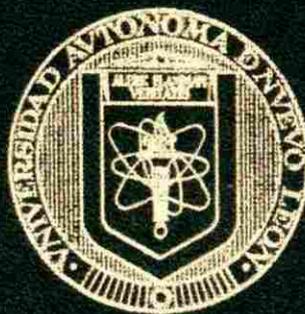


UNIVERSIDAD AUTONOMA DE NUEVO LEON
FACULTAD DE INGENIERIA MECANICA Y ELECTRICA
DIVISION DE ESTUDIOS DE POSGRADO



**"MEJORAMIENTO EN LOS PROCESOS
DE DESARROLLO, CARGA, VALIDACION
Y EXPLOTACION IMPLICADOS EN UN DATA
WAREHOUSE MEDIANTE LA CREACION DE UNA
FABRICA DE INFORMACION CORPORATIVA
APLICADA AL AREA DE VENTAS DE LA
EMPRESA GAMESA"**

POR

ING. RODOLFO PORTILLO SALINAS

TESIS

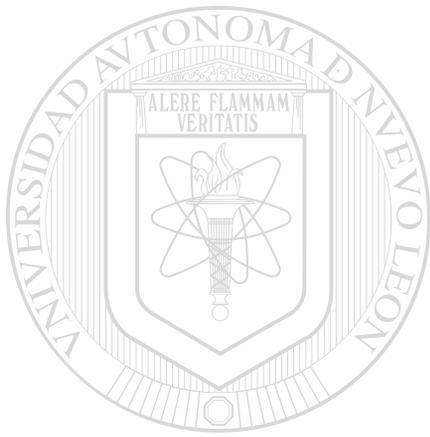
**EN OPCION AL GRADO DE MAESTRO EN CIENCIAS
DE LA ADMINISTRACION CON ESPECIALIDAD
EN SISTEMAS**

SAN NICOLAS DE LOS GARZA, N. L. JUNIO DEL 2002

"MEJORAMIENTO EN LOS PROCESOS DE DESARROLLO
CARGA, VALIDACION Y EXPLOTACION IMPLICADOS EN UN
DATA WAREHOUSE MEDIANTE LA CREACION DE UNA
FABRICA DE INFORMACION CORPORATIVA APLICADA AL
AREA DE VENTAS DE LA EMPRESA GAMESA"

TM
QA76
.9
.D37
P6
2002
c.1

2002



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

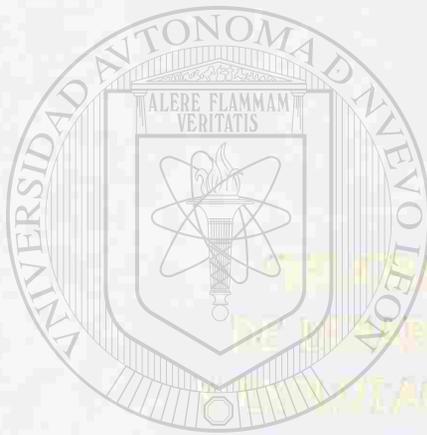


DIRECCIÓN GENERAL DE BIBLIOTECAS

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

DIVISION DEL INSTITUTO DE POSGRADO



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
APLICADA AL ÁREA DE VENTAS DE LA
DIRECCIÓN GENERAL DE BIBLIOTECAS

POR

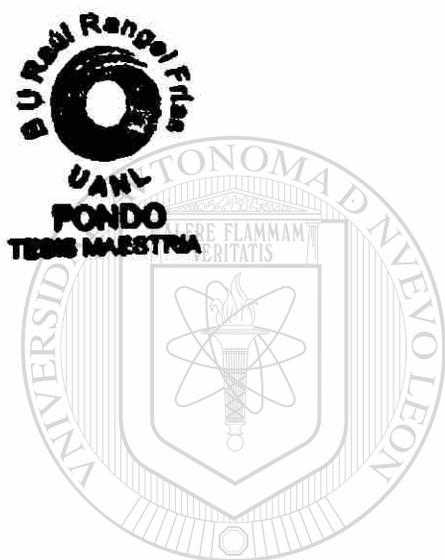
ING. RODOLFO PORTILLO SALINAS

TESIS

EN OPCIÓN AL GRADO DE MAESTRO EN CIENCIAS
DE LA ADMINISTRACIÓN CON ESPECIALIDAD
EN SISTEMAS

SAN NICOLÁS DE LOS GARZA, N. L. JUNIO DEL 2008

TM
QA76
.9
.D37
P6
2002



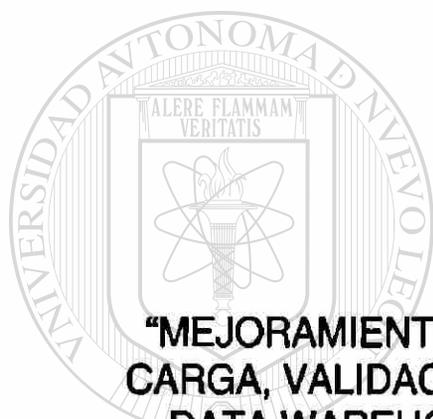
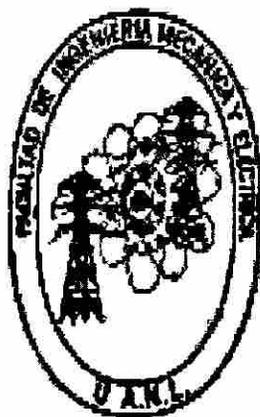
UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

®

DIRECCIÓN GENERAL DE BIBLIOTECAS

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA
DIVISIÓN DE ESTUDIOS DE POSGRADO



**“MEJORAMIENTO EN LOS PROCESOS DE DESARROLLO,
CARGA, VALIDACIÓN Y EXPLOTACIÓN IMPLICADOS EN UN
DATA WAREHOUSE MEDIANTE LA CREACIÓN DE UNA
FÁBRICA DE INFORMACIÓN CORPORATIVA APLICADA AL
ÁREA DE VENTAS DE LA EMPRESA GAMESA”**

**DIRECCIÓN GENERAL DE BIBLIOTECAS
POR**

ING. RODOLFO PORTILLO SALINAS

TESIS

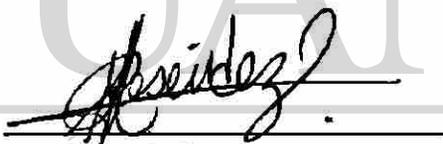
**EN OPCIÓN AL GRADO DE MAESTRO EN CIENCIAS DE LA
ADMINISTRACIÓN CON ESPECIALIDAD EN SISTEMAS**

SAN NICOLÁS DE LOS GARZA, N.L, JUNIO DEL 2002.

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA
DIVISIÓN DE ESTUDIOS DE POSGRADO

Los miembros del comité de tesis recomendamos que la tesis "Mejoramiento en los procesos de desarrollo, carga, validación y explotación implicados en un Data Warehouse mediante la creación de una fábrica de información corporativa aplicada al área de ventas de la empresa GAMESA", realizada por el alumno Ing. Rodolfo Portillo Salinas, matrícula 0763159 sea aceptada para su defensa como opción al grado de Maestro en Ciencias de la Administración con Especialidad en Sistemas.

El Comité de Tesis



Asesor

M.C. Rosa María Reséndez Hinojosa

DIRECCIÓN GENERAL DE BIBLIOTECAS



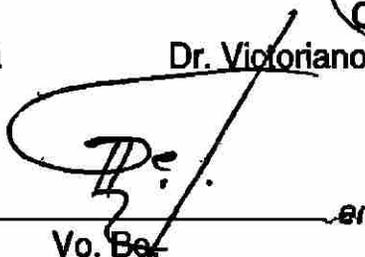
Coasesor

M.C. David Garza Garza



Coasesor

Dr. Victoriano F. Alatorre González



Vo. Be-

Dr. Guadalupe Alan Castillo Rodríguez
División de Estudios de Posgrado

San Nicolás de los Garza, N.L., Junio del 2002.

DEDICATORIAS

A Dios

*Gracias Señor, por darme la bendición de la vida
y la fortaleza para seguir adelante*



A mis padres

Irma Alicia Salinas y Rodolfo Portillo Rubalcava

Por todo su amor y su apoyo incondicional otorgado durante toda mi vida

UANL

A mis hermanos

Claudia, Luis y Jesús

Por apoyarme en todo momento y formar parte de mis logros

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
DIRECCIÓN GENERAL DE BIBLIOTECAS

A Verónica Avilés

*Por aparecer en mi vida brindándome todo su amor y su cariño, sus palabras
de aliento y toda la dicha y felicidad, que me fortalece
y me hace crecer y ser mejor cada día*

AGRADECIMIENTOS

A mi asesora la M.C. Rosa María Reséndez Hinojosa por el tiempo, la paciencia, los consejos, las correcciones y sus palabras de aliento.

A la compañía GAMESA y a mi jefe, el Ing. Daniel Badillo, por permitirme desarrollar en la empresa la investigación tratada en esta tesis.

A Ralph Kimball una de las personas pilares en el concepto de Data Warehouse, el cual influyó con algunos de sus comentarios para definir parte del tema de esta tesis.

A mis coasesores

Al Ing. Ángel Hermida por haberme iniciado en el ámbito del concepto Data Warehouse.

A mis compañeros de trabajo Lic. José Manuel Guillén Tapia, Ing. Mauricio Loaiza Prat, Ing. José Favero y el Ing. Luis Fernando Valdéz que me ayudaron en alguna medida a implementar todos los conceptos que se plantean en esta tesis.

Al Ing. Salvador Flores quién con sus aportaciones y sus experiencias me ha ayudado para incrementar la calidad de mi tesis.

Al Ing. Elva Ruth Ortiz y al Instituto Tecnológico y de Estudios Superiores de Monterrey por permitirme acceder a una parte de la bibliografía usada para la realización de este trabajo de tesis.

Y a todos aquellos compañeros que de alguna manera han aportado algo para esta tesis y que listarlos a todos me sería imposible.

PRÓLOGO

Todo lo que se encuentra en este planeta se encuentra en constante evolución, desde un pequeño y diminuto microorganismo hasta las más complejas formas de vida. Esta metáfora también aplica para todo lo que el ser humano crea en su entorno, desde la forma de realizar las cosas hasta la forma de pensar. En esta tesis ataco una de las tantas formas en que el hombre, en su intento de ser más eficiente, trata de ser competitivo con respecto a los negocios, un punto importante en estos días en los que la competitividad, productividad y velocidad de la gente y las empresas de adaptarse al cambio son primordiales para poder subsistir.

Todo empieza por un intento del hombre por automatizar procesos, recopilar información, realizar transacciones y responder casi en línea ante las diferentes tareas que abarcan una cadena de administración en las empresas, para esto, desde los tiempos más remotos, estas actividades se realizaban por medios manuales los cuales realmente consumían tiempo y se volvían procesos poco eficientes y difíciles de controlar y administrar. Con la revolución tecnológica que llega, es decir, con la entrada de computadoras personales, servidores, medios de almacenamiento, redes, y más aún, con la entrada de los sistemas de información para administrar los diversos cambios en las empresas y sus recursos en forma integrada mejor conocidos como sistemas transaccionales o "Enterprise Resource Planning" en inglés (ERP) [Ball, 99], estas tareas empiezan a tener partes automatizadas permitiendo mejorar los procesos y lo más importante, almacenar datos que son la materia prima para obtener información.

Lamentablemente, la forma en que se almacenaban estos datos no permitía un fácil entorno para la obtención de información apareciendo los trastornos en la mayoría de las empresas como la redundancia de información

en diferentes áreas, dificultad de concentrar información de diferentes fuentes con sus diferentes formatos, la falta de habilidad para poder realizar análisis con respecto al tiempo y por consiguiente los problemas de administración de la información.

El concepto de sistemas de soporte a la toma de decisiones es un tema que surge a raíz de los problemas antes mencionados tratando de solucionar éstos en la manera en que sea posible dando pie al nacimiento del concepto de "Data Warehouse" (DW) o almacén de datos el cual servía como un repositorio central por medio del cual la gente tenía una integración de información que le permitía explotarla resolviendo algunos problemas de información, pero generando nuevos debido a la desinformación, la dificultad y el entendimiento de un enfoque totalmente diferente que los nuevos paradigmas tienden a traer.

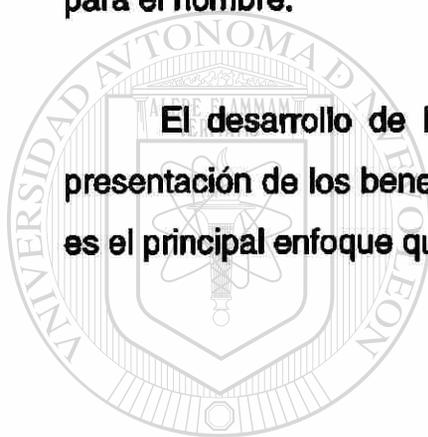
Lo que veremos en esta tesis es un ataque directo a la resolución de los problemas que se encuentran actualmente en muchas compañías (principalmente en la ciudad de México) en donde la falta de cultura hacia el enfoque de soporte a la toma de decisiones tiende a mermar la capacidad de las empresas a responder a una competencia extranjera o interna en donde aún y cuando no son perfectos, tienden a ser un buen contrincante al cual es necesario enfrentar con la mejor arma que existe actualmente "La información".

El enfoque de repositorio central evoluciona, dando paso a un nuevo concepto aceptado por muchos autores e implementado en pocos lugares, este concepto es el de la fábrica de información corporativa (FIC) o en inglés "corporate information factory". Esta tesis obtiene su principal fundamento en el entendimiento de este concepto en donde el repositorio central se complementa en una serie de arquitecturas las cuales tienen como punto central el concepto de DW mejorado, pero incluyendo nuevos componentes que permiten proporcionar al usuario diferentes medios de obtener información,

los cuales una vez implementados convierten a los usuarios explotadores de información en “agentes de cambio”.

Un medio que a valido a diferentes personas para atacar un problema complejo ha sido el relacionar este problema a una actividad natural humana dando pie a conceptos como redes neuronales, algoritmos genéticos, etcétera. En el caso de lo que se plantea con la FIC es la relación a un ecosistema natural en donde un elemento apoya al otro para poder subsistir o ser mejor, además relaciona la FIC como una fábrica verdadera en donde entra de alguna forma la materia prima (los datos) y salen ya procesados en elementos útiles para el hombre.

El desarrollo de la metodología para lograr una FIC, su diseño y la presentación de los beneficios aplicados a la empresa, en este caso GAMESA es el principal enfoque que se aplica a la presente tesis.



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

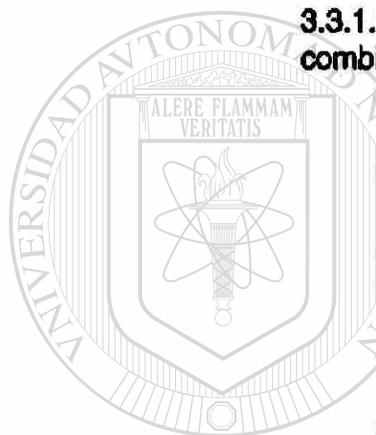


DIRECCIÓN GENERAL DE BIBLIOTECAS

ÍNDICE

	Página
SÍNTESIS	1
Capítulo 1. INTRODUCCIÓN	5
1.1 Descripción del problema.....	5
1.2 Objetivos de la tesis.....	7
1.3 Hipótesis.....	8
1.4 Límites del estudio.....	9
1.5 Justificación de la tesis.....	9
1.6 Metodología.....	11
1.7 Revisión bibliográfica.....	11
Capítulo 2. MARCO TEÓRICO	13
2.1 Generalidades.....	13
2.1.1 El enfoque "Data Warehouse".....	14
2.1.2 Enfoque actual "El ecosistema de información".....	17
2.2 Fábrica de información corporativa (FIC).....	17
2.2.1 Componentes.....	18
2.2.2 Usuarios.....	24
2.2.3 Arquitecturas.....	26
2.3 Metodologías de desarrollo de fábricas de información corporativas.....	30
2.3.1 Método propuesto por Bill Inmon.....	31
2.3.1.1 El plan estratégico.....	31
2.3.1.2 La acción estratégica.....	32
2.3.1.3 Ciclos de vida de desarrollo.....	33
2.3.1.4 Desarrollo de las bases de datos.....	34
2.3.1.5 Despliegue de información.....	35
2.3.2 Método propuesto por Ralph Kimball.....	35
2.3.2.1 Administración del proyecto y requerimientos.....	35
2.3.2.1.1 El ciclo de vida dimensional del negocio....	36
2.3.2.1.2 Administración del proyecto.....	38
2.3.2.1.3 Recolección de requerimientos.....	38
2.3.2.2 Diseño de datos.....	39
2.3.2.3 Selección de la arquitectura.....	40
2.3.2.3.1 Arquitectura de Adquisición de datos.....	42
2.3.2.3.2 Arquitectura de presentación de datos.....	44
2.3.2.4 Metadatos.....	46
2.3.2.5 Implementación.....	47
2.3.2.5.1 Complementando el diseño físico.....	47
2.3.2.5.2 El área de procesamiento.....	48
2.3.2.6 La construcción de las aplicaciones de los usuarios finales.....	48
2.4 Problemas encontrados usualmente en un DW y sus posibles soluciones.....	48
2.5 Herramientas para alcanzar la creación de una FIC.....	50
2.5.1 Herramientas de adquisición.....	50

2.5.1.1 Herramientas ETL...te...logía.....	51
2.5.1.2 Herramientas Modeladoras.....	52
2.5.2 Herramientas de almacenamiento.....	52
2.5.3 Productos de acceso.....e.....	59
2.6 El por qué una arquitectura FIC puede ayudar al problema de descentralización de información, inflexibilidad y optimización de un Data Warehouse.....f.....n.....	54
Capítulo 3. MÉTODO PROPUESTO.....	55
3.1 Metodología de desarrollo propuesta.....	55
3.2 Definición de herramientas a utilizar.....	57
3.2.1 Herramienta de Administración de Proyectos.....	58
3.2.2 Herramienta ETL.....	58
3.2.3 Herramienta modeladora.....	60
3.2.4 Herramienta de acceso.....	60
3.3 Formas de atacar objetivos de la tesis.....	61
3.3.1 Mejoramiento en proceso de desarrollo.....	61
3.3.1.1 Análisis de estatus actual.....	62
3.3.1.2 Utilización del Ciclo de Vida de Kimball en combinación con elementos de la metodología de Inmon.....	63
3.3.1.2.1 Plan de proyecto.....	63
3.3.1.2.2 Obtención de requerimientos.....	66
3.3.1.2.3 Diseño de modelo general de FIC propuesto.....	67
3.3.1.2.4 Construcción de matriz de bus.....	67
3.3.1.2.5 Realización del modelado dimensional.....	68
3.3.1.2.6 Identificación de las fuentes de cada tabla de hechos y dimensiones.....	71
3.3.1.2.7 Mapeo fuente a destinos.....	73
3.3.1.2.8 Desarrollo por medio de herramienta modeladora de modelo físico a un manejador de base de datos.....	73
3.3.1.2.9 Dimensionamiento.....	74
3.3.1.2.10 Plan y diagrama de arquitectura tecnológica y de infraestructura.....	75
3.3.2 Mejoramiento en procesos de carga.....	76
3.3.3 Mejoramiento en validación.....	79
3.3.4 Mejoramiento en la explotación.....	80
Capítulo 4. DESARROLLO DEL MÉTODO.....	82
4.1 Análisis de estado actual.....	82
4.2 Mejoramiento del proceso de desarrollo.....	83
4.2.1 Plan de proyecto.....	84
4.2.2 Obtención de requerimientos.....	86
4.2.3 Diseño del modelo general de la FIC propuesto.....	89
4.2.4 Construcción de matriz de bus.....	90
4.2.5 Realización del modelado dimensional.....	91
4.2.6 Identificación de las fuentes de cada tabla de hechos y dimensiones.....	92
4.2.7 Mapeo fuente a destinos.....	94
4.2.8 Desarrollo por medio de herramienta modeladora de modelo físico a un manejador de Base de datos.....	94
4.2.9 Dimensionamiento.....	95



4.2.10 Diagrama de arquitectura tecnológica.....	96
4.3 Mejoramiento de los procesos de carga.....	100
4.4 Mejoramiento en procesos de validación.....	101
4.4.1 Bitácora de proceso general.....	101
4.4.2 Bitácora de datos.....	103
4.4.3 Bitácora de procesos individuales.....	104
4.5 Mejoramiento en la explotación.....	106
Capítulo 5. RESULTADOS.....	107
5.1 Mejoramiento en proceso de desarrollo.....	107
5.2 Mejoramiento en procesos de carga.....	113
5.3 Validación.....	114
5.4 Explotación.....	114
Capítulo 6. CONCLUSIONES Y RECOMENDACIONES.....	115
6.1 Conclusiones.....	115
6.1.1 Para el mejoramiento en el proceso de desarrollo.....	115
6.1.2 Para el mejoramiento en el proceso de carga.....	117
6.1.3 Para el mejoramiento en el proceso de validación.....	119
6.1.4 Para el mejoramiento en el proceso de explotación.....	121
6.1.5 Conclusiones en general.....	122
6.2 Recomendaciones.....	123
BIBLIOGRAFIA.....	126
LISTA DE TABLAS.....	128
LISTA DE FIGURAS.....	129
LISTA DE ABREVIATURAS.....	130
ANEXO A Formatos.....	131
ANEXO A1 Plan de proyecto.....	132
ANEXO A2 Obtención de requerimientos.....	133
ANEXO A3 Matriz de indicadores.....	136
ANEXO A4 Modelado dimensional.....	137
ANEXO A5 Hechos derivados.....	141
ANEXO A6 Definición de fuentes.....	142
ANEXO A7 Mapeo campo a campo.....	143
ANEXO A8 Dimensionamiento.....	144
ANEXO A9 Mapa de configuración.....	145
ANEXO B Modelado dimensional aplicado.....	146
ANEXO C Hechos derivados.....	159
ANEXO D Modelo lógico - físico.....	161
ANEXO E Mapa de configuración aplicado.....	163
ANEXO F Glosario.....	164
RESUMEN AUTOBIOGRÁFICO.....	168

SINTESIS

La presente tesis tiene como objetivo fundamental expresar una metodología formal de desarrollo de una solución para el soporte a la toma de decisiones enfocada al diseño de una Fábrica de Información Corporativa (FIC), la cual aplicada permitirá analizar los diferentes procesos implicados en la fábrica y mostrar el grado de cumplimiento de los objetivos de mejoramiento en los mismos.

La presentación realizada se basa en los conceptos encontrados en diversas literaturas las cuales en combinación forman la base para realizar de una manera óptima la creación de una FIC.

En el capítulo 2 tomo todos los conocimientos y los aterrizo a la propuesta de diseño la cual básicamente se respalda en dos metodologías de las cuales una de ellas establece el concepto de fábrica de información como arquitectura definiendo los conceptos y elementos principales que conforman ésta, mientras que la otra metodología está encaminada en hacer realidad todo el concepto de FIC con la particularidad que se enfoca en el cómo debe de hacerse proporcionando herramientas para llegar al objetivo final.

La metodología propuesta en el capítulo 3 y el diseño desarrollado en el capítulo 4 proporciona las bases para lograr cumplir con los objetivos de esta tesis abordando los temas de disminución de la cantidad de tiempo y esfuerzo en un desarrollo, la aplicación de herramientas que en conjunto con el cambio de arquitectura permitirán mejorar los procesos de carga o de información a la fábrica, la propuesta de medios alternos para manejar la validación de información, así como finalmente el mejoramiento de la explotación de información por medio del aprovechamiento de la arquitectura propuesta

permitiendo al usuario realizar análisis inteligentes, enfrentando con éstos los típicos problemas experimentados por la antigua arquitectura.

Y como en todo desarrollo, o como en cualquier construcción, es necesario entender los medios por los cuales es posible llegar al resultado final, por lo que planteo además las diferentes herramientas que son útiles para este fin.

Los puntos que ataco como parte de la tesis son cuatro:

1) Mejoramiento de los procesos de desarrollo

Para esto la tesis abarca todos los puntos de un ciclo de vida de desarrollo de sistemas (CVDS) compuesto por el análisis, diseño, construcción, pruebas e implementación, con la particularidad que estos se basan en los conceptos que trabajan en el ámbito de los sistemas de soporte a la toma de decisiones.

