

El hecho de dividir en 2 partes fundamentales la creación de una Fábrica de información ayuda a mejorar el tiempo del proyecto ya que permite un desarrollo en paralelo en el que una fase de Front end no depende de la de Back End.

El formato para el proyecto incluye los siguientes conceptos: Actividad, Duración, Tiempo de inicio, Tiempo de finalización, Recursos. El formato a utilizar se encuentra en el anexo A1.

### **3.3.1.2.2 Obtención de requerimientos**

Esta es una subfase del análisis y consiste en realizar reuniones con la gente de las áreas involucradas para obtener realmente sus necesidades de información.

Para la obtención de los requerimientos se realizaron 3 reuniones recopilando los datos referentes a la información que proporciona un valor agregado para la compañía. Toda la información recopilada se organizó el documento presentado en el anexo A2 el cual incluye las siguientes características:

- Información del proyecto (Nombre, Responsables, Áreas funcionales involucradas, etcétera)
- Información de los requerimientos (Que indicadores de negocio son necesarios en la iteración, cuales son sus dimensiones, cual es su impacto en el negocio, formas de acceso)
- Información de necesidades de información (Cuanta historia de datos requiere)
- Información del documento en sí (Fechas de creación, modificación, revisión, creador)

### **3.3.1.2.3 Diseño de modelo general de la FIC propuesto**

Basado en los requerimientos obtenidos, es en esta fase donde se define la arquitectura óptima para poder satisfacer éstos. Este modelo es donde se identifican que partes de la fábrica de información corporativa se emplearán, modelo que será la base fundamental para el desarrollo del proyecto: Este modelo se plasma en un diagrama indicador general con los componentes necesarios para GAMESA.

### **3.3.1.2.4 Construcción de matriz de bus**

Una vez entendidas todas las partes involucradas en la fábrica de información, y que además hemos entendido todas los requerimientos, es necesario crear una matriz de bus, que funcione como base fundamental de indicador para definir las dimensiones posibles y que cuando sea requerido un incremento de indicadores de otras áreas de negocio y organizar nuevo Data Marts, sea fácilmente identificable las dimensiones en las cuales puede realizar el usuario el empate de conceptos. La matriz de bus se encuentra en el anexo A3.

Los conceptos que incluye el documento son:

- DataMarts
- Dimensiones
- Breve descripción de los Data Marts y las dimensiones
- Fechas de creacion, modificación, revisión y autores.

### 3.3.1.2.5 Realización del modelado Dimensional

Como punto crucial en el diseño y creación de la FIC es necesario utilizar los siguientes pasos para realizar una correcta modelación.

- **Utilización de método de los cuatro pasos para diseñar cada tabla de hechos**
  - 1) ***Se selecciona el Data Mart o los indicadores a modelar:*** Verificando esto a partir de nuestra matriz de bus. Es muy importante que si el proyecto inicia con varias áreas de negocio, se inicie con la más sencilla que equivaldría a seleccionar la que menos fuentes de datos tenga.
  - 2) ***Declaración del Grano o nivel de detalle:*** Esta parte consiste en que basados en los requerimientos de información lleguemos a determinar el mínimo nivel de detalle que encontraremos en las tablas de hechos. Un punto importante para entender la granularidad consiste en determinar cuál es el nivel de explotación de información necesario para tomar una decisión.
  - 3) ***Selección de las dimensiones:*** El punto anterior es muy importante ya que dependiendo del nivel de granularidad seleccionado se debe de visualizar que dimensiones pueden satisfacer nuestros requerimientos por lo que una vez visualizado el nivel y revisado la matriz de bus, podemos o cambiar el nivel de detalle o eliminar esta dimensión del indicador. Una vez determinadas las dimensiones es necesario analizar los diferentes atributos descriptivos de esta dimensión, el objetivo es ser comprensivo y detallado.
  - 4) ***Selección de la tabla de hechos:*** Este paso consiste en agregar tantos hechos como sean posibles para un nivel de granularidad especificado.

En esta subfase del diseño se realiza el modelado de todas las dimensiones y hechos, así como sus características. El objetivo principal de

esta actividad es tener todo el modelo tanto de ODS como de DW listo para ser transmitido a la herramienta modeladora seleccionada y poder crear el modelo de Base de datos apropiado.

- **Diagrama y detalle de tabla de hechos**

En esta parte se crea un diagrama para cada una de las tablas de hechos que se vayan a utilizar. El diagrama de tablas de hechos no solo muestra las dimensiones que le corresponden, sino que se presenta sobre el total del proyecto. Este diagrama muestra los nombres de las tablas de hechos, la granularidad y muestra todas las dimensiones tanto conectadas como no conectadas.

Para el detalle de la tabla de hechos es necesario crear una lista de todos los hechos disponibles a través de la tabla de hechos incluyendo los hechos físicos, los hechos derivados y aquellos que se calculen con la combinación de ambos. Además se deben incluir reglas de agregación para cada hecho.

Se anexa el formato utilizado para plasmar el detalle de la tabla de hechos en el anexo A4.

- **Diagrama y detalle de las dimensiones**

En esta parte se crea un diagrama para cada una de las dimensiones que se vayan a utilizar. El diagrama de dimensiones muestra los atributos individuales dentro de una sola dimensión. Cada diagrama muestra el grano explícito de cada dimensión, además de que presenta la posible cardinalidad de cada dimensión. Las relaciones entre los atributos definen la trayectoria de navegación (drills).

Para el detalle de la tabla de dimensiones se propone el documento el cual se encuentra en el anexo A4 que en complemento con la modelación de los hechos nos proporciona el modelo final.

**Los campos a llenar son:**

- **Nombre del atributo**
- **Descripción del atributo**
- **Cardinalidad (El mejor estimado de los posibles valores distintos que puede tener este atributo)**
- **Política de Cambio de dimensiones con sus cuatro tipos de valores:**
  - 1) **Tipo 0 o que el valor nunca se sobrescribe.**
  - 2) **Tipo 1 o Sobreescribir**
  - 3) **Tipo 2 o creación de un registro nuevo por cada valor**
  - 4) **Tipo 3 o solo se mantiene el valor inmediato anterior y el actual.**
- **Datos de ejemplo**

- **Hoja de hechos derivados**

Usualmente no todos los campos de la tabla de hechos son un indicador puro como tal, es decir, es necesario realizar alguna operación con ellos para llegar a un resultado final. Un ejemplo de esto es el acumulado total de venta de lo que va del año.

Para poder llevar a cabo un entendimiento de todos los cálculos de los indicadores necesarios para solventar los requerimientos de información se utiliza un formato que servirá de base para la explotación de la información. Este formato se encuentra en el anexo A5.

**Incluye las siguientes columnas:**

- **Bandera de cambio:** Útil cuando existen demasiados cálculos de hechos y por consiguiente su localización es complicada. Este nos ayuda a que no sea así.
- **Data Mart(s):** Indica los Data Marts que están involucrados con los cálculos.
- **Nombre del Hecho**

- Descripción del Hecho

pos Tipo:

- Columna (Hecho directo de la tabla)
- Restriccional (Cuando es necesario aplicarle una condición)
- Transformación (Cuando se requiere que se realice el cálculo basado en la transformación de algún atributo de la dimensión)
- Cálculo (Un cálculo algebraico)
- Columna con Límites (Especificando cuando aplica este hecho).
- Regla de agregación: Determina el nivel default de agregación.
- Fórmula: Es la fórmula para llegar al cálculo final.
- Restricciones: Las cuales se deben de aplicar a los datos para llegar a un resultado final.
- Transformaciones: Indica las transformaciones específicas que aplican para crear este hecho.

