos son representados en Diagramas de Casos de Usos (UC), donde se especifica más sobre los mismos.



**Figura No. 1.3: Diagrama de Casos de Uso del negocio (BUC).**

En la figura No 1.4 aparecen representados los Casos de Uso del sistema, en el mismo el “gestor de BD RODAS XXI”, actor del diagrama BUC, se transforma en el actor del diagrama UC. Las actividades del actor del diagrama BUC: “analizar los asientos, analizar el libro mayor y analizar las obligaciones”, se transforman en casos de uso del diagrama UC y se asocian con el actor del mismo. Estos casos de uso representan el análisis de medidas de hechos dentro de un contexto determinado: las cuentas, los análisis, los estados de las obligaciones, las entidades que contraen las obligaciones, los comprobantes y el período contable. Este contexto de análisis se representa por UC extendidos que cumplen la función de agregar medidas según las diferentes dimensiones. Los puntos de extensión representan niveles de agregación y están representados con notas.



**Figura No. 1.4: Diagrama de Casos de Uso del Sistema.**

## **Análisis**

La Dirección de Gestión Económica Financiera es la encargada de gestionar, analizar, planificar, asesorar y fiscalizar los recursos económico-financieros de la entidad, constituyendo un instrumento fundamental de la Dirección General para la toma de decisiones.

Para realizar su trabajo los especialistas del área económica utilizan como herramienta principal el sistema relacional de bases de datos RODAS XXI, diseñado por CITMATEL e implantado en la entidad desde hace siete años.

El RODAS XXI es un sistema programado con Visual Studio y utiliza Microsoft Access 97 como gestor de Bases de Datos en su versión 2.7 y Microsoft SQL 2000 en su versión 3.0. El mismo consta de cuatro módulos fundamentales: Contabilidad General, Medios Inmovilizados, Medios de Rotación y Nómina, con los cuales se realizan las operaciones diarias del área.

Luego de un estudio y análisis del sistema y conociendo las necesidades del usuario, se decidió tomar como fuente para el diseño del Data Mart de la gestión económico – financiera las bases de datos del Sistema General de Contabilidad, ya que el mismo resume y consolida todas las actividades que realizan el resto de los módulos. Para realizar el análisis se elaboró el diagrama de la Ingeniería Inversa de la Base de Datos de Contabilidad del Sistema RODAS XXI, donde aparecen las relaciones entre las tablas que componen el sistema. Ver anexo No.1.

## **Diseño**

Para saber si los requisitos capturados cumplen con las necesidades del usuario, se especifica un esquema Multidimensional (MD) conceptual, usando el profile UML[7]. El mismo se define a partir de los diagramas de UC al identificar las clases dimensión, clases base (jerarquías de clasificación) y clases hecho. Además se reutilizan los requisitos para descubrir dimensiones y hechos comunes. [8]

Según los conceptos planteados los UC principales del Diagrama de Casos de Uso se transforman en clases de hechos, por tanto tenemos que: “Analizar los Asientos” se transforma en la clase de hecho “Asiento”, “ Analizar las cuentas obligación” se transforma en la clase de hecho “Obligaciones” y “Analizar libro Mayor” se transforma en la clase de hecho “Mayor”.

Vemos también que más de un UC comparten extensiones similares: “Agregar por Tiempo” y “Agregar por Cuenta”, lo que indica que existen dos dimensiones comunes que son “Tiempo” y “Cuenta”. Las otras dimensiones que se obtienen son: “Comprobante”, “Análisis”, “Estado\_Obligacion”, “Entidad\_Obligación” y “TiempoM”.

Además por cada nota asociada a una relación de extensión se obtiene una jerarquía de clasificación de la dimensión y cada nivel de agregación se considera un nivel de jerarquía que se representa como una clase base. En nuestro caso existen tres dimensiones que tienen jerarquías de clasificación y son: la dimensión “Comprobante” con la clase base “Documento” de un solo nivel de jerarquía, la dimensión “Tiempo” con las clases base “Dia”, “Mes”, “Anyo” y la dimensión “TiempoM” con las clases base “Mes”, “Anyo”. Ver anexo No. 2 donde aparece representado el diagrama de clases multidimensional (MD), en el cual se observan las relaciones de las clases de hechos con cada una de las dimensiones y puede comprobar que analizar los asientos, el libro mayor y las obligaciones en cada período contable y en los contextos de las cuentas, los análisis, los estados de las obligaciones, las entidades que contraen las obligaciones y los comprobantes, se cumple navegando por el diagrama.

Como resultado de la etapa de diseño se obtiene además el diagrama conceptual, que se compone por tres niveles y del cual presentamos sólo una vista en el anexo No.3, donde se observa la tabla de hechos “Asiento” con sus respectivas clases de dimensiones: “Cuenta”, “Comprobante” y “Tiempo”, así como las clases bases. Para la elaboración de los diagramas se utilizó Rational Rose Interprise Edition y el Método global basado en UML para el diseño de Almacenes de Datos.