De las recopilaciones realizadas se desarrollará la metodología abarcando ésta para la preparación del "Back End" (todo lo relacionado con la construcción, alimentación y validación) y "Front End" (la explotación) de una fábrica de información. Esta metodología se explica empleando el método de Ralph Kimball y su grupo [Kimball y Reeves y Ross, 98] los cuales inician con el plan de proyecto específico para poder administrar por fases. Cabe mencionar que queda fuera de esta tesis la justificación y el análisis de factibilidad por el hecho de partir de un desarrollo existente, pero con una arquitectura no óptima. A partir de aquí se hacen propuestas de formatos para atacar cada uno de los puntos importantes en un proyecto de esta índole.

El primer punto a atacar es el análisis en el cual se plantean las actividades de planeación y la recopilación de los requerimientos y el

entendimiento de los mismos. La siguiente parte define el diseño en el cual todo inicia a partir de los datos recopilados facilitando enormemente el modelado que es la siguiente parte. El modelado que se plantea es el denominado modelado dimensional y nos es útil debido a la arquitectura que permite manejar propia de una solución de soporte a la toma de decisiones. Dentro de esta parte se definen formatos específicos para llevar a cabo este modelado definiendo las tablas tanto de hechos como de dimensiones, además de presentar un modelo conocido como matriz de bus que permite entender el panorama general de toda la fábrica de información y que permitirá en futuras iteraciones incorporar nuevos conceptos sin tener que realizar grandes cambios. Otro punto dentro de este modelado se centra en la especificación de los hechos derivados.

Una vez que se ha creado el modelo, se plantea la forma de llevarlo a cabo físicamente por medio de la creación de un documento en una herramienta de modelado especial que permitirá la creación de este modelo a la base de datos seleccionada.

Dentro de los puntos además incluidos se tratan las partes de la construcción, pruebas e implementación, en cada uno de ellos se hace una referencia como parte de la metodología para poder llevar a cabo todo el proyecto.

La forma de medir el grado de cumplimiento se basa en la recopilación de tiempos del desarrollo del anterior proyecto y la comparación proporcional en tiempo con la nueva arquitectura de FIC.

2) Mejoramiento de los procesos de carga

Este mejoramiento lo planteo como un agregado utilizando la nueva arquitectura y planteando que por medio de una herramienta especializada se realicen los desarrollos los cuales aún y cuando no es el fin atacarlos en la

tesis, si es parte de la metodología proporcionar una estrategia de carga por medio de la correcta administración de procesos los cuales se manejan en un mapa de configuración a ser explicado en el capítulo correspondiente.

La forma de medir el grado de cumplimiento se basa en la recopilación de los tiempos de carga del proceso con la arquitectura anterior y compararlo con la nueva arquitectura para obtener los resultados y tomar las conclusiones correspondientes.

3) Mejoramiento de los procesos de validación

Este proceso, aún y cuando es básico no estaba muy desarrollado anteriormente por lo que se plantean estrategias para poder llevar a cabo una correcta validación de información presentando recopilaciones de los errores encontrados en la anterior arquitectura y comparándolo con los encontrados en la nueva.

4) Mejoramiento de los procesos de extracción

Para este punto se plantea un cambio de paradigma enfocado a proporcionar una sola herramienta especializada, pero que permita la explotación de información directa y que elimine los tiempos de triangulación de información. Esto se verá reflejado en tiempos que se recopilarán de lo desarrollado actualmente y lo nuevo que será creado. Además doy un vistazo a los diferentes problemas que detectan diversos autores para complementar los ya existentes actualmente en la compañía GAMESA permitiendo un mejor entendimiento del porqué la necesidad de una mejor arquitectura para el soporte a la toma de decisiones.

Por último en los capítulos 5 y 6 presento los resultados de estas mediciones realizando las conclusiones y los comentarios adicionales.

1 INTRODUCCIÓN

1.1 Descripción del problema

En la compañía GAMESA S. de R.L. de C.V., empresa dedicada a la manufactura de galletas, se vive un problema relacionado con el manejo de su información en todos los niveles, principalmente en el área administrativa, en donde se toman generalmente las principales decisiones que rigen el curso de la compañía.

El problema se originó a partir de una implementación muy primitiva basada en uno de los modelos que en los 90's estuvo de moda, el "Almacén de datos" o "Data Warehouse" como repositorio de información, modelo que se trató de implementar pero sin llegar a solucionar los problemas inherentes con el manejo de información existentes en esa época (y que existen actualmente), esto quizás debido a que este concepto era muy nuevo y por lo mismo no existía gente con la habilidad y el conocimiento necesario sobre estos temas y sus metodologías. Los problemas que trataban de solucionar en su momento y que actualmente persisten son básicamente relacionados con la falta de centralización de la información para su posterior explotación, ya que esto provoca que no exista un mismo lugar para obtener la información y por consiguiente que no se obtuviera un mismo resultado en todas las áreas de la compañía sobre los mismos datos lo que provocó ambigüedades. Otros problemas son la falta de seguridad en la información, su falta de homogeneidad, su dificultad para explotarla haciendo la toma de decisiones mucho más complicada y sus múltiples orígenes.

La semi-implementación de este concepto trajo otros problemas graves entre los cuales se observan:

- **La falta de centralización de la información.**
 - **La falta de capacidad para satisfacer requerimientos de información a nivel operativo.**
 - **La inseguridad en los datos no proporcionando las herramientas necesarias para validarlos.**
 - **Tiempos de carga de información grandes.**
 - **Dependencia sobre un administrador para el control y el mantenimiento de los datos.**
 - **La ineficiencia al momento de extraer información útil tanto para la operación diaria, como para el proceso de toma de decisiones.**
 - **La falta de control sobre la creación de nuevas métricas o indicadores sobre lo ya existente que hace que sea prácticamente imposible crecer el repositorio.**
 - **La falta de estandarización de criterios sobre la explotación de la información.**
-
- **La falta de credibilidad de los usuarios sobre la información. Esto más que nada debido al desconocimiento sobre la forma en que se encuentra almacenada la información.**
 - **Mantenimientos caros y relativamente difíciles (Debido a las grandes cantidades de código de programación utilizado).**
 - **Cambios casi imposibles de realizar debido a la inexistencia de una documentación sólida del Almacén de datos.**

Todos estos problemas tienen una principal consecuencia:

El desinterés de los usuarios por realizar análisis inteligentes debido a que no tienen ni las herramientas necesarias ni la confianza en la información para poderlos llevar a cabo.

1.2 Objetivo de la tesis

Mi objetivo general con la presente tesis consiste en atacar los problemas mencionados con anterioridad por medio de una buena metodología, buenas herramientas especializadas y una buena arquitectura para demostrar que se pueden mejorar los problemas mencionados, desde los puntos relacionados con el desempeño de las cargas de información, la seguridad y el mantenimiento, hasta la capacidad de proporcionar la información de una forma fácil, rápida e inteligente que les permita realizar análisis en menos tiempo, abriendo la posibilidad de que lo hagan de una manera óptima.

El objetivo específico a alcanzar en la empresa GAMESA es diseñar una arquitectura de almacén de información centralizado, robusto, flexible, confiable, fácil de explotar y que proporcione las herramientas para una toma de decisiones óptima para el buen desempeño de la compañía.

Para lograr lo que he mencionado utilizaré un modelo nuevo denominado Fábrica de Información Corporativa (FIC) el cual complementaré con experiencias propias.

Esto permitirá además satisfacer los siguientes objetivos secundarios:

- Estandarización de documentación
- Rapidez de desarrollo de nuevas iteraciones en la fábrica de información,
- Mejoramiento en la calidad de la información.
- Organización en la forma de crear nuevos conceptos de negocio dentro de la fábrica de información.
- Disminución de tiempos de desarrollo.
- Mantenimientos mínimos sobre las aplicaciones de la fábrica de información.
- Mejoras en los tiempos de explotación de información para la toma de decisiones.
- Lograr un entendimiento de los usuarios de todo el potencial que se tiene dentro de la fábrica de información para su trabajo de toma de decisiones.

1.3 Hipótesis

H1:

Utilizando una metodología formal y un buen diseño de la arquitectura en creación de una fábrica de información corporativa se pretende disminuir los tiempos de desarrollo.

H2:

Por medio de la creación de una fábrica de información, en la cual van inmersas una serie de metodologías formales y una serie de herramientas especializadas se pretenden mejorar los tiempos de carga, de validación y de explotación permitiendo una toma de decisiones más sencilla y con mayor valor agregado en la compañía.

1.4 Límites del estudio

La limitante principal que planteo al realizar la tesis radica en que solo aplicaré los conceptos de fábrica de información sobre una de las áreas pilar en la compañía GAMESA, que es el área de Ventas. Esto debido a que el proceso de creación de una fábrica de información es un proceso incremental en el cual se debe de iniciar por una pequeña parte del negocio e ir aumentándola conforme al crecimiento de los requerimientos de información de otras áreas.

Además la intención de la tesis no es proporcionar un manual para el manejo de las herramientas utilizadas para crear la fábrica de información sino que solo se mencionarán como parte de la estrategia y se dará una breve explicación de éstas en su aplicación en el proyecto y como mejoran el desempeño de las cosas.

En la parte de explotación de información no es objetivo de la tesis mejorar el diseño de los reportes, sino proponer una estrategia para el mejoramiento de esta explotación independientemente de los análisis que se realicen.

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

DIRECCIÓN GENERAL DE BIBLIOTECAS

1.5 Justificación del trabajo de tesis

El presente trabajo de tesis es conveniente realizarlo debido a que ataca uno de los principales problemas que existen actualmente en la compañía, y que posiblemente se encuentre en muchas compañías en México, que está relacionado con su manejo de información para la toma de decisiones. Problema que es importante abordar ya que las compañías en la época actual dependen grandemente de la información debido a que esta es la materia prima que lleva a la compañía a alcanzar sus objetivos sirviendo de brújula para lograrlos.

***La tesis tendrá una contribución de aportar conocimiento debido a que abordaré temas que servirán de base para que la gente que no conoce o que actualmente tenga implementado un Data Warehouse observe las diferentes alternativas que existen sobre este modelo y además puedan observar como puede o no llegar a funcionar en una compañía, teniendo como base a una tan importante como lo es GAMESA.**

Esta tesis y por consiguiente este diseño tiene una fuerte implicación práctica ya que ayudará a resolver una serie de problemas que es posible que muchas compañías estén sufriendo y lo más importante es que lo verán aplicado.

Por último, la información recopilada tanto de metodología como de mediciones y pruebas, proporcionará a las gentes interesadas de conocimientos en el tema una base sólida para futuras investigaciones de optimización de procesos de desarrollos de proyectos de Fábricas de información y de explotación de información; así como servirá de guía para las personas que no tienen el conocimiento sobre este tema y que quiere intentar resolver los problemas relacionados con el manejo de su información y tengan una base.

UNIVERSIDAD AUTONOMA DE NUEVO LEON

®

DIRECCIÓN GENERAL DE BIBLIOTECAS

1.6 Metodología

Un resumen de la propuesta de la metodología para la elaboración de esta tesis se presenta a continuación:

- Definición del plan del proyecto con tiempos y entregables.
- Elaboración de un marco teórico
 - DetECCIÓN de la literatura
 - Obtención y consulta de la literatura
 - Revisión de literatura
 - Extracción y recopilación de la información
 - Organización de la información y creación del marco teórico.
- Selección de las metodologías a aplicar.
- Diseño de formatos a utilizar.
- Obtención de los requerimientos del usuario.
- Aplicación de la metodología y medición de resultados (Tiempos).
- Obtención de resultados para medir en que forma se soluciona el problema de la carga, validación y explotación en el manejo de información de la empresa.
- Redacción del reporte de investigación.
- Redacción de las conclusiones.
- Propuesta de trabajos futuros sobre la base de la experiencia adquirida.

1.7 Revisión Bibliográfica

W.H. Inmon, en conjunto con Claudia Imhoff y Ryan Sousa proponen en su libro *"Corporate Information Factory"* [Inmon e Imhoff y Sousa, 01] un enfoque diferente sobre el manejo de las soluciones para el soporte a la toma de decisiones en la cual plantean el concepto de Fábrica de Información

Corporativa identificando los diferentes elementos que conforman la FIC y proporcionando una explicación de los elementos a incluir en un modelo. Los planteamientos de arquitectura y de elementos sirvieron de base para el diseño del modelo planteado en esta tesis.

Ralph Kimball, en conjunto con Laura Reeves y Margy Ross en su libro *"The Data Warehouse Lifecycle Toolkit"* [Kimball y Reeves y Ross, 98] proponen una metodología base para ser aplicada en el desarrollo de cualquier concepto implicado en una FIC pero enfocándolo más al punto de vista de estos autores. En este libro redactan la metodología del ciclo de vida de un proyecto de Data Warehouse y lo complementan con comentarios de lo que se debe y lo que no hacer en cualquier proyecto de este tipo. Esta metodología sirvió de base para proponer un estándar en GAMEESA y como propuesta para los desarrollos en esta compañía relacionados con la FIC. Como complemento se analizó el libro *"The Data Warehouse Toolkit"* de Ralph Kimball [Kimball, 96] en el cual plantea partes de esta metodología pero aplicadas a situaciones reales de negocio útiles para ayudar a modelar la arquitectura propuesta en ésta tesis.

Finalmente, comentarios utilizados para la optimización de procesos, detección de errores en desarrollos de Data Warehouse y comentarios base sobre el diseño de la arquitectura propuesta se visualizaron de autores como W.H. Inmon y Ken Rudin en su libro *"Data Warehouse Performance"* [Inmon y Rudin, 98], Michael Corey y Michael Abbey en su libro *"Oracle Data Warehousing, Guía práctica para analizar, construir e implantar con éxito un sistema Data Warehouse"* [Corey y Abbey, 97] y del instituto de Data Warehouse el reporte *"Ten Mistakes to avoid"* [TDWI, 00].

2 MARCO TEÓRICO

2.1 Generalidades

A través de la historia de desarrollo de sistemas, el énfasis primordial había sido enfocado a los sistemas operacionales y a la información que ellos procesan. Pero una de las cosas que se ha aprendido a través del tiempo es que no es práctico mantener la información operacional indefinidamente en estos sistemas, por lo que se ideó una estructura alterna diseñada para almacenar datos que el sistema operacional había procesado. Los requerimientos fundamentales [Inmon, 96] utilizados por los 2 enfoques son diferentes por lo que la forma en que se estructuran influyen grandemente para lograr el éxito en los proyectos de sistemas para el soporte a la toma de decisiones. A raíz de estos el concepto de Data Warehouse (DW) o Almacén de datos ha evolucionado y se ha convertido en una clásica aplicación única y popular.

Los constructores de DW consideran sus sistemas componentes clave de sus arquitecturas y estrategias de tecnologías de información (TI).

Los vendedores de "Hardware" y "Software" han rápidamente desarrollado productos y servicios que se enfocan específicamente al mercado del DW, servicios y productos que vienen a sustituir el enfoque anterior en el que los desarrollos de sistemas de soporte a la toma de decisiones se desarrollaban generalmente en casa y por medio de herramientas no

especializadas las cuales hacían que los desarrollos fueran pesados, laboriosos y difíciles de manejar.

Aún y cuando un Data Warehouse no es un fenómeno nuevo, las empresas no se han dado a la tarea de administrarlos bien por lo que en cierta forma no se está preparado para el futuro en el que el crecimiento del concepto de sistemas de soporte a la toma de decisiones va a ser enorme con nuevos productos y tecnologías por lo que es importante que los desarrolladores y los administradores tengan una idea clara de lo que están buscando y que seleccionen las estrategias y los métodos que los proveerán hoy de un buen desempeño y mañana la flexibilidad necesaria. [Orr, 96].

2.1.1 El enfoque “Data Warehouse”

Existen Diversas definiciones de lo que es un Data Warehouse, algunas de ellas son:

- Un Data Warehouse en su más simple percepción no es más que una colección de las piezas llave de información usadas para administrar y dirigir el negocio para obtener el ingreso más rentable [Anahory y Murray, 97]. Pero ellos van más allá definiéndolo como el proceso que se requiere para obtener la información de una fuente a una tabla y obtener los datos de la tabla para los analistas.
- Un Data Warehouse es una colección de información corporativa derivada directamente de los sistemas operacionales y de algunos orígenes de datos externos en donde su propósito específico es soportar la toma de decisiones en un negocio, no las operaciones del negocio. [Corey y Abbey, 97, pag. 1]

- **Un Data Warehouse es una arquitectura estructurada que soporta la administración de los datos y que es orientada a individuos, integrada, variante con el tiempo, no volátil e incluye datos agrupados y detallados. Este existe para soportar la administración de las decisiones las cuales soportan los procesos de planeación estratégica de la corporación. [Inmon e Imhoff y Sousa, 01, pag. 93]**

A saber el enfoque original visto desde el punto de vista del Ralph Kimball y su grupo [Kimball y Reeves y Ross, 98, pag. 9-11] un Data Warehouse debe de cumplir con los siguientes objetivos:

- **Hacer que la información de la organización sea accesible con conceptos entendibles y navegables y con una acceso de alto desempeño.**
- **Hacer que la información de la organización sea consistente obligando a que si dos o más departamentos de la misma compañía hablan del mismo concepto, entonces las medidas o nomenclaturas de este sea la misma ya que si la información es distinta estamos proporcionando datos inconsistentes con sus futuras repercusiones entre la gente por la falta de credibilidad.**
- **Permitir ser una fuente de información adaptativa, es decir deberá estar diseñada para los cambios continuos.**
- **Debe ser un lugar que proteja el conjunto de información controlando no sólo los accesos a los datos de una manera efectiva, sino permitiendo a los usuarios tener una gran visibilidad en los usos y los abusos de estos datos aún y cuando hayan salido del Data Warehouse.**
- **Es el fundamento para la toma de decisiones.**

Como toda novedad en el mercado, y con las ansias de las empresas por desarrollar y estar a la par con las novedades tecnológicas, surge la necesidad de empezar a crear el concepto de Data Warehouse con una idea utópica de que desarrollando un almacén centralizado en el cual se tendría almacenada la

información por medio de programas sin especializar (hechos en casa o con lenguajes de programación de 3era y 4ta generación) y sobre una arquitectura perfecta ésta amoldaría a todas las situaciones posibles en un negocio. Idea con la cual podrían solventar los problemas comunes de toda empresa relacionados con el manejo, distribución y credibilidad de información que los aquejan debido a la gran diversidad de fuentes y segmentaciones en los negocios. Pero esto no fue posible.

A su vez que se pudieron arreglar algunos problemas originales existentes en el manejo de información, otros salieron a flote, problemas que van desde la dificultad de implementación de nuevos desarrollos debido a la complejidad que generalmente estos presentan, la falta de metodología, la cimentación sobre una arquitectura inflexible, problemas con la carga de información debido a lo inflexible que puede ser la programación por medio de código el cual requiere un buen nivel de conocimiento sobre estos por parte de la gente, así como un aumento en la falta de credibilidad de los usuarios debido a los problemas que generalmente se enfrentan en el mantenimiento lo que originó que el concepto como tal de Data Warehouse se viera forzado a evolucionar.

“Desarrollar un Data Warehouse es como pintar una casa con pincel, no es imposible, pero es una tarea muy lenta y desalentadora, que, cuando se termina, es necesario volver a empezar de nuevo” [Corey y Abbey, 97]

2.1.2 Enfoque actual “El ecosistema de información”

Un ecosistema de información es un sistema con componentes diferentes, cada uno sirviendo a una comunidad directamente mientras se trabaja en consenso con los otros componentes para producir un ambiente cohesivo y balanceado de información [Inmon e Imhoff y Sousa, 01, pag. 2].

Para fines prácticos del presente trabajo se presenta esta definición para explicar las características que se buscan dentro un desarrollo enfocado al área de soporte a la toma de decisiones ya que como los ecosistemas naturales, un ecosistema de información presenta características similares necesarias para un buen desempeño como lo son:

- Debe de ser adaptable
- Cambiante a la misma velocidad que sus habitantes y participantes
- Balanceados.

Cada negocio requiere que un conjunto de componentes diferentes trabajen a la par logrando que el ecosistema de información se oriente a los negocios teniendo como resultado un ambiente de información que permita a las compañías crecer en un panorama en constante cambio caracterizado por relaciones con los clientes y entrega de productos acoplada a sus necesidades.

2.2 Fábrica de información corporativa (FIC)

Este concepto aún y cuando fue introducido por W.H. Inmon en los inicios de los años 80s [Inmon e Imhoff y Sousa, 01, pag. 7] marcaba la pauta para describir el panorama físico de la noción de un ecosistema de información.

Para entender la fábrica de información corporativa es necesario visualizarla en términos de los datos que fluyen hacia la fábrica y la información que fluye fuera de ella.

La definición es la siguiente:

“Una fábrica de Información corporativa es la arquitectura física de lo que representa el ecosistema de información” [Inmon e Imhoff y Sousa, 01, pag. 7]

2.2.1 Componentes

A saber una FIC puede tener una combinación de componentes para sustentar este ecosistema de información. A saber:

- El Mundo Externo
- Aplicaciones

- Almacén de datos operativos (ADO) o "Operational Data Store" (ODS)
Nota: para cuestiones de facilidad en lo que resta de la tesis se utilizará la sigla ODS.
- Capa de integración y transformación
- Data Warehouse (DW)
- Data Mart (DM)
- Internet/Intranet
- Metadata
- Almacén para la exploración y la minería de datos
- Almacenamiento alternativo
- Sistemas de soporte a la toma de decisiones

En la figura 2.1 se representa la FIC y cada uno de sus componentes.

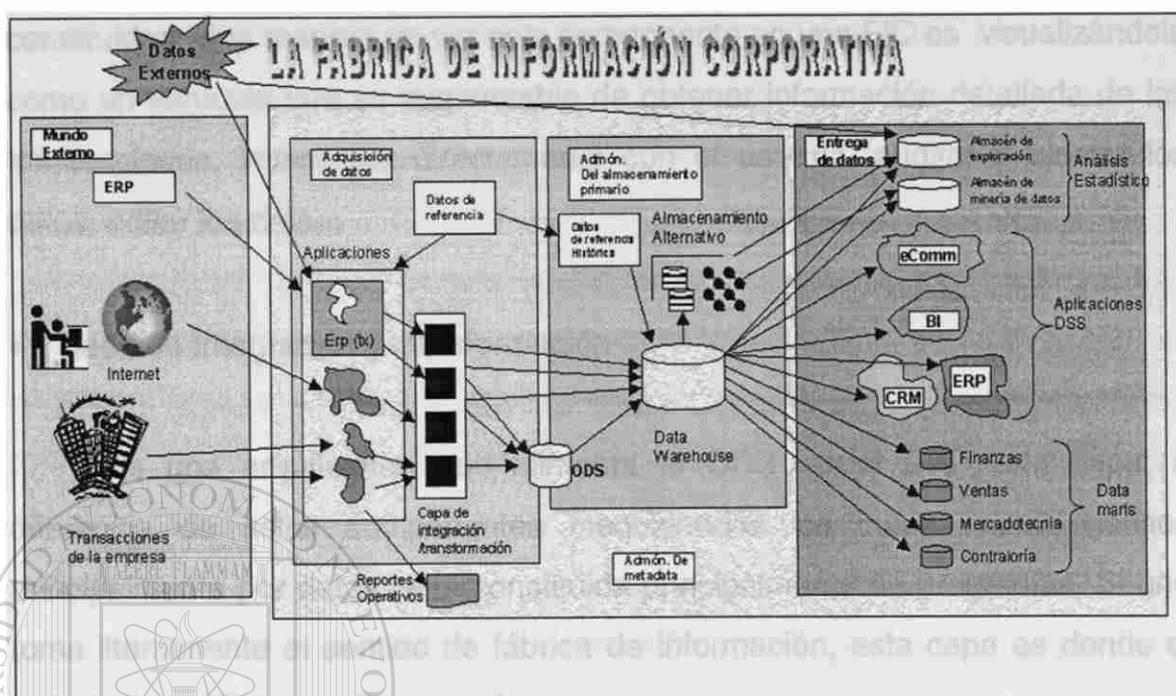


Figura 2.1 La fábrica de Información corporativa

A continuación se presenta una descripción de cada uno de estos componentes.

• Mundo Externo

Es el alfa y omega de la fábrica de información corporativa en el cual el negocio es llevado a cabo. Sin un mundo externo de comercio no existiría una FIC. Este es el componente que alimenta el combustible y consume la información que es producida por la FIC.