Es importante que este documento sea llenado antes de empezar a desarrollar ya que muchas veces el hecho de estandarizar y obtener fórmulas de negocio puede ser muy complejo y algo frustrante.

### **3.3.1.2.6 Identificación de las fuentes de cada tabla de hechos y dimensiones**

Este paso que inicia previo en el análisis pero que corresponde al diseño se utiliza para llevar a cabo un estudio de las posibles fuentes de donde puede existir la información necesaria para extraer los datos y almacenarlos en la fábrica. Consiste en comprender cada una de las fuentes de datos que serán incluidas en la iteración del ODS y del DW.

**Aquí es importante identificar el tipo de fuente ya que puede ser de dos tipos:**

- 1) Fuente formal de datos:** Una fuente formal es aquella que ya forma parte de los procesos de Informática para soportar un área funcional y que es utilizado y alimentado basado en el día a día de la compañía. Los ERP "Enterprise Resource Planning" y los sistemas hechos en casa caen dentro de este concepto.
- 2) Fuente informal de datos:** Una fuente informal es aquella que no existe en ningún sistema informático, pero que es necesario que se incluya en el DW ya que forma parte de los requerimientos. Muchas veces es necesario construir algo para poder solventar esta alimentación de datos.

**Se anexa documento que nos ayudará a identificar y entender las fuentes de información en las iteraciones en el anexo A6.**

**Los campos que la incluyen son:**

- **Fuente:** El nombre del sistema origen
- **Área de negocio responsable:** Contacto responsable de los datos
- **Responsable de Informática:** Persona responsable de los sistemas
- **Plataforma:** El ambiente de sistema operativo en que corre.
- **Localidad:** La localidad física en que se encuentra el sistema
- **Descripción**

**Para poder seleccionar correctamente una fuente nos basamos en los siguientes criterios:**

- **Accesabilidad de los datos**
- **Que tan exacto es el dato**
- **Calendarización de proyectos**

### **3.3.1.2.7 Mapeo fuente a destinos**

Una vez identificada la fuente formal de explotación continúa esta actividad de diseño que hay que considerarla como la base para el proceso de alimentación de datos a la Fábrica de información.

Básicamente consiste en unir las tablas fuentes con sus correspondientes tablas destinos campo(s) por campo.

El documento para realizar esto se encuentra en el anexo A7.

Los campos a incluir son: Nombre de la tabla, nombre de la columna, tipo de dato, longitud, descripción de la columna destino, sistema origen, tabla o archivo origen, campo/columna (s) origen, transformación(Identifica a grandes rasgos las reglas de negocio de la empresa), dimensión o DM, atributo/hecho.

### **3.3.1.2.8 Desarrollo por medio de herramienta modeladora de modelo físico a un manejador de Base de datos**

Una vez analizados todos los requerimientos, definido la arquitectura conceptual de la iteración y finalmente haber desarrollado el modelo lógico de datos, el siguiente paso consiste en diseñar por medio de una herramienta modeladora de datos el modelo que posteriormente se transformará en la estructura física de la Base de Datos a utilizar.

Por medios de la herramienta Erwin se definirán estos modelos. Esta parte se encuentra en el diseño.



### 3.3.1.2.9 Dimensionamiento

Ya que se tiene el diseño del modelo de datos es necesario estimar los tamaños de los objetos a crear dentro de la Base de datos. Esta es una actividad que tiene un poco de arte ya que cuando es una iteración nueva, no es posible tener un exacto del espacio por lo que por medios intuitivos se estiman estos tamaños. Es importante comentar que el dimensionamiento varía dependiendo de los tamaños de los datos por lo que el formato a llenar para esta tesis es creado con las características generales del manejador de base de datos a utilizar, en este caso ORACLE.

La estrategia sobre la cual se llevó a cabo el dimensionamiento es la siguiente:

- Se plantea llenar un resumen de dimensionamiento que parte de un documento en el cual se recopilan los siguientes conceptos:

Concepto	Explicación
Tabla	Nombre de la tabla
Campo	Nombre de los campos
Tipo de dato	El tipo de datos dependiendo de la B.D.
Tamaño	Tamaño del campo
Escala	Número de decimales
Valor	Si puede ser nulo o no
Bytes	# de bytes dependiendo del tipo de dato: Numéricos= (#bytes/2) +1 Fechas = 7bytes Alfanuméricos = #bytes
Total bytes por registro	Suma de los bytes calculados de todos los campos
Número de registros por periodo	Cantidad de registros por periodo de tiempo
Cantidad de periodos en línea	Cantidad de periodos a almacenar
Total de registros históricos	Total de registros a almacenar
Total de espacio de datos	Espacio requerido para almacenar los registros.

Tabla 3.1 Equivalencias de dimensionamiento generales

- Se crea el resumen de dimensionamiento con las siguientes características:

Concepto	Explicación
Índice	Nombre del índice
Tabla	Nombre de la tabla
Espacio por registro	Espacio requerido por registro.
Cantidad de registros	Cantidad de registros a almacenar
Segmento Inicial (Initial Extent)	Segmento inicial característico de ORACLE para pre-alojar espacio necesario para satisfacer requerimiento de registros por tabla. Aquí se calcula basado en el total de espacio dando un 20% de más para soportar futuros crecimientos.
Siguiente Segmento (Next Extent)	Espacio para apertura de segmentos posteriores propio de ORACLE que se calcula como un 5% del Segmento inicial.
Tablespace	Concepto propio de ORACLE en donde se alojan las tablas e índices.

Tabla 3.2 Resumen de dimensionamiento

- Finalmente basado en estas características se crean los objetos de tabla e índice.

El machote del formato se encuentra en el anexo A8.

### 3.3.1.2.10 Plan y Diagrama de arquitectura tecnológica y de infraestructura.

Una vez entendidos todos los requerimientos el último paso consiste en crear el plan tecnológico de la arquitectura y su correspondiente diagrama de la infraestructura a utilizar. Esta es la traducción técnica de los requerimientos de negocio y contiene las capacidades que se busca se cumplan para satisfacer las necesidades del mismo.

Los puntos que se considerarán son:

- **Entendimiento de los requerimientos del negocio**
  - ~ **Cuestiones de negocio**
  - **Acceso de información**
  - **Navegación**
  - **Calidad de datos**
  - **Infraestructura y utilerías**
  - ↯ **Cuestiones organizacionales**
  
- **Una revisión de la arquitectura**
- **Elementos de la arquitectura**

Una vez definido todo el modelo el último paso es basado en la arquitectura tecnológica a utilizar, es necesario realizar el bosquejo de éste para su consiguiente implementación.

### **3.3.2 Mejoramiento en procesos de carga**

Este mejoramiento se visualiza una vez que se ha dado la utilización de la metodología de desarrollo y se han llenado los correspondientes formatos para facilitar éste, pero uno de los puntos más importantes es que se realizará el cambio de herramienta por medio de la cual se desarrolló originalmente el DW (Extracción pura con programas difíciles de administrar) por la nueva herramienta de ETL la cual además de proporcionar todas las bondades que una herramienta de este tipo puede dar, su forma de trabajar ayudará a formar un mapa de configuración óptimo el cual permitirá que los procesos sean más rápidos y por consiguiente las cargas.

El hecho de utilizar una herramienta de este tipo nos hace pensar de otra manera identificando los puntos a mejorar en los procesos de carga en tres tipos:

- 1) Origen
- 2) Destinos
- 3) Proceso de transformación

Todo proceso de carga realizado para la alimentación de la fábrica de información se basó en la documentación obtenida en la parte de análisis y diseño, pero lo más importante es que se desarrolló pensando en los diferentes conocimientos y técnicas a considerar en la optimización.