### **Implementación**

Actualmente se trabaja en la fase de implementación de los procesos de limpieza, transformación y carga de los datos hacia el Data Mart, específicamente en la carga incremental del mismo. Para ello hemos utilizado el Microsoft SQL Server 2000.

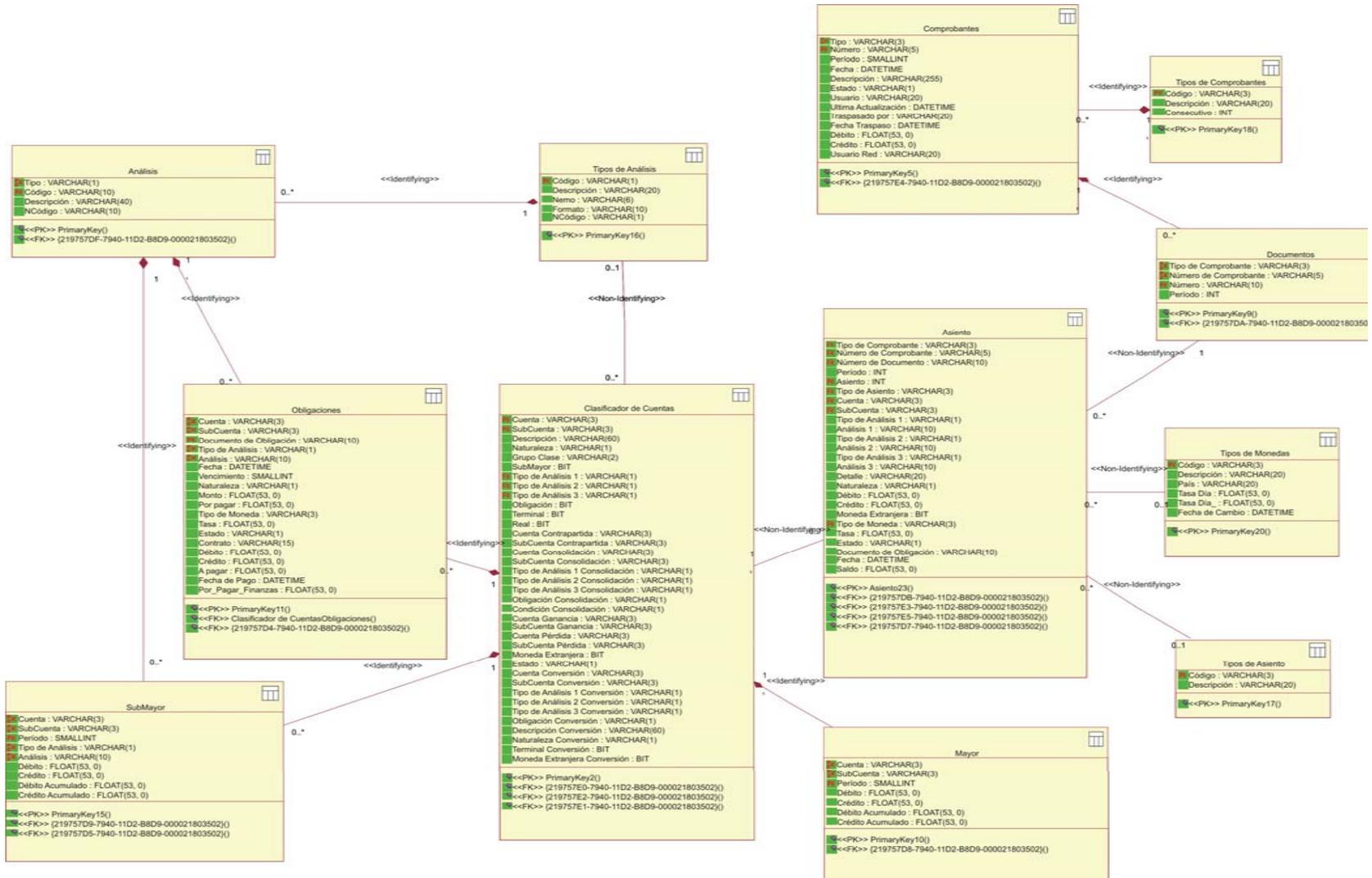
### **Trabajos futuros.**

Integrar este Data Mart al Datawarehouse Corporativo de la empresa.