Uno de los focos del mundo externo es la generación de transacciones de negocio. Una FIC debe de tomar en cuenta todas las facetas de una transacción y más.

- **Aplicaciones**

Las aplicaciones han existido desde que los primeros sistemas fueron construidos. Una manera de ver este componente en una FIC es visualizándola como un vehículo que es responsable de obtener información detallada de las transacciones, interactuar directamente con el usuario, auditar y ajustar los datos, editar los datos.

- **Capa de Integración y transformación**

Es una arquitectura que alimenta al ODS o al DW, esta capa, a diferencia de estos componentes mencionados los cuales están hechos principalmente por datos, está constituida principalmente de programas. Si uno toma literalmente el sentido de fábrica de información, esta capa es donde el trabajo de la fábrica se lleva a cabo.

La interfaz de integración y transformación es el lugar donde los datos que no están integrados son combinados y transformados en los datos corporativos.

Esta capa es fundamental para la optimización de una FIC debido a que los programas de esta capa son los que realizan las funciones de lectura de datos, transformaciones, mapeo al lugar correcto del ODS o del DW, etcétera.

Además se realizan actividades de integración como: Reformato de información, conversiones matemáticas, reordenamiento de datos, asignación de valores por constantes, el manejo de múltiples fuentes de datos, sumarización y agregación de datos.

- Almacén de datos operativos (ODS)

Un ADO (ODS en inglés por Operational Data Store) es una estructura híbrida que tiene los elementos igualmente fuertes para tener procesamiento operacional y procesamiento de soporte a la toma de decisiones dualidad que lo hace uno de las estructuras más complejas de la FIC.

Un ODS es una colección de datos detallados que satisfacen las necesidades de colectividad, integración y operatividad de una corporación. Es orientado al usuario, integrado, pero volátil con valores actuales y detallado. Comparado con un DW el ODS es muy parecido, pero únicamente en los conceptos de orientación a usuario y de que es integrado.

El ODS es alimentado directamente por la capa de integración y transformación, pero éste alimenta al DW.

Por último existen cuatro clases de ODS:

- ↳ **La Clase 1:** Asíncrona (Con 1 o 2 segundos de retraso)
- ↳ **La Clase 2:** Almacena envía (Con 2 a 4 horas de retraso)
- **La Clase 3:** Procesamiento en modo "Batch" (Nocturno)
- ↳ **La Clase 4:** Datos desde el DW.

- Data Warehouse

Esta es la componente con arquitectura más prominente de la FIC. Esta es la base para todo el procesamiento estratégico del soporte a la toma de decisiones.

El Data Warehouse es una estructura que soporta la administración de los datos y se caracteriza por ser orientada a sujetos alineada a las mayores

entidades de una organización, integrada debido a la unificación física de los datos, variante en el tiempo ya que cualquier registro en el DW es relacionado a algún momento en el tiempo, no volátil ya que no cambia constantemente los datos bajo condiciones normales, y finalmente con una combinación de datos tanto integrados como detallados. Los datos provienen tanto del ODS como de la capa de Integración y Transformación.

Cuando se compara el Data Warehouse con los demás componentes, éste es muy grande principalmente debido a los grandes manejos de datos que se tienen. Por esto que este componente tiene su mayor desafío en la parte de manejo de datos, tanto en la capa de Integración y transformación (lyT) como en la parte de la explotación.

- **Data Mart**

Un Data Mart (DM) es una colección de datos hechos a la medida de las necesidades de procesamiento para soporte a la toma de decisiones que ha sido configurada para satisfacer los requerimientos de “un departamento en específico”. Como característica principal un Data Mart contiene una pequeña cantidad de datos detallados y una porción generosa de datos resumidos; con una cantidad limitada de historia y además está preparada para solo satisfacer las necesidades del departamento a quién pertenece la herramienta seleccionada para su uso.

Existen dos clases de Data Marts:

- 1) *Los Data Marts tipo MOLAP (Multidimensional on line analytical processing)*, en donde los datos se almacenan en una forma muy estructurada y con dimensiones(Productos, tiempo, localidades) de los datos creados.

2) *Los Data Marts tipo ROLAP (Relational on line analytical processing)*, en donde el procesamiento es más general que en el MOLAP y su objetivo es proveedor de una vista multidimensional de los datos utilizando una técnica de manejo de bases de datos relacionales.

- **Almacén para la exploración y la minería de datos**

El almacén de exploración es un componente aislado del Warehouse para permitir realizar procesamientos analíticos estadísticos sin necesidad de alterar el procesamiento de las consultas del día a día.

- **Almacenamiento alternativo**

Este componente no es más que un medio que nos permite almacenar grandes cantidades de detalle e historia de datos para mantener un Warehouse eficiente y con bajo costo. Este componente surgió debido a la dificultad que se presentó con las grandes cantidades de datos que se almacenaban y al costo de la infraestructura para mantenerla. La solución para los problemas anteriormente mencionados es que se remueva la información inactiva o con muy poca actividad a medios alternos de almacenamiento con bajo costo los cuales pueden ser de dos tipos:

- *Almacenamiento secundario.*- Este medio de almacenamiento es más lento y menos caro que el almacenamiento a disco normal.
- *Almacenamiento casi en línea.*- Este medio de almacenamiento está constituido por cintas.

- **Internet/Intranet**

Esta es la línea de comunicación entre la cual los datos fluyen y los diferentes componentes de una FIC interactúan con cada uno.

El propósito de esta fábrica de comunicaciones es el de transportar datos, procesamiento distribuido, calendarización y coordinación de actividades, entregar estatus dentro de la arquitectura, proveer conectividad y exponer las capacidades de la FIC al mundo externo.

- **Metadata**

El concepto de metadata es sin duda uno de los más importantes de la FIC. Este se define como datos sobre los datos o como todo sobre los datos necesarios para promover su administración y uso.

El porqué la metadata es tan importante se explica debido a que ésta es el pegamento que mantiene la arquitectura de la FIC unida.

- **Sistemas de soporte a la toma de decisiones**

Básicamente este componente se enfoca a cumplir con el objetivo final de una FIC el cual obtener la correcta información en las manos de la gente correcta cuando lo requieran.

2.2.2 Usuarios

Un punto importante a entender en el mundo de los sistemas de soporte a la toma de decisiones es quienes son las personas que lo utilizan y sus

diferentes actividades para con él. Los diferentes usuarios pueden ser de dos tipos:

- **Usuarios de aplicación:** Estos usuarios son primordialmente trabajadores o profesionales de ventas y servicios y en algunos casos los clientes de la compañía. La preocupación principal de este tipo de usuarios está ligada con decisiones inmediatas y directas. Ellos generalmente manejan pequeñas cantidades de información y con una necesidad de un tiempo de respuesta de 2 o 3 segundos.
- **Usuarios de Sistemas de soporte a la toma de decisiones:** Estos usuarios son muy diferentes a los usuarios operativos. Ellos resuelven o investigan preguntas de largo plazo. Analizan sobre grandes cantidades de información pero saben de antemano que su información no será proporcionada en forma inmediata.

Dentro de esta categoría existen cuatro tipos de usuarios:

- 1) **Turistas:** Los cuales son analistas que se especializan en ser capaces de encontrar mucha información. Este entiende y conoce la estructura de la FIC para encontrar lo que desee. Pero es un analista no predecible.
- 2) **Granjero:** Un granjero es alguien que es predecible y conoce lo que él o ella quiere antes de realizar una consulta. Se caracteriza por ser repetitivo en la búsqueda de información y raramente busca datos que no le son familiares.
- 3) **Explorador:** Estos tienen características similares a los turistas y los granjeros, pero con la diferencia de que es alguien que opera con un alto grado de impredecibilidad e irregularidad buscando entre cantidades masivas de detalle.
- 4) **Minero:** Estos son muy similares a los exploradores y están muy relacionados ya que el explorador crea las afirmaciones y las hipótesis

mientras que el minero prueba la validez o invalidez de estas. El minero es un estadístico.

2.2.3 Arquitecturas

Existen diversas arquitecturas las cuales se forman con variantes de los diferentes componentes de una FIC. Una característica especial es que no existe una guía fija para la arquitectura a aplicar, pero dependiendo de los componentes seleccionados es el precio de deberemos pagar debido a las ventajas o desventajas que estaremos enfrentando con la arquitectura seleccionada. Se mencionan cuatro de las más tentadas a realizarse y más útiles:

1) Mundo Externo, Aplicaciones, Data Warehouse, Metadata

Esta es la arquitectura base sobre la cual se ha trabajado desde los inicios del concepto de Data Warehouse. Ésta está basada en los componentes básicos en los cuales se puede solventar una explotación para los sistemas de soporte a la toma de decisiones. Visto desde un punto de vista muy riguroso, esta arquitectura está formada por la lectura de los datos directamente de las aplicaciones, la utilización de programación en código para la carga, transformación y validación de datos y una estructura de almacenamiento sin optimización para contener los datos a diferentes niveles de detalle. Esta arquitectura se puede visualizar en la figura 2.2.

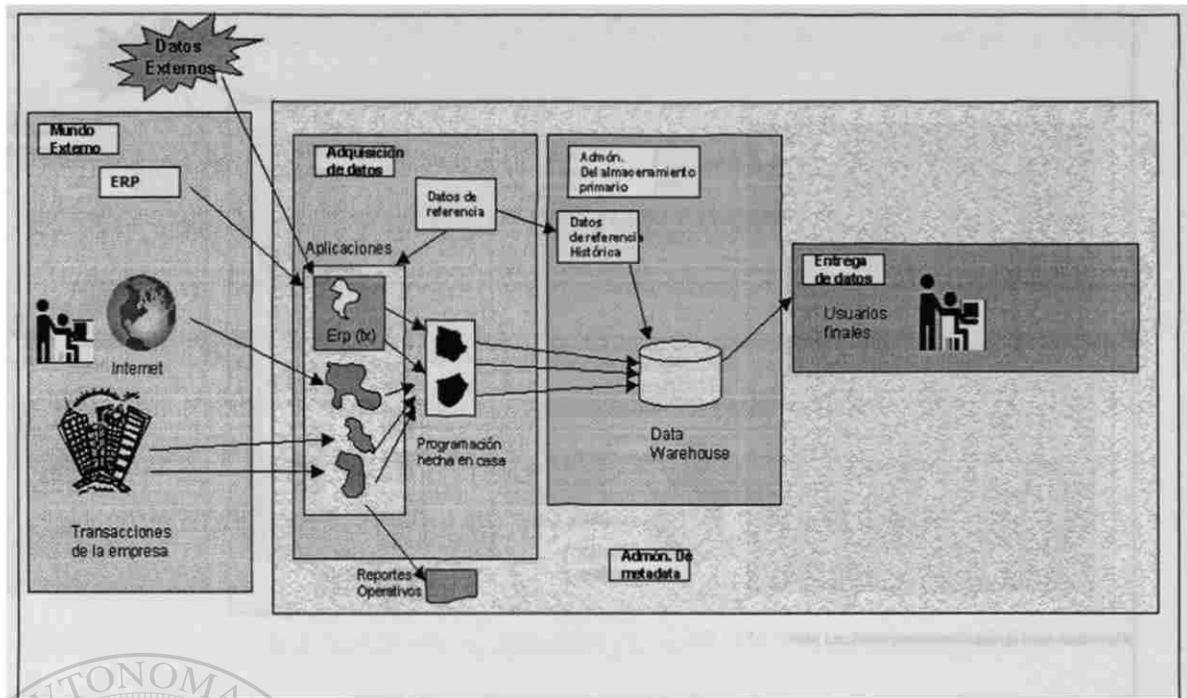


Figura 2.2 Arquitectura 1 (Mundo Externo, Aplicaciones, DW, Metadata)

2) Mundo Externo, Aplicaciones, Capa de lyT, DW, DM, Metadata

Esta es la clásica estructura para entregar inteligencia de negocios conteniendo 2 niveles de datos de interés. El DW contiene el grueso de los datos detallados que han sido coleccionados e integrados del ambiente de aplicación, mientras que el DM contiene el subconjunto de los datos a cierto nivel que ha sido preparado para el departamento.

DIRECCIÓN GENERAL DE BIBLIOTECAS

Existen varios puntos importantes a considerar en esta arquitectura a los cuales se enfrenta un arquitecto de DW y se relacionan con el orden en que se construirá el DW y el DM decisión que no parece ser muy importante pero si lo es.

- Construcción de los Data marts antes que el DW se construya

El diagrama de esta arquitectura se muestra en la figura 2.3.

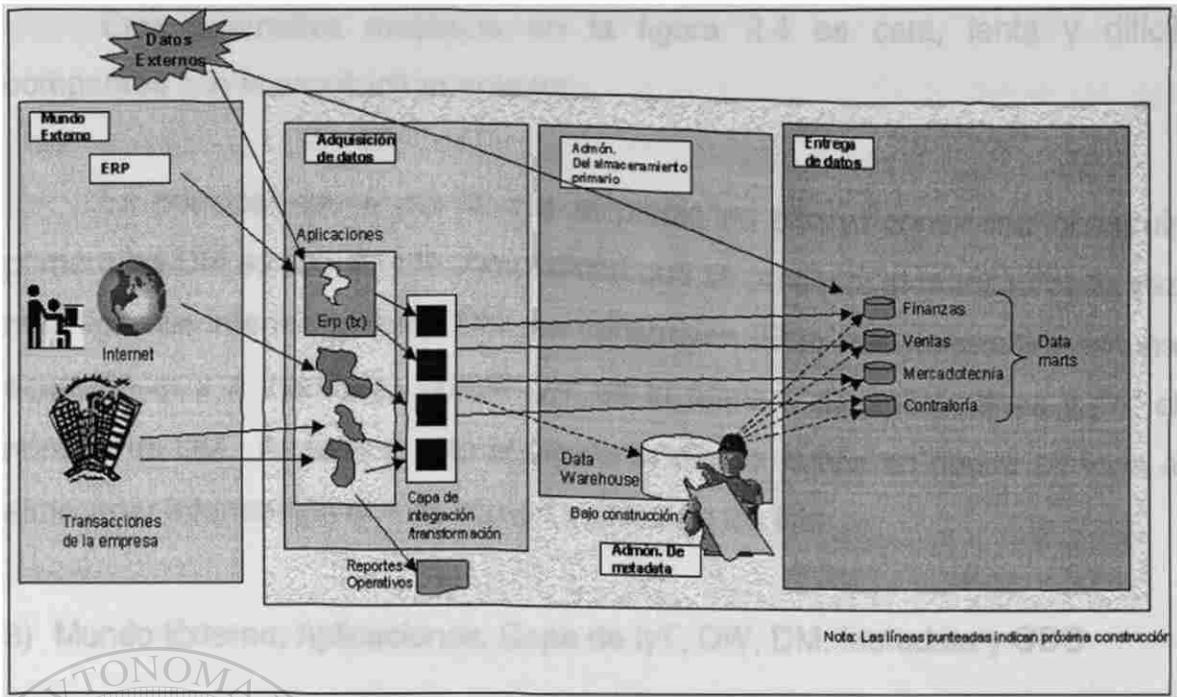


Figura 2.3. Arquitectura 2 iniciando con Datamarts
(Mundo Externo, Aplicaciones, Capa de IyT, DW, DM, Metadata)

Este se construiría directamente de las aplicaciones lo cual al principio es fácil, barato y rápido, pero con malas consecuencias.

- Construcción de los Datamarts en conjunción con el DW

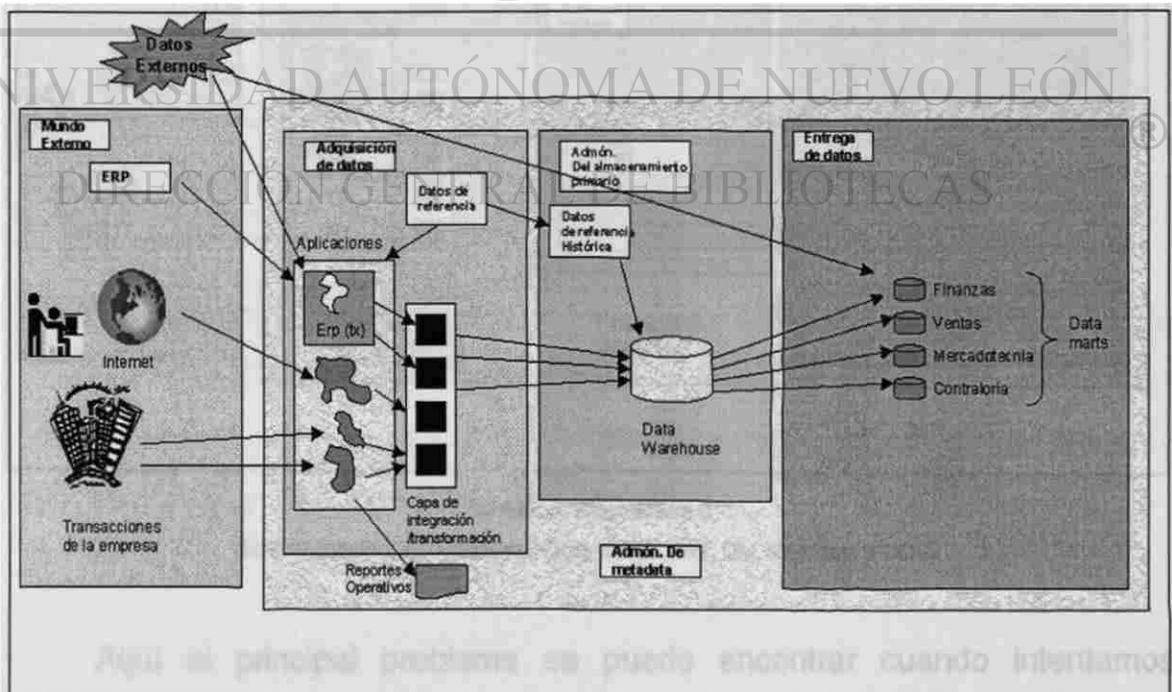


Figura 2.4 Arquitectura 2 Datamarts y Data Warehouse en conjunto
(Mundo Externo, Aplicaciones, Capa de IyT, DW, DM, Metadata)

Esta alternativa mostrada en la figura 2.4 es cara, lenta y difícil comparada con la arquitectura anterior.

La principal causa por la que el precio es alto al considerar construir primero los DM es debido a la complejidad que se presenta al manejar cada vez más ligas de interfaces como DM se construyan. Esto nos repercutirá en una ecuación $m \times n$ interfaces donde "m" es el número de aplicaciones y "n" el número de DM. Aunado a esto el siguiente dilema radica en donde se vaya a almacenar información que es común y única en los DM.

3) Mundo Externo, Aplicaciones, Capa de IyT, DW, DM, Metadata y ODS

Esta arquitectura se muestra en la figura 2.5.

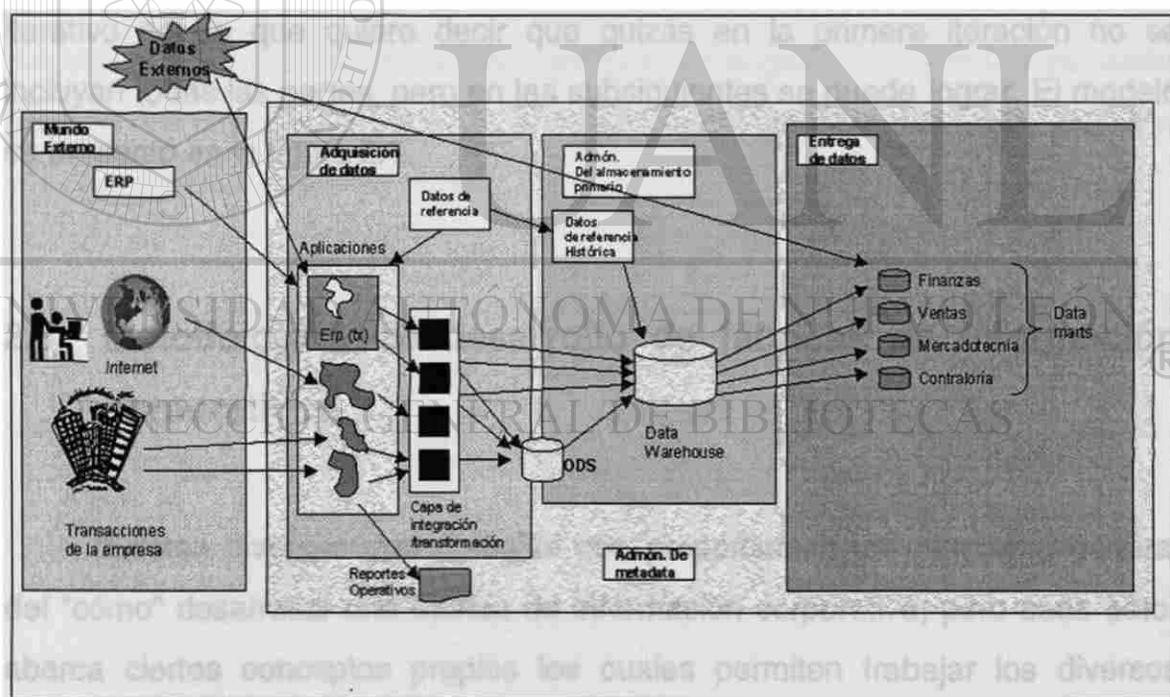


Figura 2.5. Arquitectura 3

(Mundo Externo, Aplicaciones, Capa de IyT, DW, DM, Metadata y ODS)

Aquí el principal problema se puede encontrar cuando intentamos combinar el ODS y el DW en la misma arquitectura. Esto es teóricamente posible. Pero el problema radica en que para poder soportar una arquitectura de

este tipo, es necesario tener un equipo muy potente el cual permita el trabajo de ambas arquitecturas. En pocas palabras el DW y el ODS deben de ser desarrollados en ambientes y entidades separadas físicamente para asegurar una viabilidad a largo plazo del ecosistema de información permitiendo una carga de trabajo combinada, una combinación de los diferentes usuarios que deben usar la FIC, además de una combinación de datos actuales con datos históricos.

4) Arquitectura general.

Finalmente la arquitectura general la cual aún y cuando si se cumple toda puede llegar a ser perfecta, tiene la característica de la mayoría de los proyectos de desarrollo de sistemas. No es posible dar soluciones a corto plazo, además de que hay que pensar que al construir una FIC este es un proceso iterativo por lo que quiere decir que quizás en la primera iteración no se incluyan todas las partes, pero en las subsiguientes se puede lograr. El modelo se presentó en la figura 2.1.

2.3 Metodologías de desarrollo de fábricas de información corporativas.

Existen diversas metodologías que proporcionan un panorama general del "cómo" desarrollar una fábrica de información corporativa, pero cada autor abarca ciertos conceptos propios los cuales permiten trabajar los diversos componentes que engloban esta arquitectura.

2.3.1 Método propuesto por Bill Inmon

En su estrategia de desarrollo y construcción de una fábrica de información corporativa la frase “evolucionar en la naturaleza” [Inmon e Imhoff y Sousa, 01] siempre debe estar en nuestras mentes. El desarrollo de la FIC debe de ser dirigido por acciones estratégicas enfocadas a las necesidades tácticas del negocio, cruciales para la supervivencia de los ecosistemas del negocio que soporta. El ver el concepto de evolución implica un concepto de crecimiento incremental sobre los componentes.

2.3.1.1 El plan estratégico

El primer paso para construir el CIF es el de entender que es lo que maneja al negocio. Las áreas competitivas del negocio representan las áreas necesitadas que proporcionan la habilidad a los procesos del negocio (como ventas, servicios, mercadotecnia).

Las áreas competitivas generalmente son las dedicadas a la administración de las transacciones, cuentas, productos, recursos humanos, distribución, clientes y negocios.

Esta identificación de áreas de negocio se puede observar en la figura 2.6.