Dentro del conocimiento debemos de tener en mente:

- La arquitectura de la herramienta ETL
- El conocimiento de las características de la base de datos y su arquitectura
- Características de los sistemas operativos y su arquitectura

Dentro de las técnicas tenemos:

- El saber como encontrar cuellos de botella.
- El saber como eliminarlos.

Los factores óptimos para el mejoramiento de desempeño obtenido directamente de las sugerencias de la propia herramienta ETL [Infoopt, 99] son:

- La simplificación
- Obtener datos a través de la herramienta rápidamente

Esto implica la optimización a nivel sesión(características a nivel sistema operativo y de base de datos) y mapeo (todas las características de los objetos

internos de transformación); además implica una optimización del sistema para aseguramos que la herramienta esté configurada correctamente.

- **Introducir y extraer datos de la herramienta rápidamente**

Esto implica usualmente optimizar los componentes externos de la herramienta (base de datos, sistema operativo, red). Otro punto a tener en la mente es el hecho de que la herramienta no puede correr más rápido que las fuentes y los destinos.

Para medir el cumplimiento de este objetivo se realizarán recopilaciones de tiempos de procesos actuales contra los anteriores. Basado en esto podremos definir el grado de cumplimiento.

Dentro de este punto una de las fases que se manejan es el de la construcción, el objetivo de esta tesis no es enfocarse a la mejora manera de construir ya que este depende de las herramientas a utilizar. Pero si propone una estrategia de carga en un documento denominado mapa de configuración el cual ayuda a establecer la forma en que los procesos serán organizados y por medio del cual se distribuirán las cargas y se optimizará los tiempos de procesamiento.

El formato de mapa de configuración se presenta en el anexo A9.

Además todos los desarrollos estarán basados en los siguientes puntos de optimización:

- La herramienta tiene la capacidad de trabajar en memoria por lo que es conveniente utilizarla al máximo.
- La explotación de bases de datos es importante. Muchas veces no es posible optimizar las consultas a los sistemas transaccionales por lo que se utiliza un concepto de divide y vencerás, es decir, trae los datos a un área

de procesamiento controlable (donde se puedan crear índices, uniones más rápidas entre tablas, etcétera). Este es el concepto de "Área de procesamiento".

- Una organización por pasos es la mejor opción para optimizar las cargas.
- Todas las búsquedas de llaves se trabajan por medio de objetos denominados "lookups" los cuales permiten cargar las dimensiones y que sus accesos sean más rápidos por esta misma característica.
- Siempre se busca en cada uno de los objetos de transformación que las operaciones tanto de transformación como aritméticas sean realizadas en la menor cantidad de registros por lo que aún y cuando suene muy lógico es preferible trabajar en una secuencia que suelo denominar FAP(Filtro, realizo actividades de agregación y proceso).

### **3.3.3 Mejoramiento en Validación**

Una de las deficiencias que se tenía con el antiguo modelo existente en la compañía GAMEESA era que el proceso de validación se encontraba a nivel interno de interfases por lo que únicamente era posible atacar errores lógicos de información los cuales eran detectables al siguiente día del proceso y por las personas que conocían de estas interfases, pero lamentablemente estas validaciones no permitían recuperaciones rápidas ni cuando menos una visualización por parte de los operadores encargados de los mantenimientos nocturnos.

Debido a la casi inexistencia de este tipo de procesos, en realidad el mejoramiento de los procesos actuales se verá atacado por medio de la construcción e implementación de tres tipos de bitácoras basados en los comentarios de Inmon y Rudin [Inmon y Rudin, 98]:

- 1) Bitácora de proceso general: En esta se incluirá información de las corridas de procesos para verificar su correcta y satisfactoria terminación pero por módulos.
- 2) Bitácora de datos: En esta se plantea ir a la fuente de datos y comparar los datos obtenidos con los almacenados.
- 3) Bitácora de procesos individuales: En esta se incluirá información de las corridas de procesos para verificar su correcta y satisfactoria terminación a nivel individual.

Los resultados para medir el nivel de mejoramiento se basarán en una recopilación del porcentaje de errores detectados a tiempo y fuera de tiempo antes y después de que se implementaran las bitácoras.

### **3.3.4 Mejoramiento en la Explotación**

Este punto es muy interesante ya que utiliza parte de la metodología de desarrollo que aplica para el Back End y se acopla para poder desarrollar y entender los medios de explotación (reportes) a utilizar.

Por medio de la eliminación de múltiples herramientas para realizar la explotación proporcionando una plataforma estándar (en este caso Microstrategy) como medio de explotación, además proporcionando un mejor control de accesos, es como se atacará este punto.

El medio para medir el grado de cumplimiento de este objetivo será la medición de tiempos requeridos para explotar información y realizar análisis, es decir, la disminución de tiempos que tomará el hecho de que el usuario no tenga que estar moviendo información de diferentes fuentes por medio de diferentes herramientas ya que la arquitectura permitirá hacerlo todo desde un sólo lugar.

**Cabe recalcar que la parte de explotación es enorme y que existen diversas formas de acceder a una FIC ya que ésta es la centralización de la explotación de información. Sólo se tratarán 2 reportes de explotación como parte del alcance los cuales únicamente trabajarán sobre el ODS y que servirán como base y nos ayudarán a medir los tiempos de ahorro, si es que se dan en la explotación básica que se propone de la FIC.**



## 4 DESARROLLO DEL MÉTODO

### 4.1 Análisis de estado actual.

El estado actual del DW como tal se presenta en la figura 4.1.

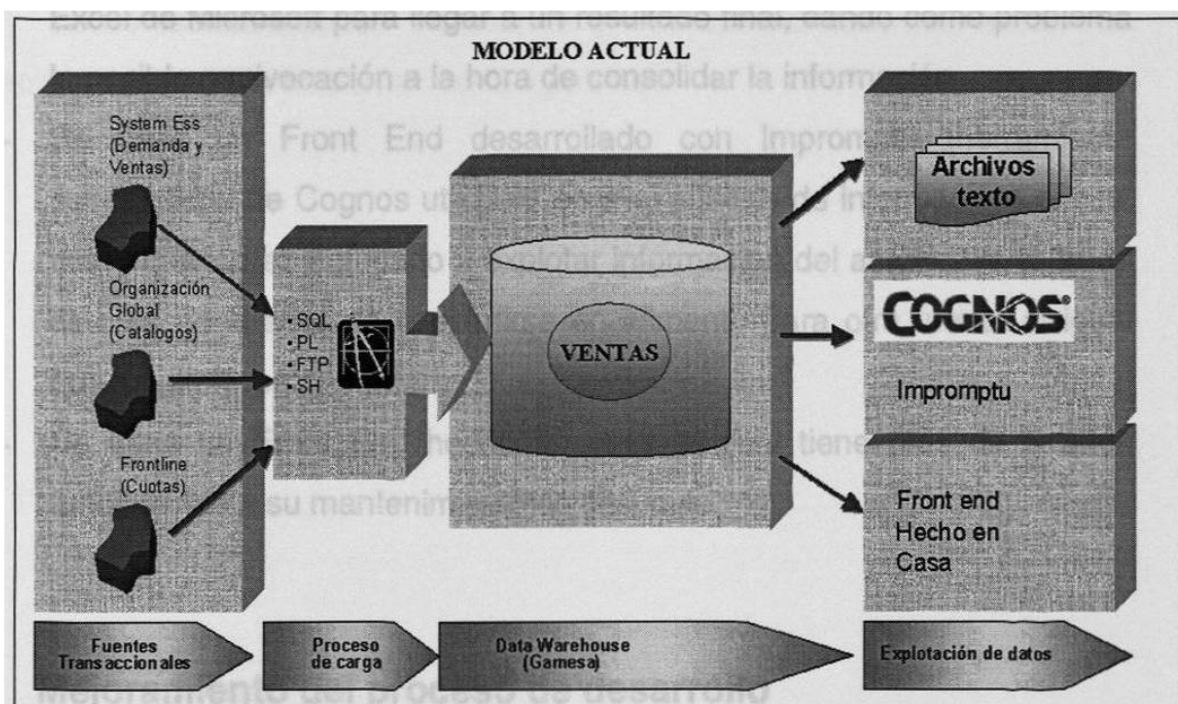


Figura 4.1 Diagrama de situación actual

En este se resume lo siguiente:

- Se tienen tres diferentes fuentes actualmente las cuales contienen la información de la demanda, las ventas y la parte de catálogos, además que se obtiene información de objetivos de venta.