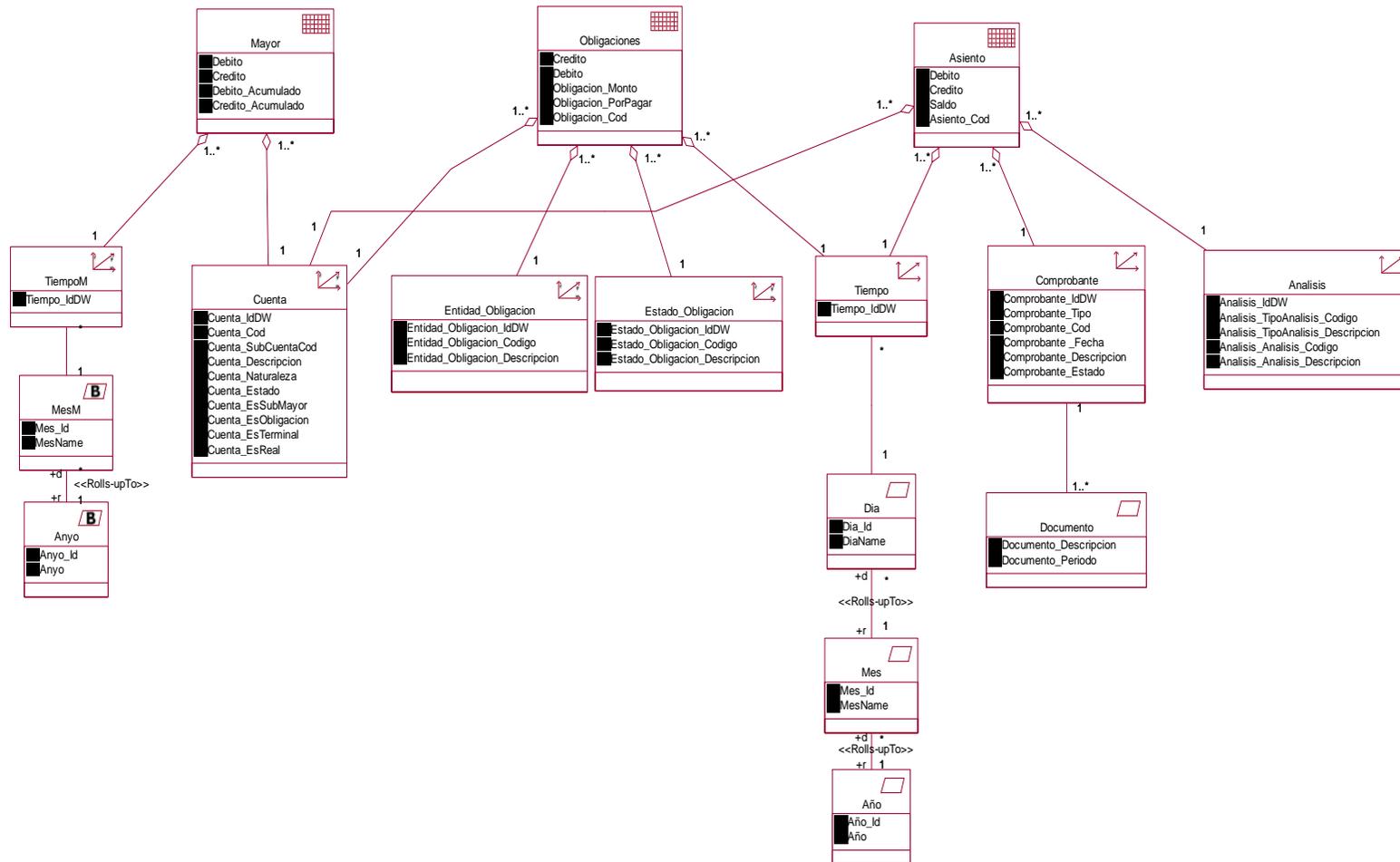
### **Bibliografía**

1. W.H., I., *Building the Data Warehouse*. 1992: QED Press/John Wiley.
2. Trujillo, J.C. *Arquitectura de un sistema de Almacenes de datos. Tema 2. in Diseño y Explotación de Almacenes de Datos. Curso Doctorado 2006. Cuba.* 2006: Departamento de Lenguajes Sistemas Informáticos de la Universidad de Alicante.
3. D.G., E., *Data Warehouse*. . 1997.
4. Velásquez Vidal, L.V.M., Mijail. , *Estudio Teórico-Conceptual sobre Data Warehouse*. 2000., Universidad de La Habana.
5. García, R.R., *Data Mart para la actividad de Ventas en la Comercializadora Cuba Ron.*, in *Facultad de Informática*. 2007, Universidad "Camilo Cienfuegos" Matanzas. p. 89.
6. Mora T.M., M.D.L., *Diseño de un Data Warehouse para la EMPAI.*, in *Facultad de Informática*. 2006, Universidad "Camilo Cienfuegos" de Matanzas, Cuba.: Matanzas. p. 104.
7. Luján-Mora S, T.J., Song I-Y, *Multidimensional modeling with UML package diagrams*. Vol. 2503. 2002.
8. Paim F.R.S., C.J.B., *DWARF: approach for requirements definition and management of data warehouse systems*. . 2003.

# Anexo No.1: Diagrama de la Ingeniería Inversa de la Base de Datos de Contabilidad del RODASXXI.



## Anexo No.2: Diagrama de clases multidimensional del Data Mart.



### Anexo No.3: Vista del Diagrama Conceptual del Data Mart.