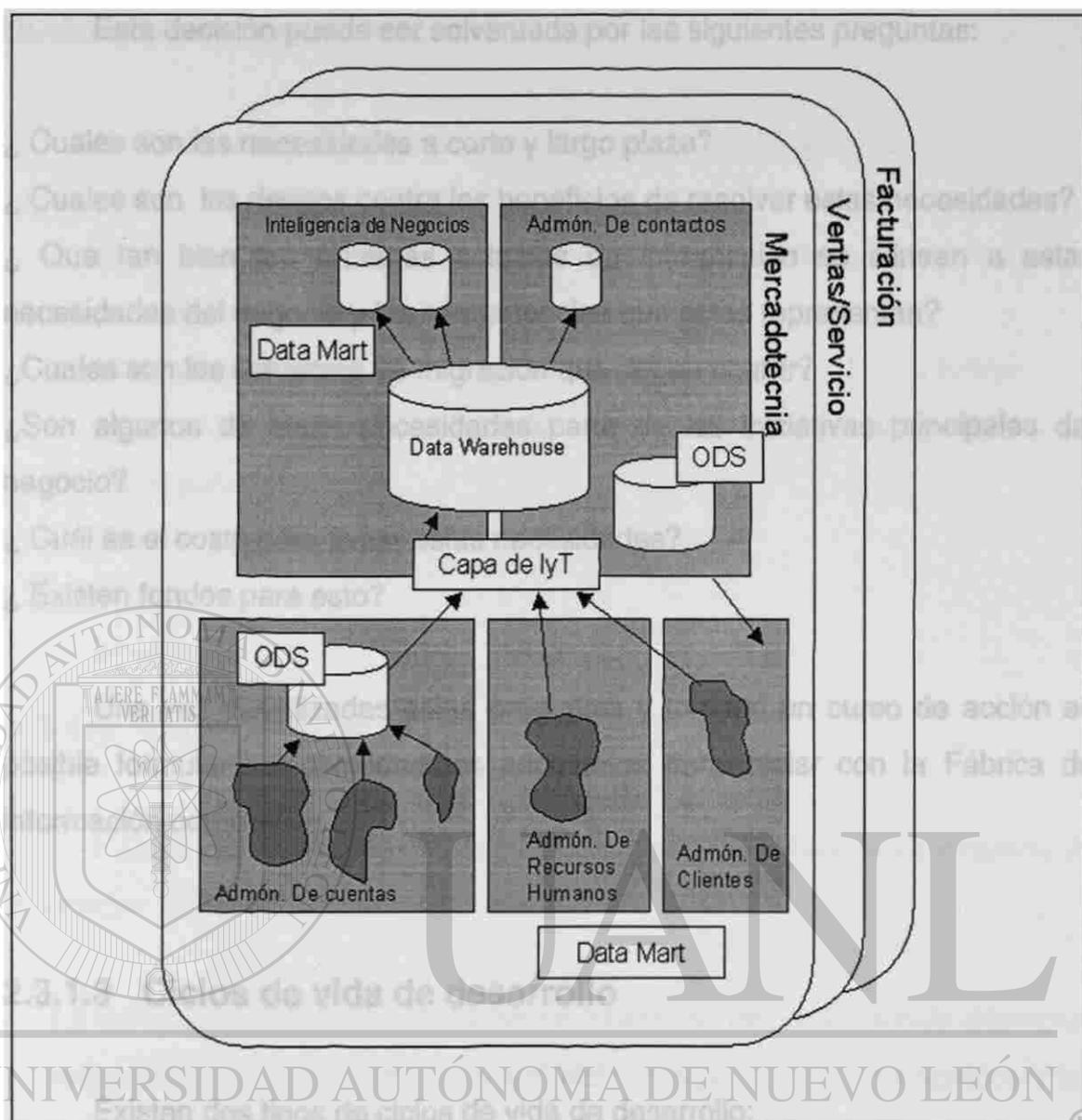


Figura 2.6. Alineación de la FIC a las áreas competitivas del negocio

2.3.1.2 La acción estratégica

Es a partir del plan estratégico que las acciones serán tomadas. El primer paso para determinar las acciones estratégicas es seleccionar de la comunidad de negocios, que áreas serán apoyadas primero.

Esta decisión puede ser solventada por las siguientes preguntas:

- ¿ Cuales son las necesidades a corto y largo plazo?
- ¿ Cuales son los riesgos contra los beneficios de resolver estas necesidades?
- ¿ Que tan bien los sistemas actuales de información se alinean a estas necesidades del negocio y las competencias que estos representan?
- ¿Cuales son los esfuerzos de migración que deben ocurrir?
- ¿Son algunos de estas necesidades parte de las iniciativas principales del negocio?
- ¿Cuál es el costo para llenar estas necesidades?
- ¿ Existen fondos para esto?

Una vez visualizadas estas preguntas y tomado un curso de acción es posible formular los componentes necesarios para iniciar con la Fábrica de información corporativa.

2.3.1.3 Ciclos de vida de desarrollo

Existen dos tipos de ciclos de vida de desarrollo:

1) El ciclo de vida de desarrollo de sistemas (CVDS) [Ruble, 97]

Esta metodología también llamada de cascada usualmente incluye las fases de obtención de requerimientos, análisis, diseño, programación, pruebas, implementación.

2) El ciclo de vida de desarrollo de sistemas (SDVC) [Ruble, 97]

Esta metodología se le conoce como SDVC o metodología en espiral por que tratan de representar la metodología CVDS en forma contraria.

Usualmente consiste en las siguientes fases

- Iniciar con datos implementados, típicamente transaccionales.
- Probar los datos.
- Escribir algunos programas de exploración para determinar que debe de ser realizado para acceder y analizar los datos.
- Una vez que los programas están escritos, realizar un diseño formal.
- Analizar los resultados del diseño y entonces ir hacia atrás y reformular y reprogramar.
- Como paso final, entender los requerimientos.

2.3.1.4 Desarrollo de las bases de datos

Para poder complementar los componentes de la fábrica de información se pueden definir dos estrategias:

- La estrategia basada en un nicho de bases de datos en donde diferentes manejadores de bases de datos son utilizados para diferentes componentes basados en un arreglo ideal. En esta estrategia cada manejador tiene sus propias especialidades para el manejo de datos dependiendo de las características de estos.
- La estrategia basada en Manejadores de propósito general en donde un mismo manejador de bases de datos contiene diferentes componentes de la fábrica de información. Aquí la base de datos puede ser configurada para solventar las diferentes características de los componentes de una FIC y sus capacidades de transaccionalidad y de soporte a la toma de decisiones.

2.3.1.5 Despliegue de información

Otro punto a considerar es los diferentes medios por los cuales la información es presentada a los usuarios dependiendo de sus necesidades.

Posiblemente, al nivel de aplicación los datos se presentan utilizando reportes, terminales tontas y aparatos de interacción directa con el consumidor; el ODS utiliza terminales tontas y reportes para el despliegue de información; el DW así como los DM utilizan una amplia variedad de despliegues, como estaciones de trabajo, hojas de cálculo y reportes.

El modo en que la información es desplegada va de la mano con el estilo del procesamiento de datos.

2.3.2 Método propuesto por Ralph Kimball

Este método aún más reconocido del propuesto por Bill Inmon debido a que entra más en detalles relacionados con el desarrollo del modelo llevando de la mano sobre como se deben de realizar los elementos de una fábrica de información corporativa.

2.3.2.1 Administración del proyecto y requerimientos

Esta metodología provee de una estructura que engloba todos los pasos estipulados para desarrollar una FIC.

Una implementación exitosa de una FIC depende de la apropiada integración de numerosas tareas y componentes. No es suficiente tener el modelo perfecto y la mejor tecnología separadas "Se deben de coordinar las

diversas etapas de un proyecto de este tipo, tanto como si un conductor debe unificar los numerosos instrumentos de una orquesta” [Kimball y Reeves y Ross, 98, pag. 32].

2.3.2.1.1 El ciclo de vida dimensional del negocio

Esta metodología se ilustra en la figura 2.7.

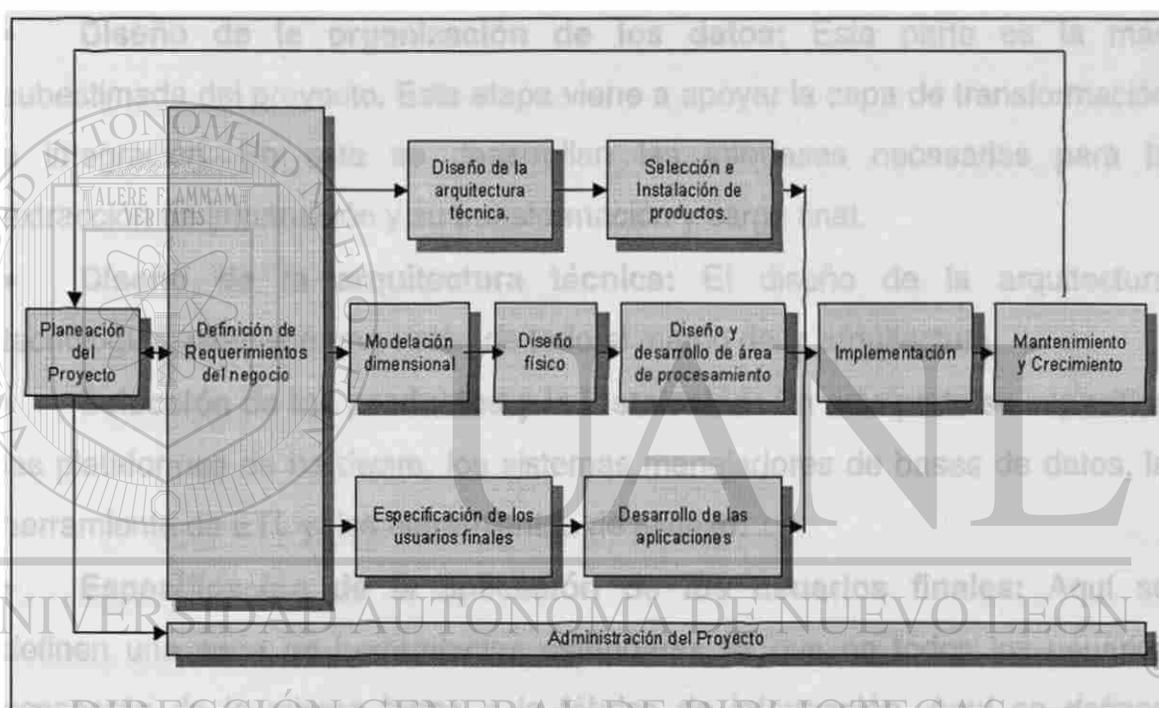


Figura 2.7 El ciclo de vida dimensional del negocio

Las fases se describen a continuación:

- **Planeación de proyectos:** Como todo proyecto el ciclo inicia con una planeación en la cual se estipulan las definiciones y los alcances del proyecto de FIC. Esta planeación se enfoca en recursos, asignaciones de tareas, duración y secuencialidad.
- **Definición de los requerimientos del negocio:** La probabilidad de que un proyecto de esta magnitud tenga éxito se ve muy relacionada con la

capacidad de entender el negocio y los requerimientos de los usuarios finales. Los desarrolladores de un DW deben entender los factores clave para manejar el negocio para entender los requerimientos del negocio y transformarlos en consideraciones de diseño.

• **Modelación dimensional:** Consiste en el diseño de los modelos de datos necesarios para soportar los análisis basados en los requerimientos analíticos de los usuarios.

• **Diseño Físico:** Esta parte se enfoca en la definición de las estructuras físicas necesarias para soportar el diseño lógico de la base de datos.

• **Diseño de la organización de los datos:** Esta parte es la más subestimada del proyecto. Esta etapa viene a apoyar la capa de transformación e integración. En esta se desarrollan las interfases necesarias para la extracción de información y su transformación y carga final.

• **Diseño de la arquitectura técnica:** El diseño de la arquitectura tecnológica establece una visión de todo el marco de la arquitectura.

• **Selección de los productos y la instalación:** En esta parte se especifica las plataformas de hardware, los sistemas manejadores de bases de datos, la herramienta de ETL y las herramientas de acceso.

• **Especificación de la aplicación de los usuarios finales:** Aquí se definen una serie de herramientas estándares ya que no todos los usuarios accederán de la misma forma a la fábrica de información. Aquí se definen Plantillas de reportes, parámetros que maneja el usuario y cálculos requeridos.

• **Desarrollo de la aplicación de los usuarios finales:** Aquí se enfoca básicamente en configurar la herramienta de metadata y construir los reportes específicos.

• **Liberación:** Esta representa la convergencia de la tecnología, datos y las aplicaciones de los usuarios finales accesibles desde las máquinas de los usuarios de negocio.

• **Mantenimiento y Crecimiento:** Abarca el trabajo restante que sigue a la hora de implantar un proyecto de esta magnitud. Aquí se debe uno de enfocar en los usuarios de negocio proveyéndolos de soporte constante y educación

- **Administración del proyecto:** Este paso asegura que las actividades del ciclo de vida de desarrollo dimensional se mantengan adelante y en sincronía.

2.3.2.1.2 Administración del proyecto

El primer paso antes de iniciar con un proyecto de FIC es obtener los cinco factores llave para saber si el negocio está listo para tener una arquitectura de este tipo:

- Un patrocinador fuerte en la administración del negocio
- Una motivación fuerte de negocios (y no de crisis internas o de supervivencia)
- Una asociación entre el negocio y el área de sistemas de información
- Una cultura analítica
- Viabilidad.

El segundo paso consiste en realizar una prueba de concepto demostrando el impacto potencial de la FIC. Posteriormente se desarrolla un alcance preliminar para finalmente concluir con la justificación necesaria para el desarrollo. Una vez pasadas estas fases es necesario definir un plan del proyecto.

2.3.2.1.3 Recolección de requerimientos

Debido a que los usuarios de negocios y sus requerimientos impactan en casi cada decisión hecha a través de la implementación del FIC, esta fase es el centro del universo del Proyecto. Los requerimientos especifican que datos deben de estar disponibles en el DW, como estarán organizados y muy seguido como serán modificados. Se recomiendan dos técnicas:

- f) Entrevistas
- 2) Utilización de facilitadores para sesiones grupales.

2.3.2.2 Diseño de datos

Aquí se realiza el diseño lógico y físico de la FIC. Este diseño transforma las fuentes en estructuras de la Fábrica de Información. En esta parte es necesario realizar una modelación dimensional como única técnica viable para el diseño lógico utilizado en el DW-ODS-DM ya que esta modelación permite presentar los datos en una arquitectura intuitiva y con accesos de alto desempeño.

Este difiere de la modelación Entidad-Relación [Chen, 76] debido a que ésta sirve para los modelos transaccionales y también puede ser utilizada en las fases de administración de datos a la hora de construir un DW, pero nunca debe ser utilizado para entregar al usuario final datos.

Cada modelo dimensional está compuesto por:

- **Tabla de hechos:** Es una tabla con una llave compuesta y contiene algo medible y que corresponde a un indicador del negocio. En pocas palabras es todo lo que puede medirse.
- **Tablas de dimensión:** Es una tabla en donde su llave única pertenece a la llave compuesta de la tabla de hechos. Son una colección de atributos que están relacionados unos con otros.
- **Atributos:** Se utilizan para describir una característica tangible de alguna cosa. Estos se organizan en las dimensiones.

Esta combinación de elementos forma una estructura llamada Diagrama Estrella mostrada en la figura 2.8.

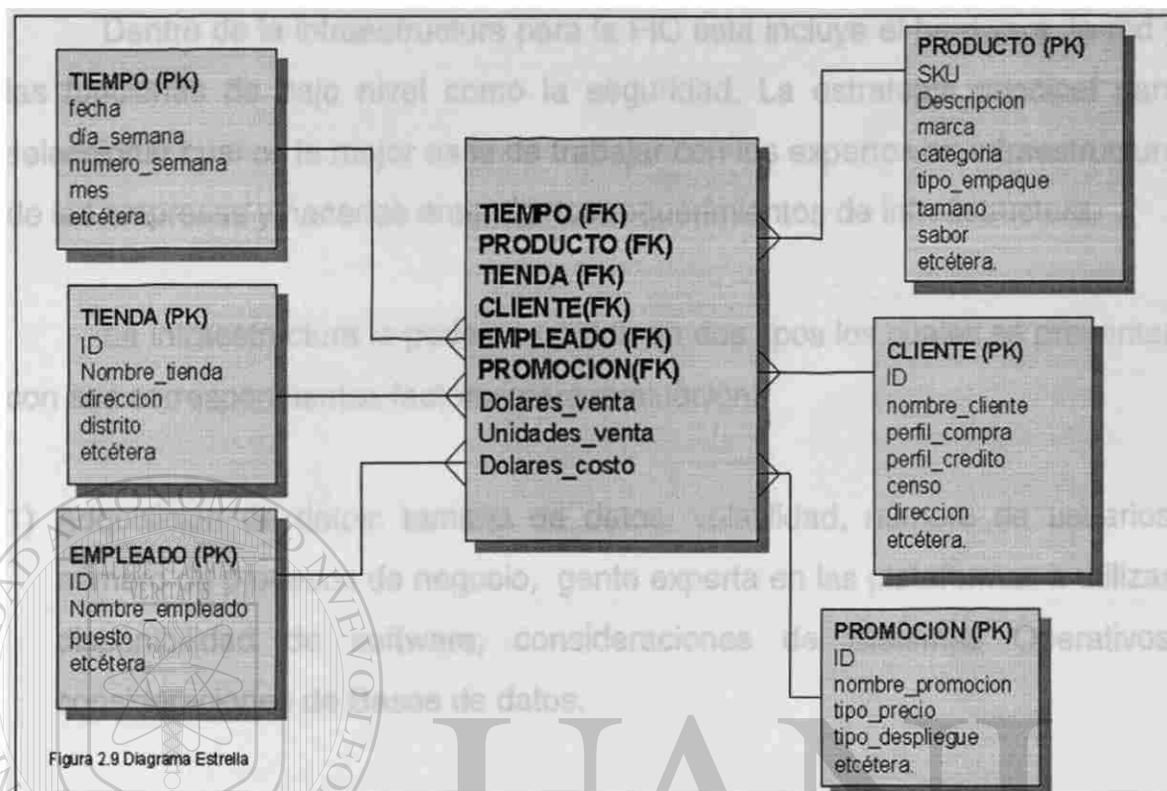


Figura 2.9 Diagrama Estrella

Figura 2.8 Diagrama Estrella

2.3.2.3 Selección de la arquitectura

En la fase de entendimiento de la arquitectura, básicamente se puede explicar con una analogía directa relacionada con los planos de una construcción de una casa. No es viable iniciar a construir si no se tienen los planos previamente hechos. Estos planos ayudan al arquitecto y al cliente a comunicar acerca de los resultados deseados y la naturaleza del esfuerzo de construcción. Además esto nos ayuda a identificar los recursos requeridos, las dependencias, los tiempos y los costos.

El trabajo a realizar a la hora de seleccionar la arquitectura es identificar las capacidades que son más importantes para la organización, una vez

realizado esto es necesario identificar los vendedores que proveen estas capacidades y determinar cual es el que mejor se acopla a la situación.

Dentro de la infraestructura para la FIC ésta incluye el hardware, la red y las funciones de bajo nivel como la seguridad. La estrategia principal para seleccionar cual es la mejor es la de trabajar con los expertos en infraestructura de las empresas y hacerles entender los requerimientos de infraestructura.

La infraestructura la podemos dividir en dos tipos los cuales se presentan con sus correspondientes factores para evaluación:

- 1) **Adquisición de datos:** tamaño de datos, volatilidad, número de usuarios, número de procesos de negocio, gente experta en las plataformas a utilizar, disponibilidad de software, consideraciones de Sistemas Operativos, consideraciones de Bases de datos.
- 2) **Presentación de datos:** Consideraciones del servidor de aplicación (Memoria, Disco, cuellos de botella), de máquinas cliente, de conectividad y de red.

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN[®]
DIRECCIÓN GENERAL DE BIBLIOTECAS

La figura 2.9 muestra algunas de las diferentes configuraciones posibles:

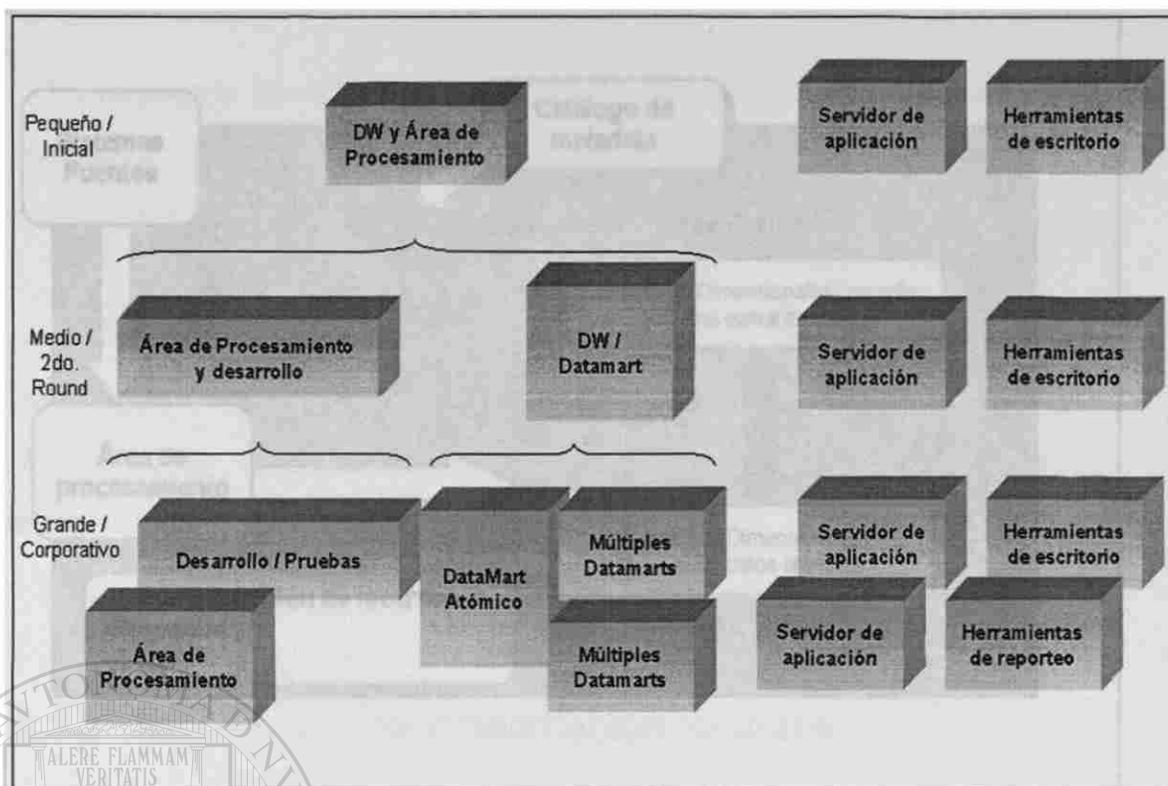


Figura 2.9 Evolución de las configuraciones de arquitectura

2.3.2.3.1 Arquitectura de Adquisición de datos.

Básicamente la parte de la arquitectura que se enfoca a la adquisición de datos inició con un esfuerzo llevado a cabo manualmente por medio de código, ésta se conoce como "Back End". Afortunadamente la industria ha desarrollado algunas herramientas y técnicas para mantener esta parte de la FIC.

La arquitectura y los servicios se presentan en la figura 2.10.