- Actualmente no se tiene una herramienta especializada para la explotación, transformación y carga por lo que se tienen que usar herramientas básicas de trabajo poco administrables como FTP "File Transfer Protocol" para transferencia de archivos , programas en UNIX (shells), programas en SQL directos, procedimientos de Oracle (PL).
- La información se almacena diariamente desfasada a un día anterior.
- Sólo existe una base de datos de almacenamiento con una arquitectura de tipo "Snowflake" o copo de Nieve.
- Existen diferentes medios para acceder la información
  - Lamentablemente una fuente es alimento para otra, por ejemplo los archivos de texto que se generan se reprocesan en herramientas como Excel de Microsoft para llegar a un resultado final, dando como problema la posible equivocación a la hora de consolidar la información.
  - Se tiene un Front End desarrollado con Impromptu (herramienta especializada de Cognos utilizada para el análisis de información) el cual aún y cuando es enfocado a explotar información del actual DW el fin de este es generalmente convertirse en alimento para otra fuente al igual que los archivos.
  - Se tiene un Front End hecho en casa el cual tiene más de 5 años funcionando y su mantenimiento es casi nulo.

## **4.2 Mejoramiento del proceso de desarrollo**

Dentro del proceso de desarrollo se prosigue a llenar la metodología. Puntos estratégicos a considerar son:

- No avanzar sin un plan bien organizado.
- Recordar que la obtención de requerimientos es fundamental ya que es el pilar de la fábrica de información.

- El proceso debe de reflejar una arquitectura incremental. Se busca la estandarización de iteraciones para que los desarrollos subsecuentes sean más sencillos de incorporar.

### **4.2.1 Plan de proyecto**

Es un hecho que los tiempos pueden variar de proyecto en proyecto, pero en este se plantean las fechas tentativas teóricas a utilizar y desarrollar.

Otro punto a considerar es que únicamente un solo recurso estuvo desarrollando las actividades.

Todas las actividades necesarias y sus correspondientes fechas de duración tentativas se presentan en la figura 4.2.

ID	Task Name	Duration	Start	01				
				T	W	T	F	S
1	FIC-Iteracion Ventas	71 days	Tue 5/1/01					
2	<b>BACKEND</b>	55 days	Tue 5/1/01					
3	<b>ANÁLISIS</b>	5 days	Tue 5/1/01					
4	Análisis de Estatus actual	5 days	Tue 5/1/01					
6	Programación Juntas y Entrevistas	5 days	Tue 5/1/01					
8	Identificación de las fuentes	5 days	Tue 5/1/01					
10	<b>DISEÑO</b>	10 days	Tue 5/8/01					
11	Diseño de modelo general (Inmon)	3 days	Tue 5/8/01					
13	Metodología Kimball	9 days	Fri 5/11/01					
14	Construcción de matriz de bus	1 day	Fri 5/11/01					
16	Realización del modelado dimensional	3 days	Mon 5/14/01					
20	Complemento selección de fuentes	1 day	Thu 5/17/01					
22	Realización de mapeo campo a campo	4 days	Fri 5/18/01					
24	Desarrollo de modelo lógico-físico	2 days	Thu 5/24/01					
26	Realización de dimensionamiento	1 day	Mon 5/28/01					
28	Diseñar la arquitectura a utilizar	2 days	Tue 5/29/01					
30	Diseñar mapa de configuración de procesos	2 days	Thu 5/31/01					
32	<b>CONSTRUCCIÓN</b>	20 days	Mon 6/4/01					
33	Crear usuarios y aspectos de bases de datos(TS)	2 days	Mon 6/4/01					
34	Crear estructura de la Base de datos ODS	2 days	Wed 6/6/01					
35	Crear estructura de la Base de datos DW	2 days	Fri 6/8/01					
36	Construcción de Interfases	14 days	Tue 6/12/01					
37	Carga dimensiones ODS	3 days	Tue 6/12/01					
38	Carga demanda ODS	3 days	Fri 6/15/01					
39	Carga Ventas ODS	2 days	Wed 6/20/01					
40	Carga Cuotas ODS	2 days	Fri 6/22/01					
41	Carga DW	1 day	Tue 6/26/01					
42	Construcción de bitácoras validadoras	3 days	Wed 6/27/01					
43	Bitácora 1	1 day	Wed 6/27/01					
44	Bitácora 2	1 day	Thu 6/28/01					
45	Bitácora 3	1 day	Fri 6/29/01					
46	<b>PRUEBAS</b>	9 days	Mon 7/2/01					
47	Pruebas individuales de procesos	3 days	Mon 7/2/01					
48	Pruebas individuales de datos	3 days	Thu 7/5/01					
49	Pruebas integrales	3 days	Tue 7/10/01					
50	<b>IMPLEMENTACIÓN</b>	2 days	Fri 7/13/01					
51	Actividades de liberación	2 days	Fri 7/13/01					
52	<b>FRONTEND</b>	16 days	Tue 7/17/01					
53	Apoyo en análisis de Back End para entender requerimientos	1 day	Tue 7/17/01					
54	<b>DISEÑO</b>	6 days	Wed 7/18/01					
55	Investigar los accesos de los reportes (quien los lee, cuantos usuarios)	3 days	Wed 7/18/01					
56	Diseño tentativo de reportes que se requerirán	3 days	Mon 7/23/01					
57	<b>CONSTRUCCIÓN</b>	5 days	Thu 7/26/01					
58	Definición de atributos, métricas	1 day	Thu 7/26/01					
59	Construcción de reporte 1	3 days	Thu 7/26/01					
60	Construcción de reporte 2	2 days	Tue 7/31/01					
61	<b>PRUEBAS</b>	3 days	Thu 8/2/01					
62	Pruebas de reportes contra información	3 days	Thu 8/2/01					
63	<b>IMPLEMENTACION</b>	1 day	Tue 8/7/01					
64	Actividades de implementación	1 day	Tue 8/7/01					

Figura 4.2. Plan de proyecto

## 4.2.2 Obtención de requerimientos

Para la obtención de requerimientos fue necesario una serie de reuniones con las gentes de ventas en las cuales por medio de apuntes se realizó la recopilación de información. Un punto importante es que no sólo se obtuvieron los requerimientos directamente del usuario, se utilizaron métodos tales como pláticas informales, entendimiento de documentación existente, por medio de las cuales se pudo lograr el entendimiento de los conceptos.

Después de aplicar el documento propuesto en el anexo A2 se presenta el siguiente resumen:

Los indicadores detectados se presentan en la tabla 4.1.