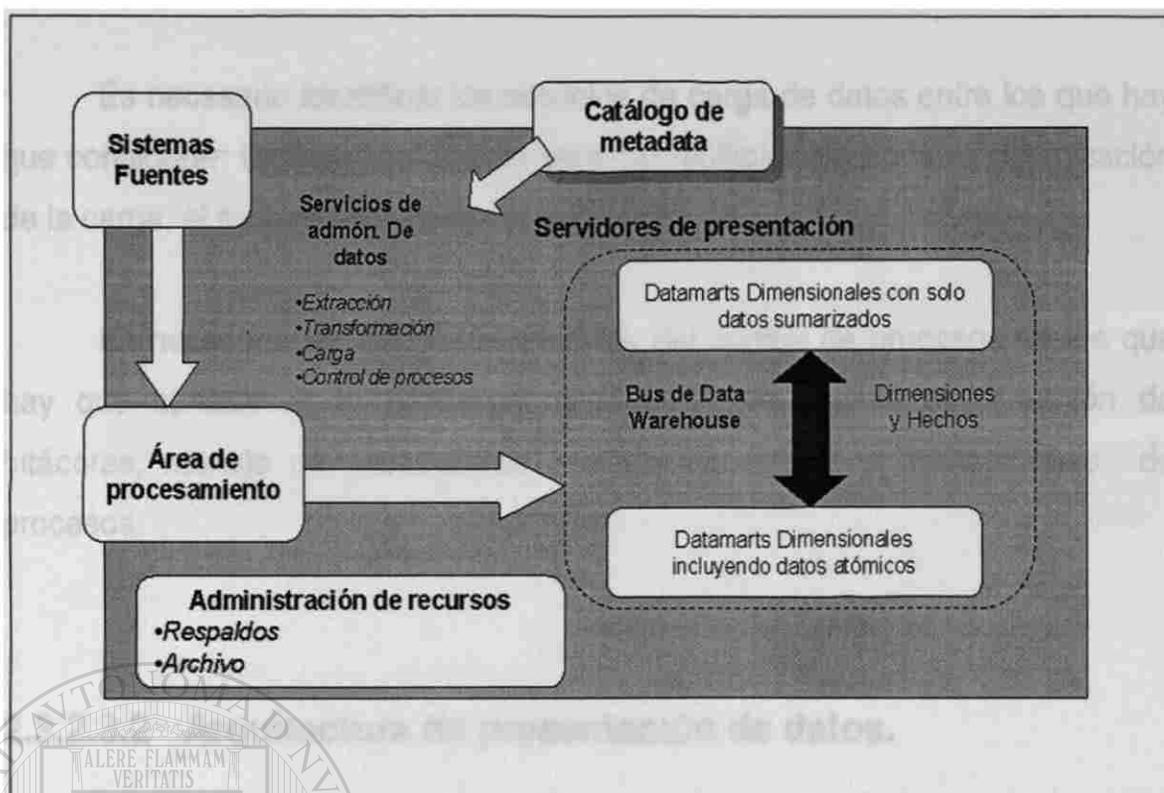


Figura 2.10 Arquitectura de Back End

Para entender esta parte de la arquitectura es necesario identificar los diferentes componentes que la conforman:

- **Sistemas fuente:** Estos son las fuentes de información del negocio
- **Area de transformación:** Ésta es el área de construcción de la FIC. No debe de ser vista por el usuario. Aquí se realizan las transformaciones necesarias para alimentar tanto como el DW como el ODS.
- **Los servidores de presentación:** Básicamente, estos son las plataformas destinos en donde los datos son almacenados para su uso por los usuarios finales, los sistemas de reporte y otras aplicaciones.

Es necesario identificar los servicios de transformación de datos que se van a requerir entre estos hay que considerar: La Integración, los chequeos de referencias de integridad, la normalización y denormalización, la limpieza y depuración, conversión, calculo, derivación y localización de datos; la agregación, la auditoría de datos, validaciones de valores nulos, etcétera.

Es necesario identificar los servicios de carga de datos entre los que hay que considerar: La funcionalidad de soportar múltiples destinos, la optimización de la carga, el proceso entero de carga.

Es necesario identificar los servicios del control de procesos en los que hay que considerar: la definición, calendarización, monitoreo, creación de bitácoras, manejo de excepciones, manejo de errores y notificaciones de procesos.

2.3.2.3.2 Arquitectura de presentación de datos.

Esta arquitectura representa la cara pública de la FIC. Esta es la que los usuarios del negocio ven y con la que trabajan día a día. Esta es la interfaz de los usuarios. Desgraciadamente los datos que ellos quieren acceder son complejos. El modelo dimensional reduce la complejidad, pero los negocios tienen reglas y excepciones que deben ser incluidos en la FIC para poderlos analizar y entender.

Esta arquitectura se presenta en la figura 2.11.

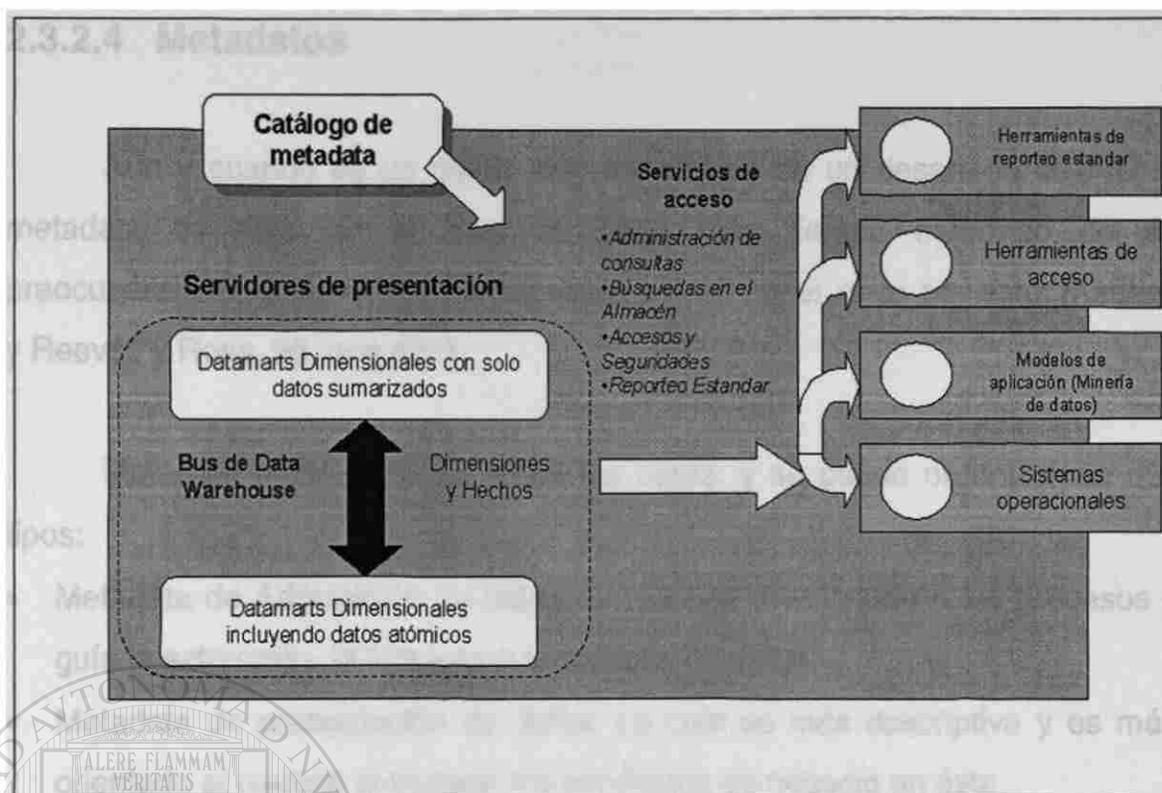


Figura 2.11 Arquitectura de Front End

Uno de los puntos principales a entender es sobre que accesos se van a realizar y los medios alternativos de almacenamiento para poder presentar la información al usuario final.

Los servicios que se deben de considerar para el acceso a los datos son: La navegación en el DW, ODS o DM, el control de accesos y la seguridad de la información, el monitoreo sobre desempeño y soporte, administración de consultas, simplificación de contenidos, la reformulación de consultas, la capacidad de poder seleccionar agregaciones, fechas anteriores sobre la explotación para comparativos, etcétera.

Uno de los puntos importantes para mejorar los accesos es entender quienes son los clientes primarios y la información que trata de obtener, además de identificar las capacidades necesarias para lograr el punto anterior. Esto es importante a la hora de seleccionar una herramienta capaz de satisfacer los requerimientos mencionados.

2.3.2.4 Metadatos

Aún y cuando es un punto muy importante en un desarrollo de FIC la metadata es algo de lo cual se pasa más tiempo hablando de él, preocupándonos y sintiéndonos culpables por no hacer nada por este. [Kimball y Reeves y Ross, 98, pag 435]

Metadata significa datos sobre los datos, y se puede organizar en dos tipos:

- **Metadata de Adquisición de datos:** La cual es relacionada a los procesos y guía la extracción, la limpieza y el proceso de carga.
- **Metadata de presentación de datos:** La cual es más descriptiva y es más orientada al usuario al trabajar los conceptos de negocio en ésta.

Dentro de estos tipos se pueden observar varias divisiones que son:

- **Metadata de las fuentes:** Especificaciones, información descriptiva y de procesos.
- **Metadata de los procesos de adquisición:** Adquisición de datos, Administración de las tablas de dimensión, Transformaciones y Agregaciones, bitácoras y documentaciones.
- **Metadata de la base de datos:** Todo lo relacionado con los objetos de ésta.
- **Metadata de la presentación de datos:** Nombres de negocio y descripciones de las columnas, tablas y agrupaciones, definiciones de consultas, seguridades.

En resumen metadata es todo el proyecto, excepto por todos los datos almacenados.

En la figura 2.12 se presentan los roles de la metadata en todo el proceso.

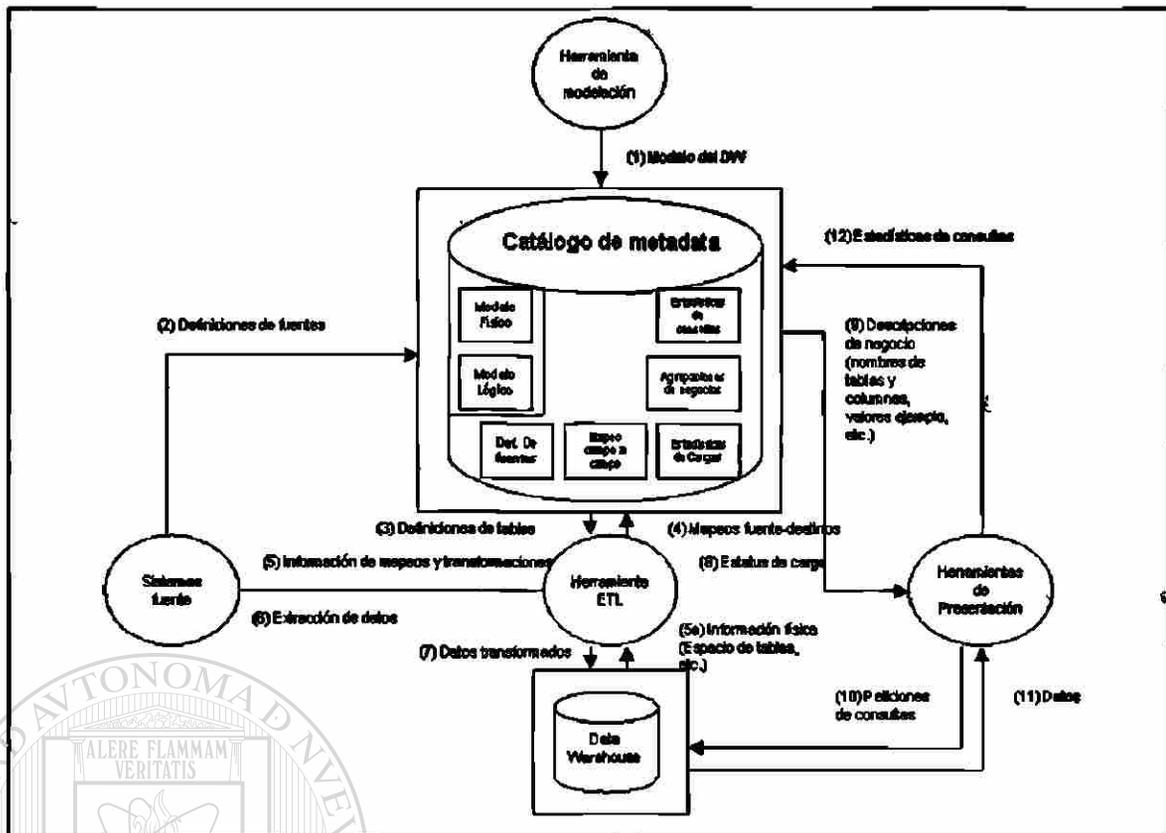


Figura 2.12. Metadata

2.3.2.5 Implementación

En este punto se habla de la forma de llevar a cabo todo lo diseñado en modelos lógicos y plantearlo en la realidad.

2.3.2.5.1 Complementando el diseño físico

Para finalizar el diseño físico se proponen una serie de conceptos para facilitar la creación física del modelo lógico obtenido que van desde desarrollo de estándares, desarrollo del modelo de datos físico, definición del plan inicial de índices, diseño y construcción de la instancia de la base de datos, diseño de la estructura física de almacenamiento

2.3.2.5.2 El área de procesamiento

El área de proceso es el iceberg en un proyecto de FIC². Aquí es donde se presentan la mayor cantidad de retos debido a la dificultad de tratar con muchas fuentes de datos que invariablemente tienden a tomar mucho más tiempo del que uno espera.

Aquí se crea el plan, se selecciona la herramienta de ETL, se aplican todas las reglas de negocio y se automatiza.

2.3.2.6 La construcción de las aplicaciones de los usuarios finales

En esta se visualiza la presentación de la información a los usuarios finales, teniendo como principal actividad la creación de un conjunto de reportes que se conocen como bases de reporte de aplicación del usuario final.

El proceso de crear estas aplicaciones se divide en dos fases:

1) Especificaciones

2) Desarrollo

2.4 Problemas encontrados usualmente en un DW y sus posibles soluciones.

Las diez principales causas de problemas al construir un DW [TDWI, 00] son:

1) Empezar con una cadena de patrocinio errónea, es decir, sin incluir al mediador que permite que se den las cosas para el proyecto entre el patrocinador ejecutivo y el administrador del proyecto de DW.

- 2) Ofrecer expectativas de cosas que no se pueden cumplir y la frustración de los ejecutivos al momento de la verdad.
- 3) Trabajar con comportamientos políticos ingenuos solo para vender la idea con comentarios como "haremos que los ejecutivos tomen mejores decisiones", en este caso dando un mensaje adverso que se traduce en "ustedes no toman buenas decisiones" y por lo tanto con el DW yo lo arreglaré.
- 4) Cargar información al DW sólo porque estaba disponible.
- 5) Creer que el diseño de una BD de DW es similar al de una base de datos transaccional.
- 6) Seleccionar un Administrador de DW orientado a la tecnología en lugar de al usuario.
- 7) Enfocarse en el concepto de almacenamiento de datos obtenidos sólo a nivel interno en lugar de almacenar datos del mundo externo de valor potencial como textos o imágenes.
- 8) Proporcionar información con Definiciones que se traslapan o que son confusas.
- 9) Creer en las promesas de capacidad, desempeño y escalabilidad.
- 10) Creer que una vez que se libera un DW los problemas han finalizado.
- 11) Creer que sólo existen 10 problemas a la hora de realizar un DW

DIRECCIÓN GENERAL DE BIBLIOTECAS

Estos son algunos de los factores críticos para que un proyecto de DW tenga éxito [Dorsey, 00]:

- El líder de proyectos debe de ser experimentado.
- Realizar el análisis y la recolección de requerimientos de usuario incluyendo los reportes en forma cuidadosa
- Tener en mente el cómo lo usuarios van a usar la nueva herramienta
- Realizar un proyecto piloto
- Tener a los usuarios felices. Ya que con solo tener un usuario tomando una mala decisión basado en una mala información todos los usuarios perderán la confianza en los datos.

2.5 Herramientas para alcanzar la creación de una FIC

Un punto importante para lograr una buena implementación de un concepto como la FIC recae en la correcta selección de las herramientas necesarias para llevar a cabo una tarea de esta magnitud.

Las herramientas a utilizar se pueden clasificar en tres categorías basadas en las actividades a desempeñar en el desarrollo de una FIC [Sakaguchi y Frolick, 96] las cuales son:

- **Herramientas de adquisición:** Las cuales son necesarias para desempeñar las tareas de modelar, diseñar y poblar un DW. Estas herramientas extraen los datos de varias fuentes y las transforman para hacer los datos utilizables en el DW. Además establecen el Metadatos donde se almacena la información acerca de los datos.
- **Herramientas de almacenamiento:** Estas están compuestas por bases de datos relacionales generalmente para permitir que los datos sean utilizados para el soporte a las decisiones en forma efectiva.
- **Productos de acceso:** Básicamente son los productos para análisis multidimensionales que apoyan a los usuarios finales al acceder y analizar los datos en las diferentes formas de verlos en un DW.

2.5.1 Herramientas de adquisición

Las herramientas de adquisición se pueden identificar en dos tipos, las herramientas de Extracción transformación y carga, y las herramientas modeladoras, ambas se complementan para poder realizar la construcción de una FIC.

2.5.1.1 Herramientas ETL

El proceso de desarrollar una fábrica de información tiene como punto primordial la necesidad de manejar efectiva y eficientemente la extracción, la limpieza la transformación y la migración de datos desde los sistemas fuentes. El porqué son necesarias estas herramientas se explica debido a que estas ayudan a obtener la eficiencia técnica y pueden realizar tareas de calidad de datos que de otra manera no serían prácticas.

Para poder decidir cuales son las herramientas más apropiadas para asistir con estas tareas es necesario contestar las preguntas acerca de las necesidades del negocio [Orli y Santos, 96] como:

- ¿La base de conocimiento acerca del sistema es adecuada o inadecuada?
 - ¿Las extracciones son de una sola vez o en forma repetitiva?
 - ¿Los asuntos sobre la calidad de datos son específicos o generales?
 - ¿Existe una herramienta que se acople a las características necesarias o es necesario adquirirla o crearla?
-
- ¿Es necesario controlar la extracción o la transportación en forma centralizada o distribuida y además vía parámetros o por codificación fija?

DIRECCIÓN GENERAL DE BIBLIOTECAS

Una herramienta para estas actividades se denomina ETL "Extraction, Transformation and Loading" y se define como un conjunto de herramientas para el propósito de extracción, transformación y carga de datos a DW [Tylor, 01]. Este enfoque es diferente que las soluciones más comunes de en donde se tenían un conjunto de programas o herramientas para cumplir con las mismas tareas.

Algunas de las herramientas ETL que podemos encontrar actualmente en el mercado para esta actividad son [DWOL, 01]: DataStage (Ascential),

DecisionBase (Computer Associates), DecisionStream (Cognos), Powermart (Informatica), Data Mirror.

2.5.1.2 Herramientas Modeladoras.

Otro tipo de herramientas que recaer en esta categoría son las herramientas de modelación de datos [Plaster, 97] las cuales pueden definirse como aquellas herramientas que permiten dibujar figuras, cajas, flechas de los elementos que conforman el negocio y que tienen facilidades para introducir información como definiciones e identificadores. Además que permite crear una completa documentación del diseño de la base de datos permitiendo además generar la base de datos en diferentes plataformas y "software" manejadores de bases de datos.

Estas herramientas están basados en el modelo definido por Peter Chen [Chen, 76] las cuales facilitan manejar el mundo real y plasmarlo en un modelo lógico entendible de datos.

Algunas herramientas en el mercado son: Erwin Modelmart (Computer Associates), Power Designer (Sybase), Visible Analyst (Visible systems Corporations).

2.5.2 Herramientas de almacenamiento

Básicamente estas son herramientas especializadas que permiten almacenar la información, pero de una manera estructurada para la toma de decisiones. Estas herramientas en cierta forma se complementan con las herramientas de adquisición de datos y generalmente son los manejadores de Bases de datos en el mercado como Oracle, Informix, Progress, etcétera.

2.5.3 Productos de acceso

Estas herramientas ayudan con la actividad general de realizar consultas y presentar datos numéricos y caracteres desde el Data Warehouse. Generalmente estos se conocen como herramientas OLAP "On line analytical Processing" o procesamiento analítico en línea las cuales permiten un análisis en línea de la información teniendo una tecnología no relacional y generalmente basado en un cubo de datos multidimensional.

Las herramientas OLAP pueden dividirse en dos categorías:

- 1) ROLAP (Relational OLAP): Conjunto de herramientas y aplicaciones que dan a una base de datos relacional, un sabor dimensional.
- 2) MOLAP (Multidimensional OLAP): Conjunto de herramientas y aplicaciones puramente dimensionales.

Estas herramientas muchas veces se complementan con herramientas de accesos especializadas para los usuarios finales las cuales invitan al usuario final a realizar sus propias consultas permitiéndoles manipular las tablas relacionales y sus relaciones.

DIRECCIÓN GENERAL DE BIBLIOTECAS

Algunas herramientas que existen actualmente en el mercado son [DWOL, 01]: Power Play (Cognos), Web Intelligence (Business Objects), Essbase (Hyperion), Business Intelligence Solutions (IBM), Microstrategy 7 (Microstrategy).

2.6 El por qué una arquitectura FIC puede ayudar al problema de descentralización de información, Inflexibilidad y optimización de un Data Warehouse

La FIC es difícilmente la única forma de construir DSS "Decision support Solutions", pero es la mejor manera de alcanzar los objetivos a largo plazo del procesamiento de información de la compañía [Inmon e Imhoff y Sousa, 2001].

¿Porqué?

"Abarca todos los puntos que deben de tomarse en cuenta a la hora de desarrollar una solución para el soporte a la toma de decisiones por medio de una arquitectura robusta y una serie de componentes los cuales permiten organizar los sistemas de información ante los problemas generales que se pueden encontrar. Los problemas como demandas de negocio en constante evolución, la constante disminución de los costos de la tecnología, incremento en los requerimientos de los usuarios y el crecimiento de las capacidades del software, hardware y red son bien atacados por esta arquitectura"

[Inmon e Imhoff y Sousa, 2001].

3 MÉTODO PROPUESTO

3.1 Metodología de desarrollo propuesta

La siguiente es la propuesta para desarrollar el medio para obtener los objetivos de mejoramiento en todos los procesos implicados en el entorno de la solución de la FIC.

Esta describe una serie de cuestiones basadas en una combinación de sugerencias de los especialistas de las soluciones para el soporte a la toma de decisiones Bill Inmon [Inmon e Imhoff y Sousa, 01] y Ralph Kimball [Kimball y Reeves y Ross, 98] en la que se manejan las diferentes etapas tanto de creación de proyecto, carga, validación y explotación de información.

Una punto importante de esta metodología es que lleva implícita en los cuatro objetivos buscados de esta tesis la implementación del concepto de metadata pilar en todo proyecto de FIC.

Para entender mejor la distribución del modelo y la metodología se presentan las 2 divisiones de servicios propuestas para la FIC de GAMESA.

- **Back End.** Que es todo lo relacionado con la arquitectura interna de almacenamiento, mantenimiento y administración de datos. En pocas palabras el creador de la materia prima para la FIC.
- **Front end.** Que es todo lo relacionado con la arquitectura de explotación y sus respectivos servicios.

Esta metodología aplicada está compuesta por los siguientes pasos los cuales en conjunto nos servirán para la creación del modelo de la FIC. Cada uno de estos atacan alguna de las cuatro áreas de problemas detectadas en el concepto de FIC identificándose de la siguiente manera:

1) Mejoramiento en el proceso de desarrollo (Back End)

- Establecimiento del estado actual
- Realización del plan de proyecto (A nivel Front End también)
- La obtención de requerimientos
- Diseño del modelo lógico de la FIC (Basado en método Inmon [Inmon e Imhoff y Sousa, 01])
 - Construcción de matriz de bus
 - Utilización del método de los cuatro pasos para diseñar cada tabla de hechos
 - Diagramación de la tabla de hechos
 - Diagramación de la tabla de dimensiones y su correspondiente detalle
- Hoja de hechos derivados
- Identificación de las fuentes de cada tabla de hechos y dimensiones
 - Mapeo de los campos fuentes a los campos destinos
 - Desarrollo por medio de herramienta modeladora del modelo físico a un manejador de Base de datos
 - Dimensionamiento
 - Planeación y Diagrama de arquitectura tecnológica.

2) Mejoramiento en los procesos de carga (Back End)

- Utilización de herramientas y desarrollo
- Mapa de configuración

3) Mejoramiento en los procesos de validación (Back End)

- Definición de los tipos de validación a utilizar y los formatos de presentación de las bitácoras.

4) Mejoramiento en el proceso de explotación (Front End)

- Establecimiento de la arquitectura de explotación

3.2 Definición de herramientas a utilizar

El hecho de seleccionar una herramienta para las actividades específicas que aplicarán sobre la FIC cae fuera del ámbito de estudio de esta tesis. Existen técnicas aplicadas que pueden ser utilizadas para la selección de herramientas que mejor se acoplen a los requerimientos, características y capacidades de la empresa o proyecto que en un momento desee crear una FIC.

En base a las características de la FIC se proponen herramientas para el cumplimiento de los objetivos de esta tesis. Estas herramientas son del tipo:

- Administración de Proyectos
- ETL
- Modeladora
- Acceso

3.2.1 Herramienta de Administración de proyectos

Para esta tesis una de las herramientas CASE empleadas es Microsoft® Project 98. [MICROSOFT, 2001]

La utilización de esta herramienta se observa para la creación y administración del proyecto. Por medio de ésta, es posible definir las diferentes actividades del ciclo de vida del desarrollo, asignando fechas de inicio y fin, asignando recursos, etcétera.

Es importante hacer notar que esta asignación varía dependiendo del tamaño del proyecto.

3.2.2 Herramienta ETL

En el mercado existen actualmente una serie de herramientas las cuales en los años 90 eran muy especializadas y por lo mismo no muy conocidas o no muy avanzadas. A la fecha, estas herramientas han ido evolucionando proporcionando características muy útiles para la realización de la Extracción, Transformación y Carga.

La herramienta ETL seleccionada para ser utilizada en esta tesis y en esta iteración es Informatica Powermart™ v4.6 (R87 D17). [INFORMATICA, 2001].

La característica primordial de esta herramienta es que puede conectarse a las diferentes fuentes y destinos necesarios para la alimentación de la FIC. Además proporciona las diferentes capacidades que una ETL debe contener.

Esta herramienta se divide en tres aplicaciones:

- *Designer(Diseñador)*: Esta aplicación tiene como función principal permitir el desarrollo de las interfases en un ambiente totalmente gráfico que disminuya la cantidad de codificación que se deba de realizar, permitiendo a su vez un rápido entendimiento de los desarrollos debido a su forma de organizar la carga. Todo desarrollo es conocido como mapeo.

Esta parte del Designer se presentan nueve objetos posibles a utilizar en el desarrollo los cuales se listan a continuación:

- 1) **"Source Qualifier"**: Este objeto representa la extracción de información desde la fuente (Son las consultas a bases de datos o las extracciones a los archivos texto)
- 2) **"Normalizer"**: Este objeto transforma las fuentes estilo COBOL (normalizadas a la primera forma normal) a una normalización tercera forma normal.
- 3) **"Expression"**: Este objeto realiza los cálculos aritméticos, condicionales, etcétera.
- 4) **"Aggregator"**: Este objeto realiza operaciones de agregación (sumas, máximos, mínimos)
- 5) **"Filter"**: Este objeto se utiliza como un filtro condicional
- 6) **"Lookup"**: Este objeto permite realizar actividades de búsqueda cargando en memoria la información y permitiendo una búsqueda basada en llaves para extraer su correspondiente atributo de relación.
- 7) **"Stored Procedures"**: Permite llamada a Procedimientos almacenados de las bases de datos.
- 8) **"Sequence Generator"**: Genera secuencias de números.
- 9) **"Joiner"**: Este objeto permite la unión de fuentes diferentes (diferentes tipos de bases de datos y archivos textos) en las cuales el motor de procesamiento de alguna base de datos no la puede realizar.

- **Server Manager(Administrador del servidor)** : Esta aplicación tiene como función principal configurar las conexiones, accesos, bitácoras, calendarizaciones de las interfases desarrolladas en el "Designer", en pocas palabras permite administrar los procesos realizados de una manera amigable. Toda configuración de cada mapeo con todas sus características se crea en un concepto denominado sesión. Es posible agrupar cada sesión en unos objetos denominados "batches" que nos permiten administrar las

sesiones de tal manera que si un proceso de carga contiene varias sesiones, estas puedan correr de un modo secuencial o concurrente pudiendo ser calendarizados en un solo lugar (el batch).

- **Repository Manager:** Esta aplicación es el corazón de toda la herramienta ya que cumple con la función de administrar toda la metadata que existe en el desarrollo y la cual puede ser utilizada para cumplir con la metadata del proyecto de toda la parte de Back End.

Esta herramienta cumple con las necesidades de una área de procesamiento ya que todas las actividades o servicios que debe proporcionar la parte de Back End se ven solventadas por la propia herramienta.

3.2.3 Herramienta Modeladora

La herramienta modeladora seleccionada para cumplir con los objetivos de esta tesis es Platinum Technology® Erwin v3.5.2. [CAI, 2001].

Esta herramienta cumple con las necesidades de ser independiente de la plataforma en que se vaya a crear la estructura de la Base de datos ya que permite definir todas las características del modelo sin ninguna complicación. ®

3.2.4 Herramienta de acceso

La herramienta de acceso seleccionada es Microstrategy™ V.7.1.114. [MICROSTRATEGY, 2001].

Esta herramienta cumple con las características deseadas de explotación.

Es una herramienta ROLAP, la cual permite explotación en línea sin necesidad de creación de Cubos ya que se dirige directamente al sistema

manejador de base de datos (DBMS). Otra de sus características es que permite análisis de una manera amigable sin necesidad de instalar nada en las máquinas cliente ya que su arquitectura se basa en Tecnología de Web (Internet), por lo que lo único que se necesita es un Navegador de Internet. Además permite definir roles, usuarios.

3.3 Formas de atacar objetivos de la tesis

Cada uno de los objetivos a alcanzar conforma una parte de todo el proceso de creación de una FIC. Estos objetivos se resumen en:

- **Mejoramiento en el proceso de Desarrollo**
- **Mejoramiento en procesos de Carga**
- **Mejoramiento en procesos de Validación**
- **Mejoramiento en procesos de Extracción o Explotación**

Cada uno de estos objetivos utilizan cierta metodología la cual se complementa con gráficas y estrategias para medir el grado de cumplimiento de los mismos.

Estos se desglosan a continuación.

3.3.1 Mejoramiento en proceso de desarrollo

Para poder llevar a cabo el proceso de desarrollo de una fábrica corporativa y sus siguientes iteraciones es necesario una metodología formal para cumplir el objetivo. Dentro de lo que se busca en esta tesis es utilizar cada uno de los puntos de la metodología de Kimball [Kimball y Reeves y Ross, 98] e Inmon [Inmon e Imhoff y Sousa, 01] para llegar al mejoramiento del proceso de desarrollo de una FIC. El corazón de esta tesis se basa en que una buena

metodología y una buena estrategia de desarrollo permitirá lograr solucionar casi todos los problemas que se presentan en el ambiente DSS ya que son la base de los datos a almacenar y posteriormente transformar en información.

Dentro de este punto uno de los conceptos que se desea medir es el grado de mejoramiento del proceso de desarrollo, el indicador a utilizar es la disminución de tiempos en cada una de estas etapas. Para esto se compararán los tiempos de desarrollo en el proyecto original, con los nuevos una vez que se ha aplicado la metodología propuesta.

Otro de los puntos que se pueden observar, pero que son muy subjetivos de medir son la facilidad de desarrollo (el cual puede interpretarse por los tiempos de desarrollo) y la estandarización de una metodología que una vez dominada facilitará el desarrollo de futuras iteraciones.

Como parte fundamental de este proceso de mejoramiento en los siguientes apartados se presenta la metodología para llenar cada uno de los puntos necesarios para crear un estándar de documentación y de estrategia de desarrollo.

3.3.1.1 Análisis de estatus actual.

El objetivo en este paso es identificar la arquitectura que se tiene actualmente del DW en GAMESA, verificando como se encuentra y definiendo un diagrama generalizado que incluya el estado actual con su modelo lógico de información.

3.3.1.2 Utilización de Ciclo de Vida de Kimball en combinación con elementos de la metodología de Inmon

Para poder crear la FIC utilizaré la arquitectura propuesta por Inmon de la cual se tomarán los elementos indispensables para poder llegar a cumplir con los requerimientos de información de la gente funcional del área de Ventas y dejar listo además un modelo incremental en el cual sea sencillo incorporar las demás áreas de negocio.

Como metodología de la arquitectura propongo utilizar el método de Ralph Kimball [Kimball y Reeves y Ross, 98] el cual incluye todas las fases importantes para poder lograr el objetivo a alcanzar.

Toda la metodología empleada formará una base sólida del concepto de metadata la cual es muy importante para cumplir con los objetivos de esta tesis.

3.3.1.2.1 Plan de proyecto

Para poder realizar el plan del proyecto me basé en el modelo básico propuesto como ciclo de vida de desarrollo de sistemas en forma de cascada y espiral [Ruble, 97] tomando como base los conceptos importantes y característicos de una FIC para desarrollar una metodología robusta que permita considerar todos los puntos necesarios para un buen desarrollo de este tipo de estrategias.

El plan de proyecto se establecerá en dos fases:

- 1) Fase de Desarrollo de Back End
- 2) Fase de Desarrollo de Front End

Cada fase esta dividida en subfases las cuales se globalizan en macro actividades identificadas por:

- **Análisis**

- **Para el Back End**

Dentro de esta macro actividad se desglosan las actividades que incluyen desde el entendimiento del estatus actual, de las fuentes, la obtención de los requerimientos por parte del usuario, la arquitectura y finalmente el desarrollo formal del proyecto.

- **Para el Front End**

Aquí se incluyen las actividades pertenecientes al entendimiento de la forma de explotar la información, los roles de usuarios, la cantidad de usuarios, sus necesidades de información.

- **Diseño**

- **Para el Back End**

En esta macro actividad se incluyen las actividades que se encargan de la modelación lógica de la FIC. Se trabajan los modelos de ODS, DW, DM así como sus correspondientes Hechos y Dimensiones. Es la parte en donde se realiza el mapeo campo a campo de las fuentes hacia la fábrica de información.

- **Para el Front End**

Aquí se incluyen las actividades pertenecientes al diseño de los reportes que permitirán a los usuarios realizar su explotación de información.

- **Construcción**

- ↳ **Para el Back End**

Aquí es donde se realiza el pase del modelo lógico al modelo físico por medio de la herramienta modeladora seleccionada. Además se incluyen las actividades de construcción de la fábrica de información por medio de la correspondiente ETL.

- **Para el Front End**

Se incluyen las actividades de desarrollo de reportes por medio de la herramienta seleccionada.

- **Pruebas**

- **Para el Back End**

Desglose de actividades las cuales permitirán asegurar que las ligas y los mapeos realizados a los sistemas transaccionales sean los correctos.

- ↳ **Para el Front End**

Incluye las actividades que permitirán validar la información con el usuario final a través de los reportes establecidos.

- **Implementación**

- **Para el Back End**

Actividades para llevar a cabo la liberación.

- **Para el Front End**

Actividades para llevar a cabo la liberación.

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 creación, 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.

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

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.

DIRECCIÓN GENERAL DE BIBLIOTECAS

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.

DIRECCIÓN GENERAL DE BIBLIOTECAS

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.

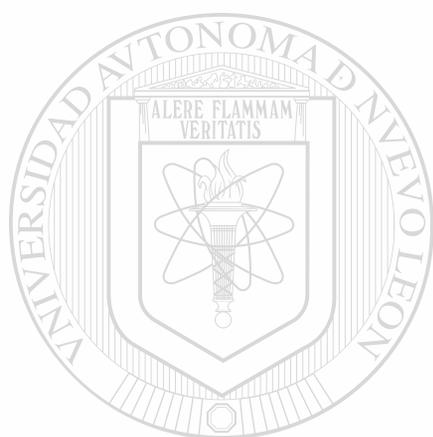
3.3.4 Mejoramiento en la Explotación

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.



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



DIRECCIÓN GENERAL DE BIBLIOTECAS

4 DESARROLLO DEL MÉTODO

4.1 Análisis de estado actual.

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

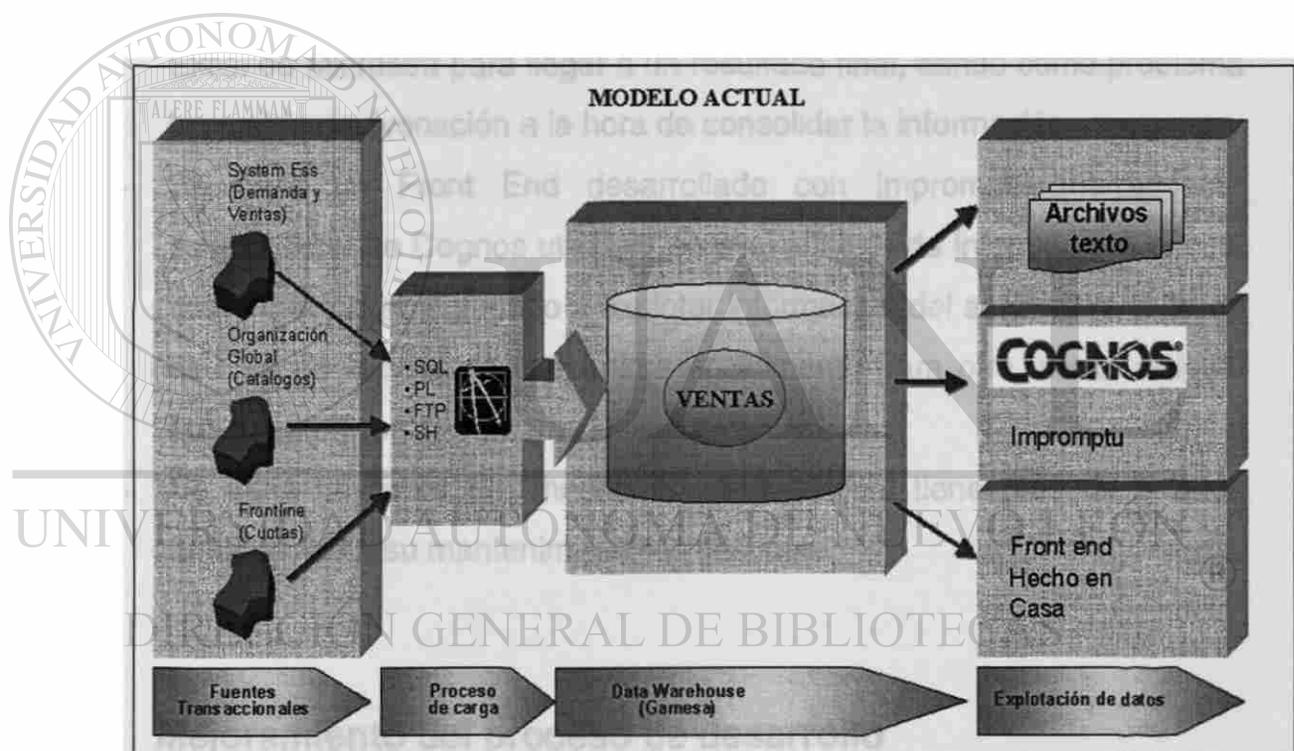


Figura 4.1 Diagrama de situación actual

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.

DIRECCIÓN GENERAL DE BIBLIOTECAS

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.

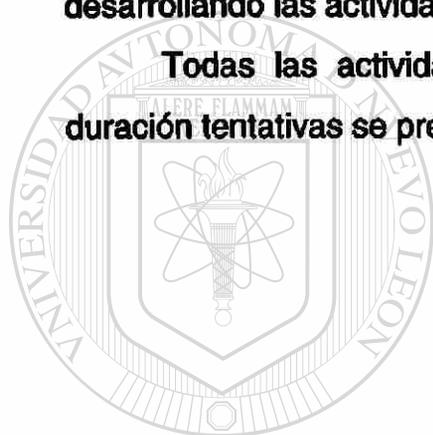
- g. El proceso debe de reflejar una arquitectura incremental. Se busca la estandarización de iteraciones para que los desarrollos subsecuentes sean más sencillos de incorporar.

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.



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

DIRECCIÓN GENERAL DE BIBLIOTECAS



ID	Task Name	Duration	Start	D1						
				T	W	T	F	S		
1	FIC-Iteración Ventas	71 days	Tue 6/1/01							
2	BACKEND	65 days	Tue 6/1/01							
3	ANÁLISIS	5 days	Tue 6/1/01							
4	Análisis de Estatus actual	5 days	Tue 6/1/01							
6	Programación Juntas y Entrevistas	5 days	Tue 6/1/01							
8	Identificación de las fuentes	5 days	Tue 6/1/01							
10	DISEÑO	19 days	Tue 6/6/01							
11	Diseño de modelo general (Inmon)	3 days	Tue 6/6/01							
13	Metodología Kimball	9 days	Fri 6/11/01							
14	Construcción de matriz de bus	1 day	Fri 6/11/01							
16	Realización del modelado dimensional	3 days	Mon 6/14/01							
20	Complemento selección de fuentes	1 day	Thu 6/17/01							
22	Realización de mapeo campo a campo	4 days	Fri 6/18/01							
24	Desarrollo de modelo lógico-físico	2 days	Thu 6/24/01							
26	Realización de dimensionamiento	1 day	Mon 6/28/01							
28	Diseñar la arquitectura a utilizar	2 days	Tue 6/28/01							
30	Diseñar mapa de configuración de procesos	2 days	Thu 6/31/01							
32	CONSTRUCCIÓN	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 interfaces	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	PRUEBAS	9 days	Mon 7/2/01							
47	Pruebas individuales de procesos	3 days	Mon 7/2/01							
48	Pruebas individuales de costos	3 days	Thu 7/5/01							
49	Pruebas integrales	3 days	Tue 7/10/01							
50	IMPLEMENTACIÓN	2 days	Fri 7/13/01							
51	Actividades de liberación	2 days	Fri 7/13/01							
52	FRONTEND	16 days	Tue 7/17/01							
53	Apoyo en análisis de Back End para entender requerimientos	1 day	Tue 7/17/01							
54	DISEÑO	6 days	Wed 7/18/01							
55	Investigar los accesos de los reportes (quién los lee, cuántos usuarios)	3 days	Wed 7/18/01							
56	Diseño tentativo de reportes que se requerirán	3 days	Mon 7/23/01							
57	CONSTRUCCIÓN	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	PRUEBAS	3 days	Thu 8/2/01							
62	Pruebas de reportes contra información	3 days	Thu 8/2/01							
63	IMPLEMENTACION	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.

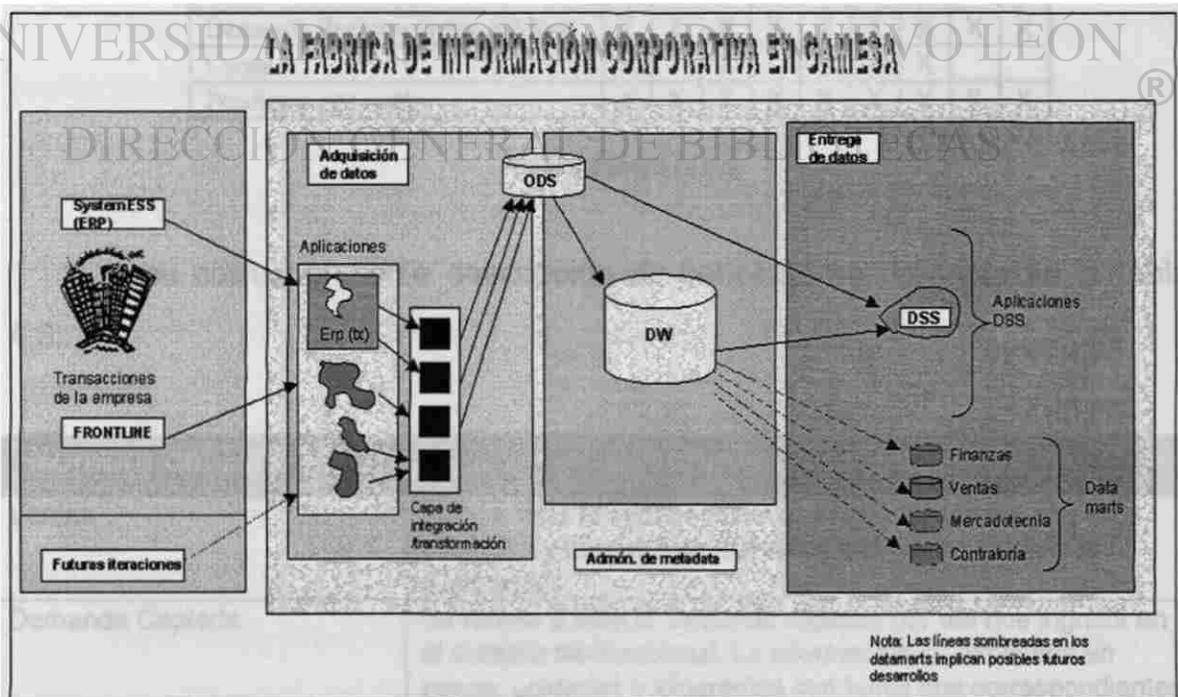


Figura 4.3. FIC propuesta

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.

DIRECCIÓN GENERAL DE BIBLIOTECAS

Fuente	Área de negocio responsable	Responsable 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	Responsable 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.

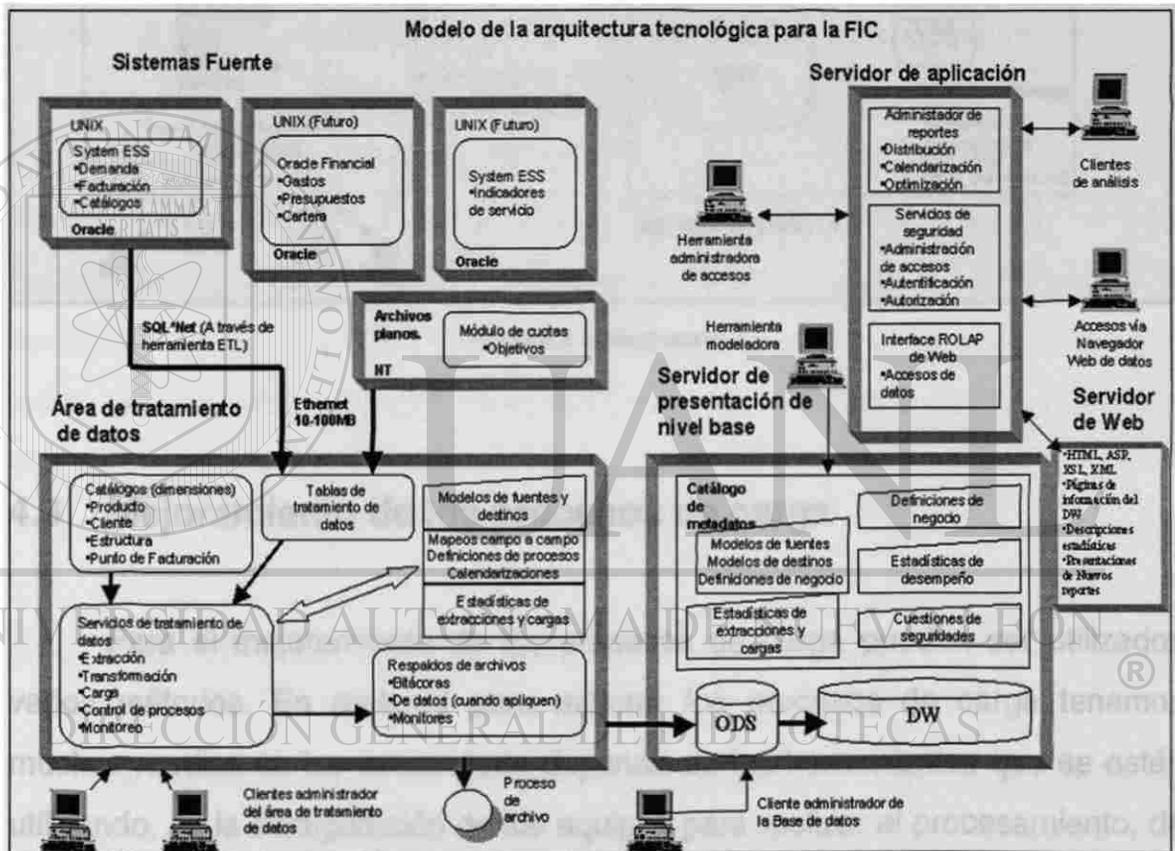


Figura 4.4. FIC arquitectura propuesta

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

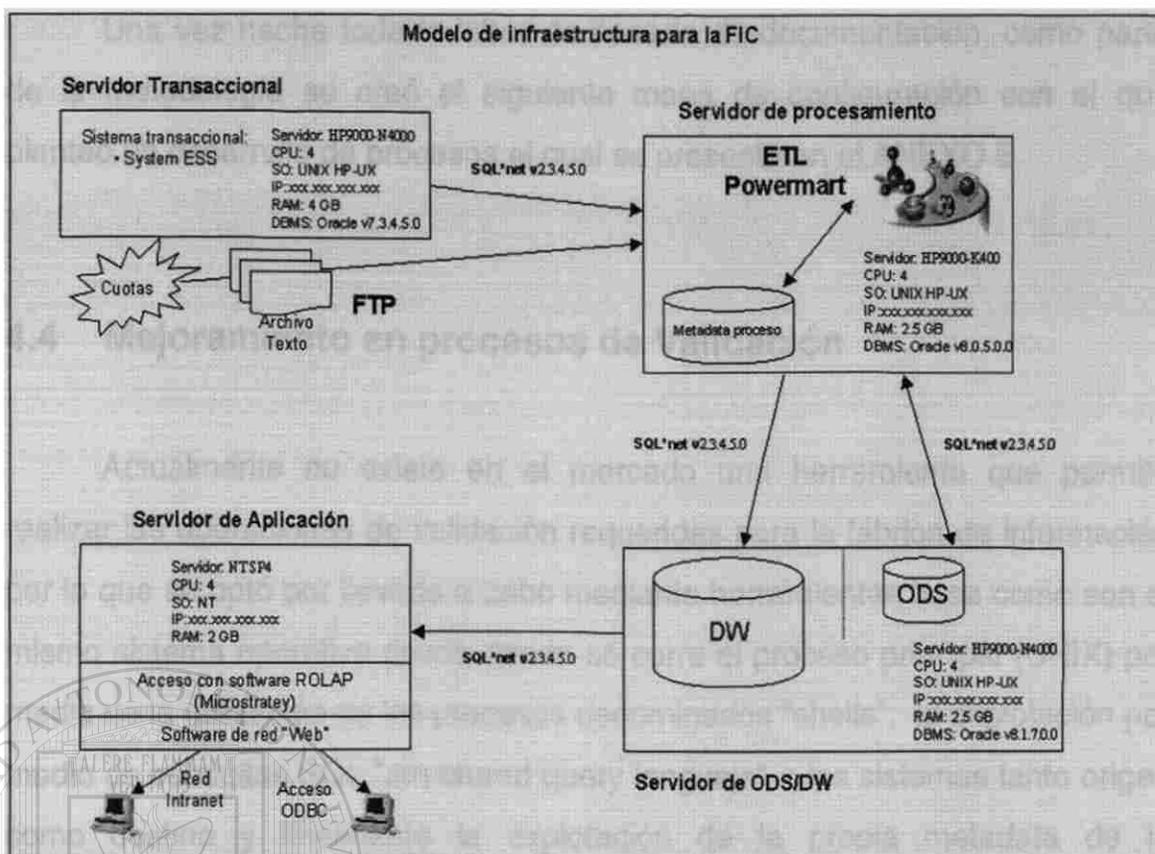


Figura 4.5. Infraestructura

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.