Indicadores	Dimensión	Dimensi ón	Dimensión	Dimensión	Dimensión
Ventas Brutas(\$)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Ventas Brutas (Kgs)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Ventas Brutas (Uds)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Ventas Netas (\$)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Ventas Netas (Kgs)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Ventas Netas (Uds)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Venta Promocion (\$)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Venta Promocion (Kgs)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Venta Promocion (Uds)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Ventas Descuentos Promocional, Especial, Línea (\$)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Demanda* Bruta (\$)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Demanda* Bruta (Kgs)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Demanda* Bruta (Uds)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Demanda Neta (\$)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Demanda* Netas (Kgs)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Demanda* Neta (Uds)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Demanda* Promocion (\$)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Demanda* Promocion (Kgs)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Demanda* Promocion (Uds)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Demanda* Descuentos Promocional, Especial, Línea (\$)	Producto	Cliente	Estructura	Punto de Facturación	Tiempo
Cuota Kgs	Línea Categoría	Cliente	Estructura	-	Tiempo

Tabla 4.1. Requerimientos

\* La demanda se considera para los diferentes tipos que existen: Captada, Autorizada, Cancelada, Depurada, Rechzada por créditos, Pendiente por surtir.

- **Necesidades de información histórica**

Las necesidades de información se resumen en dos tipos:

- 1) Información a nivel diario para la actividad operativa con 3 periodos del calendario general de GAMESA (Cada periodo consta de 28 días)
- 2) Información histórica de 5 años a la fecha para análisis estadísticos y de tendencias en sus conceptos de Demanda y Venta.

- **Requerimientos analíticos**

- **Panorama del requerimiento formal**

El objetivo a buscar es tener una solución que permita al área de ventas tener una visualización de la información relacionada con la demanda y sus ventas para poder llevar a cabo análisis tanto a un nivel operativo (para solventar las decisiones del día a día) y a un nivel más consolidado permitiendo análisis de tendencias y la utilización de esta información para la toma de decisiones de negocio.

- **Potencial de impacto en el negocio**

El potencial de esta implementación radica en la facilidad de las diferentes áreas de la compañía de poder realizar sus análisis debido a que actualmente los medios por el cual obtienen la información, además de ser muy tardados no son los oficiales y por lo tanto la información presentada puede presentar inconsistencias ya sea por las diferentes fuentes o por errores o manipulaciones de las personas que realicen la explotación.

- **Análisis típicos de información**

Los análisis son:

- Comparativos día a día de las ventas.
- Comparativos del día a día contra periodos o años anteriores.
- Análisis de cumplimiento de objetivos (cuotas) contra lo facturado.
- Análisis de todo el panorama de demanda con su venta.

- **Comparativo de todo lo pendiente por surtir contra lo facturado y contra la cuota.**

Todas estas explotaciones necesitan de un mínimo nivel de detalle debido a que la información se requiere poder explotar en las diferentes dimensiones del negocio. Por ejemplo a nivel Cliente para poder realizar análisis por canal de ventas o por cadenas y grupos, a nivel producto para analizar por las diferentes jerarquías de producto y más aún por Estructura de Ventas para analizar por las diferentes jerarquías de la estructura geográfica en que se encuentran.

- **Tipos de accesos requeridos**

- **Análisis multidimensionales, Accesos en línea, Navegaciones (drills), posibilidad de importar a otras herramientas de trabajo común, Alarmas.**

Se requiere de una herramienta flexible que tenga las funcionalidades de poder realizar todas las navegaciones requeridas a todos los niveles necesarios, permitiendo una rápida obtención de información. Se necesitan accesos en línea con información y de preferencia vía "Web".

- **Análisis preliminar de las fuentes de información**

- *System ESS.*

Este es una herramienta del tipo ERP que permite llevar a cabo toda la administración de la demanda. En ella se incluyen conceptos que van desde el mantenimiento de las ordenes, su seguimiento hasta finalizar el ciclo de la facturación. Esta es una herramienta montada en una plataforma de UNIX HP-UX 11.0 cuyos datos se encuentran almacenados en una base de datos relacional ORACLE v7.3.4.5.0.

La herramienta trabaja durante todo el día teniendo 2 respaldos diarios de 5 minutos por lo que su disponibilidad es relativamente alta.

#### 4 Frontline (Archivos Planos) para cuotas.

Frontline se encarga del ingreso de la demanda en campo a través de herramientas como LOTUS NOTES, Z-merge, etcétera. La parte que se requiere es la referente con las cuotas u objetivos la cual se proporcionará por medio de archivos texto.

#### - Organización Global

Este es una extensión del ERP por lo que tiene todas sus características. Se utiliza para extraer los catálogos de la compañía.

### 4.2.3 Diseño de modelo general de la FIC propuesto

El modelo propuesto implica los siguientes componentes (mostrados en la figura 4.3): El mundo externo, la capa de aplicación, la capa de lyT por medio de una herramienta especializada, el DW, el ODS, el pilar de metadata por medio de las propias herramientas utilizadas y por la documentación propuesta.

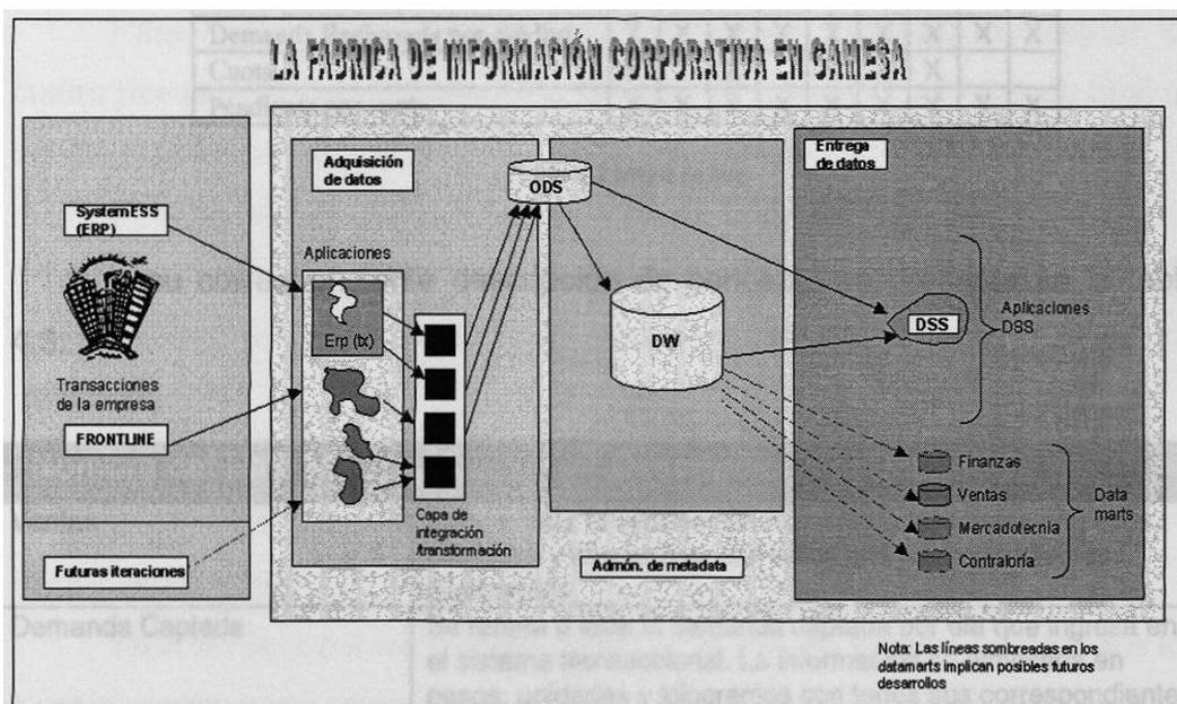


Figura 4.3. FIC propuesta



#### 4.2.4 Construcción de matriz de bus

La complejidad que se ve en este desarrollo es que la cuota u objetivo no se encuentra en todos los niveles de navegación que se requiere en el ODS y en el DW por lo que cualquier combinación con cuotas se verá limitada por esta deficiencia.