Una vez hecha toda la labor de llenado de documentación, como parte de la metodología se creó el siguiente mapa de configuración con el que planteó un desarrollo de procesos el cual se presenta en el ANEXO E

4.4 Mejoramiento en procesos de Validación

Actualmente no existe en el mercado una herramienta que permita realizar las operaciones de validación requeridas para la fábrica de información por lo que se optó por llevarla a cabo mediante herramientas base como son el mismo sistema operativo desde donde se corre el proceso principal (UNIX) por medio de la utilización de los procesos denominados "shells", la explotación por medio de consultas SQL "structured query lenguaje" a los sistemas tanto origen como destino y finalmente la explotación de la propia metadata de la herramienta ETL para obtener información por medio de consultas SQL a sus tablas propietarias y conocer características de las sesiones como tiempos de proceso, horas de corrida, etcétera.

4.4.1 Bitácora de proceso general

La llamada a todos los procesos principales de la carga de la fábrica de información corporativa se realiza mediante un programa principal en UNIX.

La bitácora desarrollada tendrá las características mostradas en la tabla

4.7

Comentario de proceso	Fecha y hora	Explicación
>>> INICIA PROCESO,		Bandera de proceso. Marca el inicio.
>>> Revisando Servicios de XXX,		Revisión de servicios (donde XXX puede ser el servicio de la base de datos o de la herramienta Powermart)
>>> Servicios de XXX OK. ó >>> Levantar Servicio XXX.		Bandera de estatus de servicios. Si el servicio está arriba entonces OK sino, un mensaje será enviado a la bitácora y a los operadores.
>>> Inicia Carga de YYY,		Hora de inicio de carga YYY. Donde YYY puede ser de catálogos, facturación, cuotas o demanda o cualquier concepto de negocio.
>>> Cargando YYY,		Tiempo en que inicia la carga específica del módulo.
>>> Finaliza Carga de YYY, ó >>> Problema en Carga de YYY,		Bandera de estatus de carga. Finaliza indica que todo estuvo correcto y si aparece algún error la palabra problema se le asignará a la bandera.
>>> PROCESO TOTALMENTE TERMINADO,		Bandera de finalización de proceso

Tabla 4.7 Cuerpo de bitácora principal

Los comandos necesarios dentro del proceso principal serán:

```
echo ">>> INICIA PROCESO, $(date)<<<" > /tmp/maintrade.stat
```

donde:

echo = Comando de UNIX que da como salida la información que necesitamos.

\$(date) = La fecha en un formato "Día Mes #día HH:MI:SS CST YYYY"

ejemplo: "Sun Mar 3 05:00:00 CST 2002"

" = Encierran el comentario a agregar a la bitácora

> = Símbolo de redireccionamiento a un tipo de salida específica, en este caso se direcciona al archivo de bitácora maintrade.stat

4.4.2 Bitácora de datos

En esta bitácora se utilizan sentencias "SQL" para generar la información a un nivel de detalle muy pequeño agrupando por totales por día sobre los conceptos importantes a validar.

La estrategia planteada es:

- Creación de tabla temporal MON_TRD_CARGADIARIA. Los valores a almacenar se muestran en la tabla 4.8.

Campo	Tipo	Explicación
FUENTE	VARCHAR2(7)	La fuente pudiendo ser en este caso ESS, DW, ODS
CONCEPTO	VARCHAR2(11)	El concepto sobre el cual actúan los rubros a mostrar. Por ejemplo: FACTURAC que equivale a facturación, PCAPTADO que equivale a pedido captado, etcétera.
FECHA	DATE	El día a validar
UNIDADES_BRUTAS	NUMBER	Sumarizado por día del concepto
UNIDADES_NETAS	NUMBER	Sumarizado por día del concepto
KILOGRAMOS_BRUTOS	NUMBER	Sumarizado por día del concepto
KILOGRAMOS_NETOS	NUMBER	Sumarizado por día del concepto
PESOS_BRUTOS	NUMBER	Sumarizado por día del concepto
PESOS_BRUTOS_INTERMEDIOS	NUMBER	Sumarizado por día del concepto
PESOS_NETOS	NUMBER	Sumarizado por día del concepto
PESOS_ESPECIAL	NUMBER	Sumarizado por día del concepto
PESOS_PROMOCIONAL	NUMBER	Sumarizado por día del concepto
PESOS_LINEA	NUMBER	Sumarizado por día del concepto

Tabla 4.8 Conceptos a almacenar para validación de la carga diaria

Es importante mencionar que este tipo de tablas se deberán crear dependiendo de cada desarrollo específico por iteración de la fábrica de información corporativa.

- Creación de programa "SQL" especial para extraer la información de los sistemas tanto fuente (el transaccional comúnmente) como de los destinos (comúnmente la FIC). Esta información se escribirá en la tabla anteriormente mencionada.

Al igual que en el punto anterior la forma en que se realice la explotación de información dependerá del sistema origen y por consiguiente de la iteración específica.

- Una vez llena la tabla se procedió a explotar la información realizando un formateo con los campos presentados en la tabla 4.9.

Campo	Explicación
FECHA	Es la fecha de carga
CONCEPTO	Es el concepto de negocio a validar (Facturación, Demanda, Cuota)
RUBRO	Es el rubro específico a validar (Kilos, Pesos, Unidades tanto brutas como netas)
ORIGEN	Sistema origen contra el que estamos comparando
DESTINO	Sistema destino
DIFERENCIA	Total de diferencia.

Tabla 4.9 Conceptos a presentar en bitácora de carga

4.4.3 Bitácora de procesos individuales

Este tipo de bitácora es muy específica para el tipo de herramienta que estamos utilizando ya que se utiliza principalmente gracias a la funcionalidad de metadata que guarda la herramienta de ETL en una base de datos definida y que nos permite tomar la información de las corridas de procesos.

Lo que se requiere para esta bitácora es:

- **Definición de tres tipos de apartados para la bitácora**
 - **Apartado 1:** Todos los procesos que estén incluidos dentro del mapa de configuración
 - **Apartado 2:** Procesos principales en donde se realice alguna carga al ODS o al DW
 - **Apartado 3:** Procesos con error tanto de BD o de cualquier otro tipo.
- **La información de los procesos se presentará con los campos presentados en la tabla 4.10.**

Campo	Explicación
TABLA	Es la tabla de donde se escriben los registros
INSTANCIA	Si la misma tabla aparece en el mismo proceso, entonces se crean instancias de ésta para diferenciarla.
CARGADOS	Total de registros Cargados
RECHAZADOS	Total de registros Rechazados
SESION	Nombre de la sesión
INICIA	Fecha y hora en que inicia el proceso
FINALIZA	Fecha y hora en que finaliza el proceso
MINUTOS	Minutos transcurridos desde que inicial el proceso a que termina.

Tabla 4.10 Conceptos a presentar en bitácora de procesos

- **Finalmente se creó un programa que accesa la información como se muestra en la tabla 4.10.**

4.5 Mejoramiento en la Explotación

La estrategia de explotación se plantea en la figura 4.6.



Figura 4.6. Estrategia de explotación

El objetivo principal de ésta arquitectura de Front End es el de buscar evitar los problemas actuales existentes en la distribución de información que se originan primordialmente con la existencia de las múltiples fuentes de datos del mismo DW en herramientas de presentación diversificadas y muchas veces con curvas de aprendizaje demasiado grandes, además evitaremos que los usuarios reprocesen los datos obtenidos por los diferentes medios con lo que se busca eliminar la cantidad de errores manuales que se pueden dar.

La estrategia se centra en la implementación de una plataforma denominada "Microstrategy" la cual apoyada sobre la base de metadata y del ODS y DW, permitirá el acceso a la información de una forma ordenada y con la disminución de tiempos de reproceso de los usuarios permitiéndoles dedicarse a su tarea o función que es principalmente la de analistas de información.

5 RESULTADOS

5.1 Mejoramiento en proceso de desarrollo

Para poder medir el grado de cumplimiento al aplicar la metodología fué necesario realizar una investigación interna sobre los resultados en tiempos de desarrollo del proyecto inicial en GAMESA sobre la creación del original DW.

El tiempo total utilizado en el año de 1996 fue de 375 días aproximadamente los cuales al tratar de distribuirlos sobre cada una de las actividades que se proponen en la metodología, esto permite obtener un tiempo de equivalencia el cual se aplica para medir el grado de cumplimiento. Cabe mencionar que de los 375 días sólo se tomarán 290 que son los tiempos comparativos de las fases de análisis, diseño y construcción.

Los tiempos del proyecto inicial fueron tomados a partir del momento en que se aceptó el proyecto de DW como parte de la solución de la compañía.

Los resultados obtenidos a raíz de la aplicación de la metodología en cuestión de tiempos, comparando los diferentes módulos se presentan a continuación.

- **Tiempos en la etapa de análisis**

Dentro de esta etapa de análisis se consideraron los puntos importantes sobre el entendimiento de la situación actual, la programación y ejecución de las juntas y las entrevistas necesarias, así como el entendimiento de los conceptos a plasmar en la FIC y la identificación de las posibles fuentes. Los resultados se presentan en la tabla 5.1.

Tiempo de duración (Proyecto Original)	Tiempo de duración (FIC)
28 días	5 días

Tabla 5.1 Comparativo Análisis Proyecto Original - FIC

- **Mejoramiento en etapa de diseño**

Para la etapa de diseño se estableció la propuesta de estrategia de FIC sobre la anterior de DW en la cual se definieron las formas de llevar a cabo el modelo, la aplicación de la metodología Kimball para el modelado dimensional y el mapeo campo a campo, se diseñó el modelo tanto en su forma lógica como física, el dimensionamiento y la parte de diseño de arquitectura de infraestructura y el diseño y la configuración de interfases.

- Para el Back End.

Los tiempos se presentan en la tabla 5.2.

Tiempo de duración (Proyecto Original)	Tiempo de duración (FIC)
20 días	19 días

Tabla 5.2 Comparativo Diseño BE Proyecto Original - FIC

Aquí se presentó un pequeño desfase de tiempos con respecto al tentativamente estipulado debido principalmente al entendimiento formal del modelado dimensional y a su correcta aplicación.

Es importante mencionar que el planteamiento del modelo ya está realizado por lo que en sucesivas iteraciones ya no será necesario utilizar tanto tiempo.

- Para el Front End.

En esta parte se comparó solo la estrategia de creación de indicadores en la nueva herramienta y sólo 2 reportes para la explotación de la FIC. Los tiempos se presentan en la tabla 5.3.

Tiempo de duración (Proyecto Original)	Tiempo de duración (FIC)
42 días	6 días

Tabla 5.3 Comparativo Diseño FE Proyecto Original - FIC

La diferencia en tiempos se presenta debido a que no existía en su momento una herramienta especializada, sino que tuvieron que diseñar una para poder presentar información al usuario.

DIRECCIÓN GENERAL DE BIBLIOTECAS

- Mejoramiento en etapa de construcción

En esta etapa es donde se notó una gran mejoría que se vió reflejada principalmente por la aplicación de la herramienta especializada la cual disminuyó considerablemente los tiempos de construcción tanto para el Back end (mostrado en la tabla 5.4.) como para el Front End (tabla 5.5.) ya que todas las banderas, validaciones, alarmas y demás cuestiones tenían que ser creadas a través de métodos codificados los que dificultaron el rápido desarrollo originalmente.

- Para el Back End.

Tiempo de duración (Proyecto Original)	Tiempo de duración (FIC)
120 días	20 días

Tabla 5.4 Comparativo Construcción BE Proyecto Original - FIC

- Para el Front End.

Tiempo de duración (Proyecto Original)	Tiempo de duración (FIC)
80 días	5 días

Tabla 5.5 Comparativo Construcción FE Proyecto Original - FIC

- **Etapas de pruebas:**

Finalmente el establecimiento de una etapa de pruebas a diferentes niveles con respecto al desarrollo original representó una mejor aceptación en la futura implementación debido a la casi inexistencia de errores. En la tabla 5.6 se muestran los resultados para el Back End y en la tabla 5.7 para el Front End.

- Para el Back End

Tiempo de duración (Proyecto Original)	Tiempo de duración (FIC)
10 días	9 días

Tabla 5.6 Comparativo Pruebas BE Proyecto Original - FIC

- Para el Front End

Tiempo de duración (Proyecto Original)	Tiempo de duración (FIC)
15 días	3 días

Tabla 5.7 Comparativo Pruebas FE Proyecto Original - FIC

- Tiempo de implementación

Este punto, aún y cuando no era parte del objetivo de la tesis se plantea como comparativo en la cual se trabaja la implementación con el usuario aún y cuando ésta no se hizo a nivel nacional ya que solo se realizó con el superusuario de la información. Los resultados se muestran en la tabla 5.8.

Tiempo de duración (Proyecto Original)	Tiempo de duración (FIC)
60 días	3 días

Tabla 5.8 Comparativo Implementación Proyecto Original - FIC

Uno de los puntos importantes obtenidos de aplicar la metodología es que todo lo desarrollado queda postrado sobre la base una metadata sólida y confiable que permitirá en un momento dado servir de armas para poder continuar con la construcción de la FIC en sentido incremental.

El comparativo de los tiempos del desarrollo actual contra el anterior se presenta en la figura 5.1.

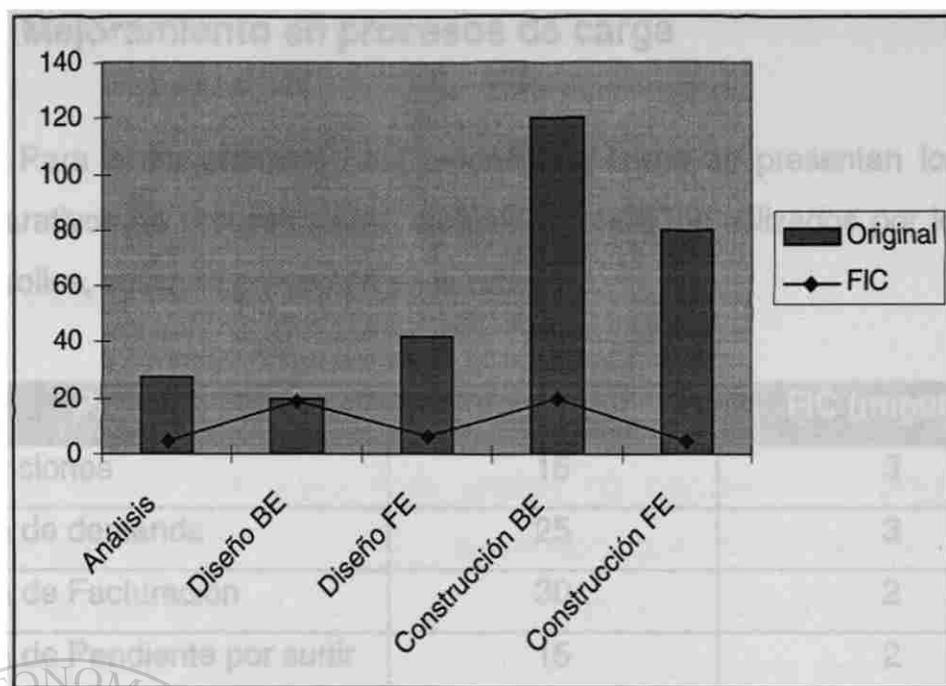


Figura 5.1. Gráfica comparativa de tiempos de proceso de desarrollo Original - FIC

Finalmente se resume que originalmente se utilizaron 290 días para las actividades de análisis, diseño y construcción las cuales se disminuyeron a 55 para un proyecto de las mismas dimensiones en una arquitectura de FIC.

5.2 Mejoramiento en procesos de carga

Para el mejoramiento del proceso de carga se presentan los tiempos comparativos de procesamiento actuales, contra los utilizados por los nuevos desarrollos, estos se presentan en la tabla 5.9.

Procesos	Original (minutos)	FIC (minutos)
Dimensiones	15	3
Carga de demanda	25	3
Carga de Facturación	30	2
Carga de Pendiente por surtir	15	2
Carga de cuotas	15	5

Tabla 5.9 Comparativo tiempos de proceso Proyecto Original - FIC

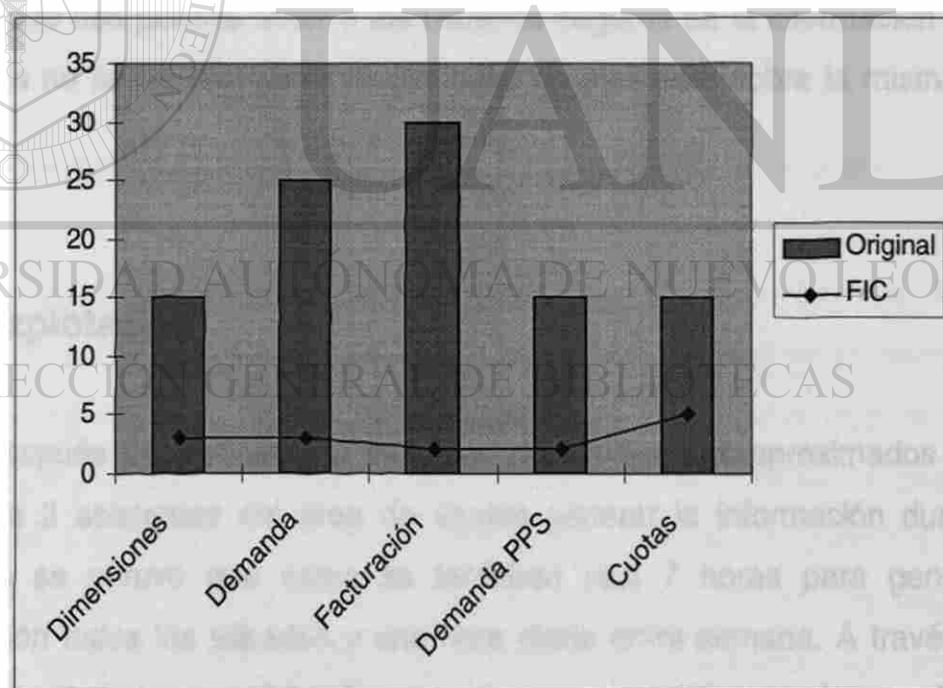


Figura 5.2. Gráfica comparativa de tiempos de proceso de carga Original - FIC

La información recopilada de los tiempos anteriores se obtuvo por observación de 30 días con respecto a los procesos originales y con un paralelo sobre la nueva arquitectura. En la figura 5.2 se puede observar la gráfica

comparativa de tiempos de proceso de la carga del desarrollo original contra el nuevo.

5.3 Validación

A través de una bitácora generada a partir de Agosto hasta Diciembre del 2001 se obtuvo una relación de errores no reportados de proceso y de información cargada contra los detectados por medio de las bitácoras obteniendo que de las 8 veces que se reportó un problema en ese transcurso de tiempo en el sistema original, en la FIC ya se había detectado antes y tomado las acciones correctivas necesarias.

Los resultados quizás no se presenten tan fácilmente en números pero lo desarrollado nos permite tener a los usuarios seguros de la información ya que a la fecha no se ha levantado ningún caso de problema sobre la misma en el área.

Después de realizar una medición de los tiempos aproximados que le tomaba a 2 asistentes del área de ventas generar la información durante 6 semanas se obtuvo que éstos se tardaban casi 7 horas para generar la información todos los sábados y una hora diaria entre semana. A través de la nueva estrategia la consolidación se realiza en automático por lo que de las 7 horas que dedicaban anteriormente sólo se llevan 2 minutos en generar la información.

6 CONCLUSIONES Y RECOMENDACIONES

6.1 Conclusiones

Se presentan las conclusiones identificándolas para cada uno de los objetivos buscados en el proyecto y finalmente un apartado de conclusiones generales.

6.1.1 Para el mejoramiento en el proceso de desarrollo

Un punto importante a discutir es el grado de cumplimiento de este objetivo. Algo que queda muy claro es que tomando en cuenta los tiempos de proyecto que tomó inicialmente con respecto a los tiempos nuevos con la metodología propuesta podemos llegar a concluir que la hipótesis H1 planteada al inicio de la tesis se cumple ya que a raíz de un tiempo de desarrollo de 290 se logró disminuir a 55 días lo cual se refleja en un 81.03% de mejora en tiempos logrando además una documentación adecuada (la cual era inexistente) y una metodología formal en la cual se plantea disminuir los tiempos de futuras iteraciones.

A raíz de este desarrollo además podemos concluir una serie de cuestiones que me gustaría comentar:

- Existe un indicador que no propuse para medir el grado de cumplimiento de este objetivo debido a que este consiste en algo no mensurable, este mide la ventaja que puede traernos el hecho de crear una documentación formal sobre todos y cada uno de los pasos dentro del desarrollo de una solución ya sea de DSS o de cualquier otra área de desarrollo de proyectos. Si nos remontamos a los problemas que existían en la compañía inicialmente y que formaban el porqué crear un concepto como lo es la FIC podemos verdaderamente pensar en que sin una base firmemente documentada no es posible realizar cambios de una manera sencilla, además de que dependeríamos de una persona ya que todo el conocimiento recaería en ella, si esa persona se va, entonces el proyecto pudiera desmoronarse. Soy muy enfático en este punto debido a que yo entré a esta área visualizando una creciente necesidad por parte de las diferentes áreas del negocio de tener un rápido crecimiento de lo que se podía decir se tenía como solución para el diseño de soporte a la toma de decisiones, pero debido a la falta de documentación, a la falta de una metodología y finalmente de una arquitectura robusta que permitiera este rápido desarrollo nos llevó casi 1 año empezar a trabajar sobre nuevos desarrollos. Es un hecho que después de implantar el concepto de FIC en la compañía el desarrollo de iteraciones ha ido en aumento.
- Actualmente utilizando esta metodología hemos empezado a crecer la FIC incrementando el alcance en el área de ventas, pero además incluyendo nuevas iteraciones como son indicadores de servicio, indicadores de análisis de mercado, indicadores de gastos, etcétera. La lista sigue creciendo. Lo más importante es que hemos incrementado la cantidad de indicadores necesarios para el negocio y la facilidad con la que lo hemos hecho ha dependido ahora únicamente del análisis ya que con la metodología aplicada el diseño y la construcción se vuelve muy sencilla.

- Un punto importante es que la metodología no es una solución mágica que va a dar resultados inmediatamente. Es necesario que la gente que entre al área tenga algunos conocimientos básicos de DW y manejo de bases de datos ya que ésta es la base para un rápido desarrollo.
- Los tiempos de desarrollo son relativos al tamaño del proyecto, pero lo que aseguramos es que por medio de la metodología los tiempos de desarrollo de nuevas iteraciones sean menores siempre y cuando se aplique correctamente.