La matriz de bus aplicada se presenta en la tabla 4.2 obtenida del anexo A3:

Dimensiones →

Conceptos de negocio (Datamarts)	Estructura	Punto de facturación	Cliente	Producto	Línea	Categoría	Tiempo	Factura	Folio pedido
Ventas	X	X	X	X	X	X	X	X	
Demanda Captada	X	X	X	X	X	X	X		X
Demanda Autorizada	X	X	X	X	X	X	X		X
Demanda Cancelada	X	X	X	X	X	X	X		X
Demanda Depurada	X	X	X	X	X	X	X		X
Demanda Rechazada por crédito	X	X	X	X	X	X	X	X	X
Cuota	X		X		X	X	X		
Pendiente por surtir	X	X	X	X	X	X	X	X	X

Tabla 4.2 Matriz de bus

Y su correspondiente descripción de concepto se presenta en la tabla 4.3.

Nombre del Data Mart	Descripción del Datamart
Ventas	Incluye toda la información referente a las ventas en pesos, unidades y kilogramos, con todos sus correspondientes descuentos
Demanda Captada	Se refiere a toda la demanda captada por día que ingresa en el sistema transaccional. La información se almacena en pesos, unidades y kilogramos con todos sus correspondientes descuentos
Demanda Autorizada	Se refiere a toda la demanda por día que se autoriza en el

Nombre del Data Mart	Descripción del Datamart
	sistema transaccional. La información se almacena en pesos, unidades y kilogramos con todos sus correspondientes descuentos
Demanda Cancelada	Se refiere a toda la demanda por día que es cancelada debido a problemas administrativos de Gamesa. La información se almacena en pesos, unidades y kilogramos con todos sus correspondientes descuentos
Demanda Depurada	Se refiere a toda la demanda por día que es cancelada debido a problemas del Cliente. La información se almacena en pesos, unidades y kilogramos con todos sus correspondientes descuentos
Demanda Rechazada por crédito	Se refiere a toda la demanda por día que es cancelada debido rechazos por crédito. La información se almacena en pesos, unidades y kilogramos con todos sus correspondientes descuentos
Cuota	Se refiere a los objetivos en kilogramos que se busca cumplir con la venta
Pendiente por surtir	Se refiere a toda la demanda por día que se encuentra pendiente por surtir. La información se almacena en pesos, unidades y kilogramos con todos sus correspondientes descuentos

Tabla.4.3 Desglose de conceptos en Matriz de bus

#### 4.2.5 Realización del modelado Dimensional

Para la realización del modelado dimensional a partir de aplicar los cuatro pasos para llevar a cabo esto (siendo uno de ellos la matriz de bus) se concluyó lo siguiente:

Para la venta y los diferentes tipos de demanda se requiere una granularidad de todos los niveles de almacenamiento a nivel ODS y DW. El detalle va a quedar almacenado a un nivel diario como atributo de la dimensión de tiempo.

Para la demanda pendiente por surtir, cómo es un concepto volátil y sirve para tomar decisiones en forma operativa, se almacenará únicamente a nivel ODS y representará una foto del día en que se realice la extracción.

Para la cuota por ser un caso especial en que no se tiene sobre todos los niveles, entonces se almacenará en 2 tablas individuales. Teniendo diferentes

tipos de granularidad siendo la almacenada por categoría a nivel periodo GAMEESA y la almacenada por línea a un nivel diario.

Debido a que realizar el documento con el diagrama y el detalle de la tabla de hechos y de las dimensiones tanto para el ODS, como el DW es muy extenso, esta se presenta en el anexo B (Modelado dimensional).

El documento con los hechos derivados se presenta en el anexo C.

#### 4.2.6 Identificación de las fuentes de cada tabla de hechos y dimensiones

Las fuentes identificadas se resumen en dos donde una de ellas servirá para obtener la información de Indicadores y de dimensiones, mientras que la otra servirá para obtener el indicador de cuota u objetivo.

Basándonos en la aplicación del formato presentado en el anexo A6 se presentan los siguientes resultados. Las fuentes se presentan en la tabla 4.4.

Fuente	Área de negocio responsable	Reponsable de informática	Plataforma	Localización	Descripción
System ESS	Desarrollo y Mantenimiento al Comercio		UNIX		Este sistema es el encargado de la administración de la demanda. Desde que se ingresa el pedido, hasta que se factura.
System ESS	Organización global		UNIX		Se encarga de la administración de todos los catálogos necesarios para el funcionamiento de los sistemas.

Fuente	Área de negocio responsable	Reponsable de informática	Plataforma	Localización	Descripción
Sistema administrador de objetivos. (Archivos Texto)	Ventas		NT		Este sistema se encarga de administrar todos los objetivos de ventas.

Tabla 4.4 Sistemas Identificados

Dentro de los criterios de selección se obtuvieron los siguientes resultados:

- **Accesibilidad de los datos**

Para el caso de "System ESS" es un sistema totalmente transaccional y por consiguiente es posible realizar accesos a casi cualquier hora del día siempre y cuando no concordemos con los horarios de respaldo. En este sistema se encuentran accesibles la mayoría de los indicadores necesarios para la iteración a utilizar. Además la accesibilidad a los catálogos para generar las dimensiones se refleja en el hecho de que todos los sistemas se apoyan en éstos. Tenemos la ventaja que los datos se encuentran almacenados en una base de datos Oracle por lo que su acceso es fácil.

Para el caso del Sistema de Cuotas u Objetivos es la única fuente por medio de la cual se pueden tomar estos indicadores. Su accesibilidad es relativa y dependemos de que se nos envíe la información por vía archivo texto ya que la información está almacenada en Lotus Notes y por consiguiente en un ambiente totalmente diferente a uno relacional.

- **Exactitud de los datos**

La información de System ESS es la actual transaccional por lo que su exactitud coincide con la realidad del sistema y las variaciones de su exactitud se puede deber a posibles errores de captura únicamente.

Para el caso de Cuotas u Objetivos podemos enfrentarnos al hecho de que la información puede ser leída un día y modificada en días posteriores.

- **Calendarización de proyectos**

No existe calendarización programada para migraciones de estos sistemas.

#### **4.2.7 Mapeo fuente a destinos**

Un punto muy importante y de los más laboriosos es el mapeo de cada una de las fuentes a sus correspondientes destinos.

El desplegar el mapeo campo a campo a detalle no genera un valor agregado para el objetivo de esta tesis por lo que no se abordará a detalle.

#### **4.2.8 Desarrollo por medio de herramienta modeladora de modelo físico a un manejador de Base de datos**

El modelo lógico-físico de la base de datos se desarrolló pensando en las 2 arquitecturas tanto ODS y el DW.

Para facilidad de la lectura de esta tesis el modelo lógico - físico de la B.D. se presenta en el anexo D. En este se presenta por concepto de negocio los diagramas estrella a utilizar. Uno de los puntos importantes a recalcar es que las referencias de integridad se presentan únicamente en el modelo ODS mientras que las tablas de DW no guardan una relación con las dimensiones por cuestiones de optimización. La granularidad utilizada se definió como diaria para el ODS y semanal para el DW.

La tabla 4.5 muestra los diferentes tipos de tablas a encontrar en el modelo.

Abreviatura	Tipo de tabla
CF_	Dimensión
ODS_	Tabla de hechos del ODS
DW_	Tabla de hechos del DW
TMP_	Tabla de carga temporal al ODS

Tabla 4.5 Tipos de tablas del modelo

## 4.2.9 Dimensionamiento

Para realizar el dimensionamiento se aplicaron las reglas mostradas en el Capítulo 3 punto 3.3.2.1.9. El resumen del dimensionamiento se presenta en la tabla 4.6.

Índice	Tabla	Initial Extent*	Next Extent*	TS
-	ODS_PEDIDO	277,236,000	13,861,800	N/A
-	ODS_PENDIENTE_SURTIR	626,400,000	31,320,000	N/A
-	ODS_FACTURACION	92,412,000	4,620,600	N/A
-	ODS_CUOTA_CLIENTE	1,425,600	71,280	N/A
-	ODS_CUOTA_CATEGORIA	138,240	33,000	N/A
-	CF_CLIENTE	23,940,000	1,197,000	N/A
-	CF_PRODUCTO	7,980,000	399,000	N/A
-	CF_ESTRUCTURA	160,200	33,000	N/A
-	CF_CONCEPTO	33,000	33,000	N/A
-	CF_TIEMPO	125,362	33,000	N/A
-	CF_PUNTO_FACTURACION	33,000	33,000	N/A
-	TMP_ESS_PEDIDO	5,736,000	286,800	N/A
-	TMP_ESS_FACTURACION	5,184,000	259,200	N/A
-	TMP_ESS_PENDIENTE_SURTIR	3,984,000	199,200	N/A
-	TMP_CUOTA_CATEGORIA	723,840	36,192	N/A
-	TMP_CUOTA_CLIENTE	344,400	33,000	N/A
-	DW_FACTURACION	2,500,000,000	125,000,000	N/A
-	DW_PEDIDO	6,105,715,200	305,285,760	N/A
-	DW_PENDIENTE_SURTIR	389,376,000	19,468,800	N/A
-	DW_CUOTA_CATEGORIA	17,971,200	898,560	N/A
-	DW_CUOTA_CLIENTE+B2	29,952,000	1,497,600	N/A
ODS_XPK_PEDIDO	ODS_PEDIDO	75,276,000	3,763,800	N/A
ODS_XPK_FACTURACION	ODS_FACTURACION	25,092,000	1,254,600	N/A
ODS_XPK_PENDIENTE_SURTIR	ODS_PENDIENTE_SURTIR	70,200,000	3,510,000	N/A
ODS_XPK_CUOTA_CATEGORIA	ODS_CUOTA_CATEGORIA	103,680	33,000	N/A
ODS_XPK_CUOTA_CLIENTE	ODS_CUOTA_CLIENTE	1,166,400	58,320	N/A

CF_XPK_CLIENTE	CF_CLIENTE	216,000	33,000	N/A
CF_XPK_PRODUCTO	CF_PRODUCTO	72,000	33,000	N/A
CF_XPK_ESTRUCTURA	CF_ESTRUCTURA	33,000	33,000	N/A
CF_XPK_CONCEPTO	CF_CONCEPTO	33,000	33,000	N/A
CF_XPK_TIEMPO	CF_TIEMPO	33,000	33,000	N/A
CF_XPK_PUNTO_FACTURACION	CF_PUNTO_FACTURACION	33,000	33,000	N/A
DW_XPK_FACTURACION	DW_FACTURACION	1,000,000,000	50,000,000	N/A
DW_XPK_PEDIDO	DW_PEDIDO	1,657,843,200	82,892,160	N/A
DW_XPK_PENDIENTE_SURTIR	DW_PENDIENTE_SURTIR	101,088,000	5,054,400	N/A
DW_XPK_CUOTA_CATEGORIA	DW_CUOTA_CATEGORIA	13,478,400	673,920	N/A
DW_XPK_CUOTA_CLIENTE	DW_CUOTA_CLIENTE	23,961,600	1,198,080	N/A

Tabla 4.6 Dimensionamiento

\* La información de los segmentos initial y next está presentada en Bytes

#### 4.2.10 Diagrama de arquitectura tecnológica.

El plan de la arquitectura tecnológica se trazó basado en los requerimientos del negocio y analizando cada uno de los puntos a considerar de lo cual se obtuvo la siguiente información:

- Entendimiento de los requerimientos del negocio

Se requiere de un lugar de almacenamiento el cual permita concentrar todos los datos referentes a actividades relacionadas con el negocio y que permita una explotación de indicadores que proporcionen un mayor valor agregado a la compañía de una manera rápida y segura.

- Acceso de información

Se requiere disponibilidad para poder tener accesos rápidos sobre información relativamente operativa los cuales deben de ser ya sea fijos con reportes previamente codificados, o con la disponibilidad de crear los reportes necesarios sin dependencia directa del área de IT. Además se debe tener acceso a información histórica en la cual se permita un análisis de indicadores pero que se visualice como una futura fuente de información para descubrir indicadores importantes inexistentes del negocio.

- **Navegación**

La navegación que se busca debe de ser intuitiva para el usuario por medio de la interfaz gráfica permitiendo una navegación hacia abajo, arriba o a través de las dimensiones. Se busca que el usuario sea relativamente autónomo. Se busca además que la navegación se realice vía "Web".

- **Calidad de datos**

La información deberá de estar verificada por medios de validación tales que permitan conocer errores y que sean corregidos tan pronto sean detectados. Los datos deberán de ser los más correctos posibles ya que cierta información que se distribuya afectará operativamente en ciertas áreas en las cuales se puede ver reflejado los errores en pagos erróneos.

- **Cuestiones organizacionales**

Se plantea que la distribución del software para la explotación sea transparente ya que se realizaría vía Navegación de "Web". Se plantea además una capacitación sobre las diferentes formas de navegación y el funcionamiento del explotador. Las comunicaciones serían actualmente vía TCP/IP en donde utilizaríamos la Intranet para poder conectar al servidor principal de explotación.



## • Una revisión de la arquitectura

Textualmente la arquitectura se basa en un flujo que inicia con la conexión a las diferentes fuentes de datos para su posterior procesamiento en el área de procesamiento por medio de un servidor necesario para la herramienta ETL seleccionada el cual se encargará de almacenar al ODS-DW para su futura explotación. Para la explotación se plantea un servidor el cual se encargará de la administración de la misma y en donde se realizarán cálculos mínimos necesarios para darles un mayor valor agregado a los reportes.

## • Elementos de la arquitectura

La arquitectura a utilizar necesita de los siguientes elementos para poder funcionar:

Una herramienta especializada de ETL en la cual se tienen los servicios de control de versiones, construcciones de sistemas de desarrollo y producción, una interfaz gráfica intuitiva que permita una mayor productividad, que trabaje basado en su metadata y que mejor aún crea su propia metadata explotable por la gente desarrolladora. Además proporciona los servicios de extracción permitiendo la conexión a múltiples fuentes, la generación de código básico, diferentes tipos de extracción.

Aunado a estos servicios se utilizarán además servicios de transformación para integración, chequeo de referencias de integridad, conversión de tipos de datos, limpieza de los mismos y depuración, así como cálculos y agregaciones; servicios de carga como el soporte de múltiples destinos, optimización de cargas; servicios de calendarización, definición de trabajos, monitoreo, manejo de excepciones y de errores.

Dentro de los servicios para el acceso de los datos se proporcionará una navegación del DW y del ODS, incluyendo servicios de seguridades y control de accesos. Además también se incluyen los servicios de monitoreo de explotaciones, servicios estándar de reporte (ambiente de desarrollo, servidor

de ejecución, capacidades de manejar parámetros, calendarización de reportes, entrega flexible de reportes, distribución masiva y herramientas de administración).

Finalmente los almacenamientos se plantean en bases de datos distintas para poder ser configuradas con los respectivos requerimientos que cada una exige.

El modelo de la arquitectura tecnológica se presenta en la figura 4.4.

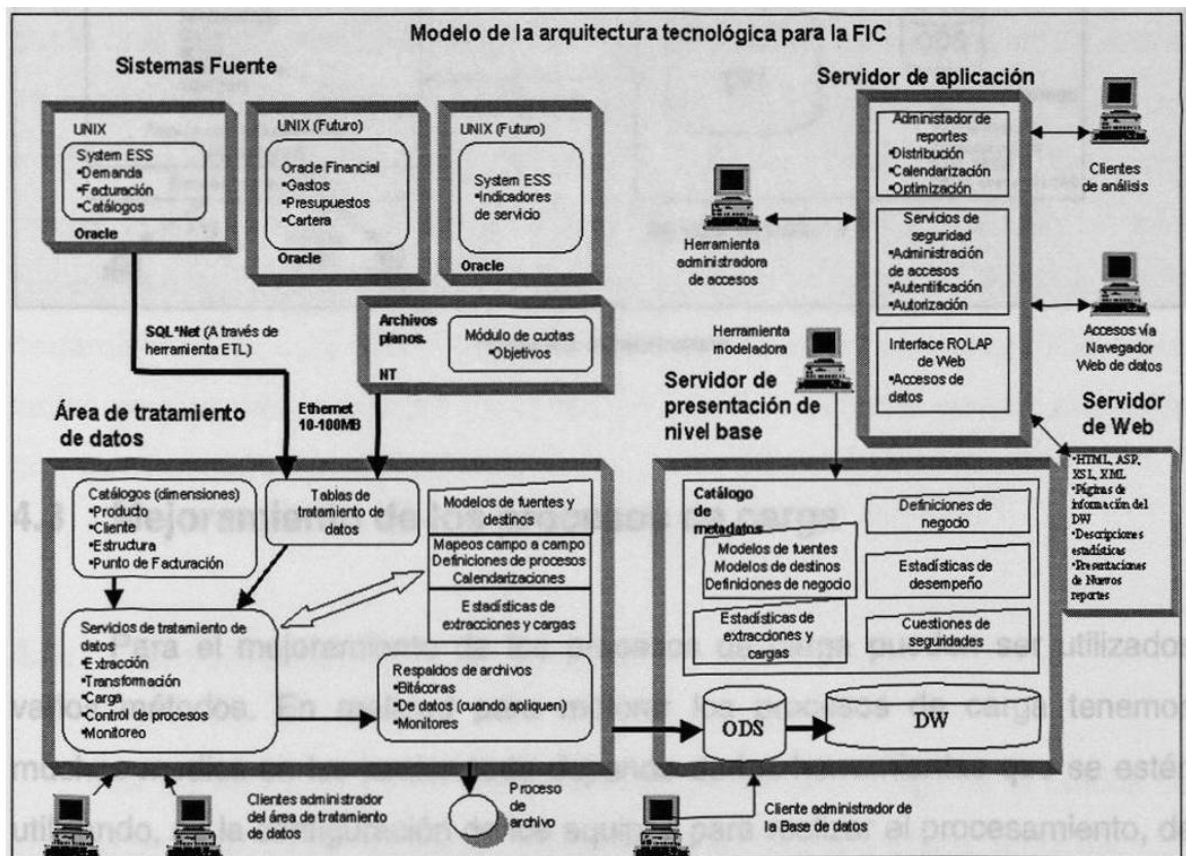


Figura 4.4. FIC arquitectura propuesta

Finalmente el modelo de la infraestructura se presenta en la figura 4.5:

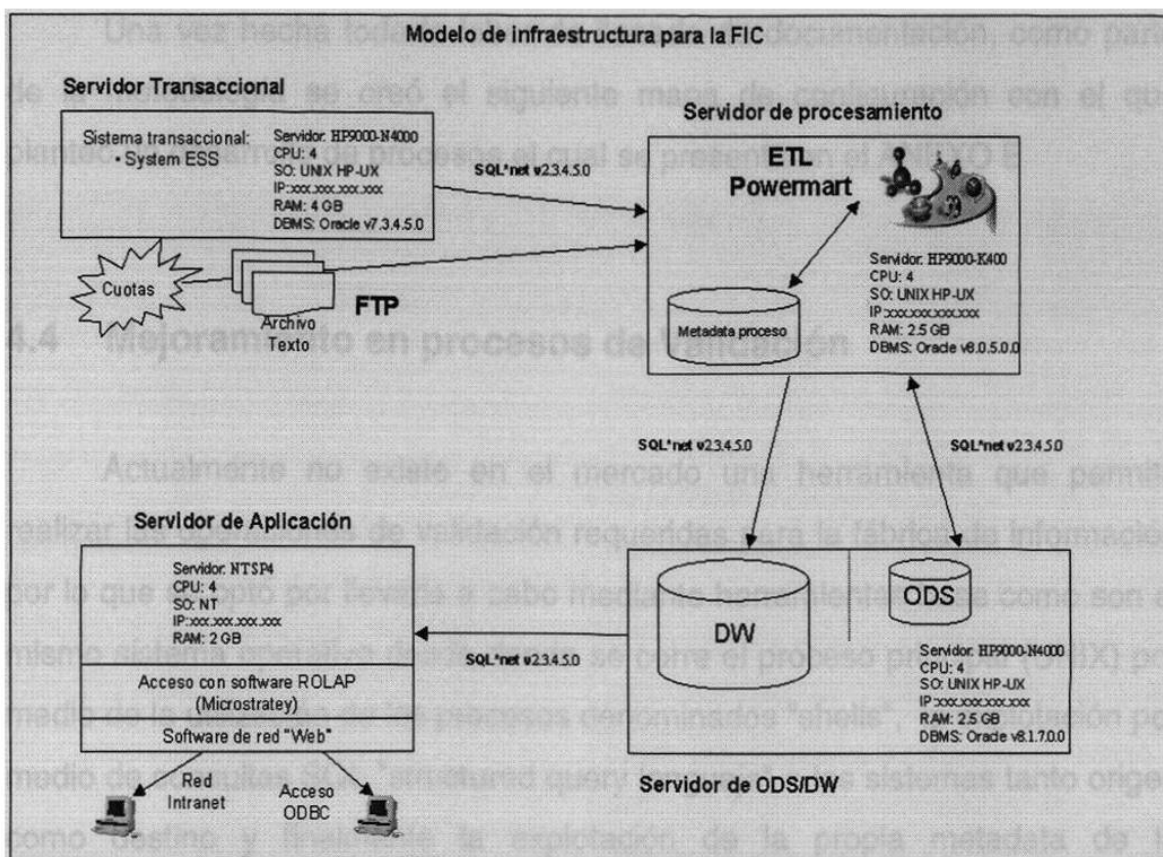


Figura 4.5. Infraestructura

### 4.3 Mejoramiento de los procesos de carga

Para el mejoramiento de los procesos de carga pueden ser utilizados varios métodos. En realidad para mejorar los procesos de carga tenemos muchos medios en los cuales todo depende de las herramientas que se estén utilizando, de la configuración de los equipos para realizar el procesamiento, de las características de red, de las cantidades de información a procesar y transformar, y otros factores que pueden optimizar las diferentes actividades que requiere el proceso de carga.

En mi propuesta únicamente me enfoco en la organización de estos procesos por medio de una herramienta especializada para el desarrollo de los mismos.