6.1.2 Para el mejoramiento en el proceso de carga

A raíz de la organización del proceso de carga y de la realización de las interfases por medio de una herramienta especializada pudimos constatar que cumplimos con el objetivo de mejoramiento en estos procesos debido a que de los 100 minutos que se tardaba la carga de datos se disminuyó a 15 lo cual refleja un 85% de mejora con respecto al tiempo original.

Dentro de los puntos importantes cabe mencionar:

- El mejoramiento se obtuvo debido a varios factores, uno de ellos principalmente fue la herramienta utilizada. En la actualidad existen muchas herramientas en el mercado las cuales tienden a realizar las mismas funciones, algunas un poco mejor que otras, pero todas tienen un mismo concepto en común, evitamos la realización de codificación difícil de mantener por medio de una interfaz gráfica que permita organizar los procesos de tal forma que estos sean fácilmente entendibles y que guarden todo el desarrollo dentro de una metadata. Si una herramienta no posee estas características, entonces no es una herramienta útil para el objetivo

que estamos buscando ya que estaríamos cayendo en los problemas que se tenían originalmente:

- La misma herramienta da origen a una nueva forma de organizar los procesos. El mapa de configuración realizado no es más que una concepción de todos los procesos que de otra forma tendrían que estar codificados en programas a los cuales todos los servicios requeridos por el área de Back end tendrían que acoplárseles por medio de codificación especial. Una herramienta ETL nos permite concebir a los procesos como objetos encapsulados que en este caso se denominan mapeos.
- Uno de los puntos más importantes que quizás no se visualizaron dentro del concepto principal de la tesis fueron la optimización de los tiempos de construcción de interfases. Estos tiempos son dependientes de muchos factores, pero principalmente algo que los viene a mejorar enormemente es la creación de la documentación previa (mapeo campo a campo, entendimiento de las fuentes) que vienen a resumir el trabajo en realiza el análisis y después solo traducirlo en procesos. Aún así se observó que los tiempos originales de construcción con respecto a los nuevos se mejoraron en un 83.3%.
- Otra cuestión que también es importante mencionar es que el hecho de haber mejorado los tiempos de cargas en GAMESA puede aplicarse en otros lugares o en otras iteraciones, pero es importante tener en cuenta que un mejoramiento puede partir desde que se mejora la forma de estructurar los procesos (mapa de configuración), pero también depende del hardware en donde van a correr los procesos y muchas veces del sistema operativo que se tiene. Es posible tener la misma herramienta ETL en 2 sistemas operativos diferentes (NT y UNIX) por ejemplo y tener algunas mejoras considerables que van desde el mejor manejo de memoria, hasta la optimización de recursos internos. Es importante comentar que los tiempos obtenidos son con una infraestructura definida para las necesidades de GAMESA.

- Otro punto a considerar es el DBMS el cual juega un importante rol cuando se empiezan a manejar grandes volúmenes de información, y además la forma en que se diseñó la arquitectura de la base de datos cuando se utiliza la modelación dimensional en ambiente relacional como fue en este caso tiende a afectar los tiempos de carga y más aún los tiempos de explotación.
- La forma de estructurar los procesos con el mapa de configuración sirvió en la empresa para mejorar los tiempos, pero es importante recalcar que si hubiéramos tenido que partir del mismo desarrollo pero sin herramientas especializadas quizás también hubiéramos optimizado los tiempos de proceso, pero no hubiéramos creado el pilar para complementar el concepto de la FIC, que consiste en la metadata de todo esto y mejor aún, en la inclusión de todos los servicios necesarios para satisfacer todos los movimientos de datos.

6.1.3 Para el mejoramiento en el proceso de validación

Éste es quizás el punto más difícil de validar en esta tesis. El porqué es debido a que previamente el nivel de validación que se tenía era casi nulo y no existía una estrategia formal en la cual se pudieran apoyar para poder detectar errores. La forma en que se atacó, aunque suene un poco obvia, muchas veces no se aplica en las compañías por razones que van desde la imposibilidad de poder implementarlas debido a la forma en que organizan sus procesos, la falta de conocimiento, los desarrollos a vapor, etcétera.

La forma en que opté por medir el grado de cumplimiento fue comparando las veces que tuvimos problemas no detectados a tiempo con las veces que si se detectaron a partir de que se implementaron las bitácoras comprobándose en el corto tiempo que tienen funcionando que constituyen un importante indicador para la correcta y oportuna detección de errores. Muestran un mejoramiento sustancial el cual se puede comprobar por los cero reportes obtenidos hasta la fechas sobre los datos.

Puntos importantes a concluir son:

- **Aún y cuando este sea un punto básico a realizar en cualquier tipo de desarrollo, es increíble tener muchas veces tan poco desarrollado sobre este rubro, tal es el caso como en GAMESA.**
- **La forma en que se atacó este punto fue una estrategia definida e implementada basándonos en consejos de Inmon y Rudin [Inmon y Rudin, 98] pero lo más importante es que sí funciona y es un punto que no debe menospreciarse.**
- **Los puntos que se pudieron atacar se basan básicamente en la integridad de la información y en la confianza que los usuarios tengan sobre la misma. Casos como que un día se presentara un número y al día siguiente este mismo cambiara para la misma fecha generó problemas enormes en la confianza de los usuarios, además otro punto importante es que no se tenían medios para comprobarles si los errores existían debido a errores de carga, o errores en los datos fuente.**
- **Un concepto que se debe de tomar en cuenta que nos obligó a crear esta estrategia de validación el cual lo he vivido en carne propia es el enfrentamiento contra un sistema origen el cual, como todo sistema de información, permite a la gente meter mano sin siquiera informar o enterar a los sistemas involucrados de los cambios. En un concepto de ODS y DW es imprescindible estar enterado de estos cambios ya que la estrategia de carga generalmente está basada en criterios de fechas, por lo que fechas pasadas usualmente no se recargan al menos que se realice un requerimiento formal de esto. Si no se tienen las armas para poder demostrar los cambios en el sistema origen y a la hora que los usuarios empiezan a comparar información, es en este momento cuando se ven los problemas.**

6.1.4 Para el mejoramiento en el proceso explotación

Los resultados obtenidos en el mejoramiento de procesos por medio de la estrategia definida, verdaderamente trajeron cambios contundentes basados en los análisis de tiempos de ahorro presentados. Desde el hecho de que de 7 horas se baje a 2 minutos el proceso de explotación de información y además la posibilidad de que un Vicepresidente o un Director de la compañía tenga el acceso a la información sin necesidad de intermediarios es el principal beneficio obtenido de esta estrategia.

- Dentro de los puntos a considerar es que el proceso de explotación tiene diversos factores que permiten medirlos los cuales van desde los tiempos de respuesta dependiendo de la herramienta utilizada para esta explotación, la capacidad de navegación, uno muy importante que es la presentación de los datos 100% confiables, y el punto evaluado para esta tesis, que es la centralización de información que permita una explotación central y que disminuya los tiempos del proceso de explotación proponiendo una estrategia diferente de generación de información evitando la triangulación de la misma.
- Esta estrategia considero que debe ser utilizada en cualquier proyecto de DSS ya que si no, se seguiría teniendo el problema que se presentaba originalmente (la falta de integridad de la información) por posibles errores manuales o por la posible comparación de datos con criterios diferentes en diferentes sistemas o fuentes de datos (desde archivos textos generados, hojas de cálculo, etcétera).
- Lo difícil para lograr este objetivo llega cuando se requieren integrar las diferentes áreas con indicadores en común y que por consiguiente se requieren tener almacenados en la FIC. Un punto importante es que esto se logrará una vez que el concepto y el desarrollo de la FIC este muy robusto. Mientras tanto si es posible la integración pero tiene que utilizarse el

desarrollo incremental, tendiendo a atacar los indicadores principales del negocio y finalmente los adicionales.

6.1.5 Conclusiones en general

A raíz de todos los puntos trabajados podemos constatar que una metodología robusta, un buen conocimiento de lo que se tiene en mente pero además un ingrediente que no he mencionado y que considero sumamente importante y que viene a complementar lo antes mencionado es la visión que se tenga cuando se propone/diseña/construye una FIC.

Una FIC es un concepto que va más allá del simple almacenamiento de datos, este es el corazón y cerebro de la compañía el cual tiende a dirigir las diversas áreas del negocio las mismas que en conjunto hacen salir adelante a la compañía. Una compañía sin información en la actualidad, es una compañía con un problema serio que tarde que temprano será eliminada por los competidores que si tuvieron la visión y se armaron correctamente.

Una FIC es un concepto que también va más allá de la simple habilitación de información ya que debe de proporcionar información con valor agregado para la compañía y debe de proporcionar ventajas competitivas, y lo más importante, debe de ser lo suficientemente robusto para soportar las diversas áreas en conjunto.

Una FIC no se debe considerar como un sistema como muchos tienden a hacerlo, sino como una solución a la mayoría de los problemas que han surgido por el manejo de cantidades industriales de datos y su posterior transformación a información.

Como mencionaba anteriormente, esta metodología propuesta permitió atacar los problemas mencionados originalmente, podemos considerar que

ésta puede ser utilizada sin pensarse para diversas compañías y en otras áreas pero hay que considerar que el tamaño, la situación propia de la empresa, la forma en que se encuentra estructurada, etcétera, son los que rigen para poder repetir los mismos resultados obtenidos. Como también mencionaba anteriormente, este concepto actualmente está creciendo a pasos agigantados en la compañía, trayendo algunos nuevos problemas, pero algo que es importante recalcar es que, bien definidas las bases, lo demás es cuestión de acoplamiento mínimos para alcanzar el resultado esperado.

Un problema que ha presentado la mayoría de las iteraciones que hemos estado realizando y que hemos estado viviendo es que la metodología propone las bases, pero el punto crítico está en la integración de las fuentes debido a los diversos problemas internos que tienen cada uno de estos sistemas y que por cuestiones internas hacen difícil y conforman un escollo que puede poner en peligro la construcción de una FIC.

Finalmente como toda metodología siempre hay pasos adicionales que pueden acoplarse debido a que cada empresa es un mundo diferente pero pudimos constatar que teniendo como base los conceptos mencionados se tendrá un pilar (la metadata) que vendrá a soportar los futuros cambios que puedan avecinarse.

6.2 Recomendaciones

En el desarrollo de esta tesis se han tratado muchos aspectos relacionados con la implementación de la FIC, pero es importante recalcar que una FIC es una solución que soporta el desarrollo de soluciones para la toma de decisiones, pero el concepto como tal de FIC tiene muchos hallazgos, nuevas tendencias y nuevas formas de pensar que pueden complementar el desarrollo actual. Y más aún, la arquitectura de FIC no es la única solución

existente que puede solucionar los problemas del manejo de información en las empresas.

A continuación se listan algunos trabajos sobre los que se puede ahondar para futuras investigaciones:

- La construcción de un DW Distribuido y su impacto en el negocio.
 - La implementación del concepto de minería de datos a diversos niveles en donde se debe buscar desde las herramientas especializadas para llevarla a cabo, la correcta integración de los datos y la búsqueda del conocimiento.
 - El diseño de un Data Mart de Rentabilidad en donde se combine información de Ventas y Gastos para sacar las verdaderas ganancias
 - El concepto de Data WebHouse
 - El impacto de los medios de comunicación actuales sobre la distribución de información.
 - La utilización de las tecnologías SAN (Storage Area Network) y NAS (Network Attached Storage) en el ambiente de la FIC.
 - El diseño y la implementación de un ODS distributivo en las empresas para el manejo centralizado de la información y la correcta distribución de ésta.
-
- Un estudio en México sobre como las soluciones implementadas han tenido éxitos o fracasos, la determinación del porqué de éstos, y si fueron fracasos, entonces el cómo el concepto de FIC puede llegar a beneficiarlas.
 - Como se robustece el concepto de metadata por medio de la aplicación de la metodología regida por el concepto de cadena de valor.
 - El análisis del valor agregado que se obtiene por medio de la utilización del concepto de DW en la planeación estratégica.
 - Como lograr una distribución eficiente del conocimiento de lo que existe dentro de una FIC permitiendo al usuario aprender del negocio a partir de ésta.

Además, algunos conceptos que por cuestiones de tiempo o porque no formaban parte del objetivo de esta tesis no fueron atacados con más detalle pero que corresponden a piezas del todo en el rompecabezas de la FIC y que podrían ser buenos temas para futuras investigaciones son:

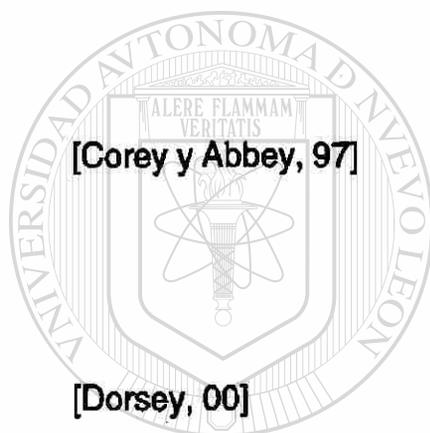
- Optimización de la organización tecnológica para el mejoramiento en los tiempos de respuesta tanto en la carga, como en la distribución y explotación de información en una compañía en particular.
- Las ventajas que se pueden obtener por medio del diseño del ODS con una arquitectura relacional.
- La correcta implementación de mecanismos de detección de datos no muy utilizados y su administración.
- El diseño y la implementación de una metodología para la correcta evaluación de herramientas de ETL para el buen funcionamiento del DW.
- Estrategias avanzadas para la validación de la información por medio de la implementación del concepto "calidad de datos" y la utilización de herramientas especializadas para este fin.
- El diseño y la implementación de una estrategia más avanzada con herramientas especializadas para el control y manejo de la metadata.

DIRECCIÓN GENERAL DE BIBLIOTECAS

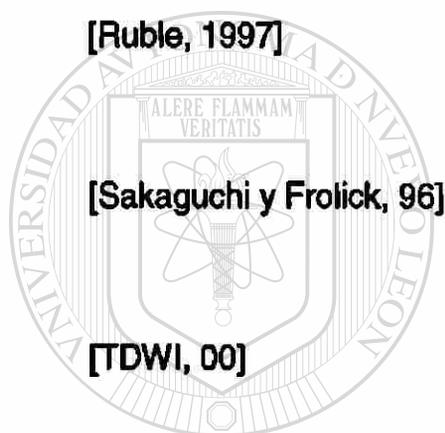
Como esta tesis partía de un concepto que ya la compañía había aceptado y que se optó por mejorar no fué necesaria una estrategia de definición de costo/beneficio y la temática de venta del mismo, pero un buen tema de investigación para empresas que no conozcan o que no hayan tenido la visión de crear una FIC podría ser el establecimiento formal de los indicadores que pudieran establecer el porqué una solución como el DW o más aún de la FIC beneficiaría a la compañía.

BIBLIOGRAFÍA

- [Anahory y Murray, 97] Anahory, Sam; Murray, Dennis; *"Data warehousing in the real world : a practical guide for building decision support systems"*; Pie imprent Harlow, England ✆ ISBN:0201175193; Addison-Wesley, 1997
- [Ball, 99] Ball, Leslie; *"What is an ERP?"*; Artículo; <http://intra.com.umass.edu/somit/public/pages/ERP%20Final%20Presentation/index.htm>; 1999
- [CAI, 2001] <http://www.cai.com>
- [Chen, 76] Chen, Peter; *"The Entity-Relationship Model Toward a Unified view of data"*; Artículo; Instituto Tecnológico de Massachusetts; 1976
- [Corey y Abbey, 97] Corey, Michael; Abbey, Michael; *"Oracle Data Warehousing, Guía práctica para analizar, construir e implantar con éxito un sistema Data Warehouse"*; USA; ISBN:0078822424; Oracle Press, Mc Graw Hill, 1997
- [Dorsey, 00] Dr. Dorsey, Paul; *"Data Warehouses, ad hoc query tools and other ways to destroy your company"*; Dulcian Inc.; 2000
- [DWOL, 01] <http://www.datawarehousingonline.com> ; The data Warehouse Institute on Line
- [Infoopt, 99] Informatica; *Conceptos de optimización*; 1999
- [INFORMATICA, 2001] <http://www.informatica.com>
- [Inmon e Imhoff y Sousa, 01] Inmon, W.H.; Imhoff, Claudia; Sousa, Ryan; *"Corporate Information Factory"*; Editorial Wiley; ISBN: 0-471-39961-2; USA; 2001
- [Inmon y Rudin, 98] Inmon, W.H.; Rudin, Ken; *"Data Warehouse Performance"*; Editorial Wiley; ISBN: 0471298085; USA; 1998
- [Inmon, 96] Inmon, W.H.; *"Building the Data Warehouse"*; Editorial Wiley; USA; 1996
- [Kimball y Reeves y Ross, 98] Kimball Ralph; Reeves, Laura; Ross, Margy; *"The Data Warehouse Lifecycle Toolkit"*; USA; ISBN: 0471255475; Wiley, 1998



- [Kimball, 96] Kimball Ralph; *"The Data Warehouse Toolkit"*; USA; ISBN: 0471153370; Wiley, 1996
- [MICROSOFT, 2001] <http://www.microsoft.com/office/project/default.asp>
- [MICROSTRATEGY, 2001] <http://www.microstrategy.com>
- [Ori y Santos, 96] Ori, R.; Santos, F., *"Data Extraction, Transformation, and Migration Tools"*; Basado en un documento del gobierno de E.U., 1996
- [Orr, 96] Orr, Ken; *"Data Warehousing Technology a White Paper by Ken Orr"*, The Ken Orr Institute; USA; Copyright 1996, edición revisada en el 2000.
- [Plaster, 97] Plaster, J. Donald; *"Enterprise Data Administration"*; Notas de sesión ICCM; 1997
- [Ruble, 1997] Ruble, David; *"Análisis y diseño práctico de sistemas"*; Editorial Prentice Hall; ISBN:9701701224; 1997
- [Sakaguchi y Frolick, 96] Sakaguchi, Toru; Frolick, Mark; *"A Review of the Data Warehousing Literature"*; University of Memphis; Abstract, 1996
- [TDWI, 00] The Data Warehouse Institute; *"Ten Mistakes to avoid"*; 2000
- [Tylor, 01] Tylor, Art; *"Data Warehouse and the ETL tool"*; Archivo CIBER, Inc.; <http://www.tdan.com>; 2001



LISTA DE TABLAS

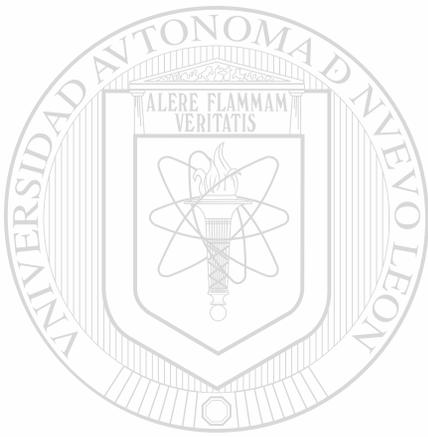
Tabla	Título	Página
3.1	Equivalencias de dimensionamiento generales	74
3.2	Resumen de dimensionamiento	75
4.1	Requerimientos	86
4.2	Matriz de bus	90
4.3	Desglose de conceptos en Matriz de bus	91
4.4	Sistemas identificados	92
4.5	Tipos de tablas del modelo	95
4.6	Dimensionamiento	95
4.7	Cuerpo de bitácora principal	102
4.8	Conceptos a almacenar para validación de la carga diaria	103
4.9	Conceptos a presentar en bitácora de carga	104
4.10	Conceptos a presentar en bitácora de procesos	105
5.1	Comparativo Análisis Proyecto Original - FIC	108
5.2	Comparativo Diseño BE Proyecto Original - FIC	109
5.3	Comparativo Diseño FE Proyecto Original - FIC	109
5.4	Comparativo Construcción BE Proyecto Original - FIC	110
5.5	Comparativo Construcción FE Proyecto Original - FIC	110
5.6	Comparativo Pruebas BE Proyecto Original - FIC	110
5.7	Comparativo Pruebas FE Proyecto Original - FIC	111
5.8	Comparativo Implementación Proyecto Original - FIC	111
5.9	Comparativo tiempos de proceso Proyecto Original - FIC	113

LISTA DE FIGURAS

Figura	Título	Página
2.1	La fábrica de información corporativa.	19
2.2	Arquitectura 1 (Mundo Externo, Aplicaciones, DW, Metadata)	27
2.3	Arquitectura 2 iniciando con Datamarts (Mundo Externo, Aplicaciones, Capa de IyT, DW, DM, Metadata)	28
2.4	Arquitectura 2 Datamarts y Data Warehouse en conjunto (Mundo Externo, Aplicaciones, Capa de IyT, DW, DM, Metadata)	28
2.5	Arquitectura 3 (Mundo Externo, Aplicaciones, Capa de IyT, DW, DM, Metadata y ODS)	29
2.6	Alineación de la FIC a las áreas competitivas del negocio	32
2.7	El ciclo de vida dimensional del negocio	36
2.8	Diagrama Estrella	40
2.9	Evolución de las configuraciones de la arquitectura	42
2.10	Arquitectura de Back End	43
2.11	Arquitectura de Front End	45
2.12	Metadata	47
4.1	Diagrama de situación actual	82
4.2	Plan de proyecto	85
4.3	FIC propuesta	89
4.4	Arquitectura propuesta	99
4.5	Infraestructura	100
4.6	Estrategia de Explotación	106
5.1	Gráfica comparativa de tiempos de procesos de desarrollo Original - FIC	112
5.2	Gráfica comparativa de tiempos de procesos de carga Original - FIC	113

LISTA DE ABREVIATURAS

Abreviatura	Significado
ADO	Almacén de datos Operativo
BD	Base de datos
CPU	Central Processing Unit
CVD	Ciclo de vida dimensional
CVDS	Ciclo de vida de desarrollo de Sistemas
DBMS	Database Management System
DM	Data Mart
DSS	Decision Support Systems o Solutions
DW	Data Warehouse
ERP	Enterprise Resource Planning
ETL	Extraction, Transformation and Loading
FAP	Filtrar, Agregar y Procesar
FIC	Fábrica de Información Corporativa
FTP	File Transfer Protocol
GB	Giga Bytes
IP	Internet Protocol
IyT	Integración y Transformación
MOLAP	Multidimensional on line analytical processing
NAS	Network Attached Storage
ODBC	Object DataBase Connectivity
ODS	Operational Data Store
OLAP	On line analytical Processing
PL	Procedural language
PPS	Pendiente por surtir
RAM	Random Access Memory
ROLAP	Relational on line Analytical Processing
SAN	Storage Area Network
SDVC	Utilizado para distinguir el CVDS pero en sentido inverso
SO	Sistema Operativo
SQL	Structured Query Language
TCP	Transfer communication protocol
TI	Tecnologías de información
TS	Table Space
UNIX	Siglas de sistema Operativo, familia de sistemas operativos



ANEXO A

FORMATOS

UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

®

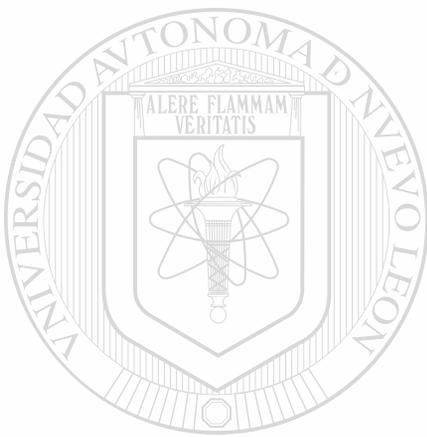
DIRECCIÓN GENERAL DE BIBLIOTECAS

ANEXO A1 PLAN DE PROYECTO

ID	Task Name	Duration	Start	Finish	% Complet	Predec	'0	T
1	FIC-Iteracion XXX	98 days	Tue 5/1/01	Thu 9/13/01	0%			
2	BACKEND	74 days	Tue 5/1/01	Fri 8/10/01	0%			
3	ANÁLISIS	15 days	Tue 5/1/01	Mon 5/21/01	0%			
4	Análisis de Estatus actual	5 days	Tue 5/1/01	Mon 5/7/01	0%			
6	Programación Juntas y Entrevistas	5 days	Tue 5/8/01	Mon 5/14/01	0%	4		
8	Identificación de las fuentes	5 days	Tue 5/15/01	Mon 5/21/01	0%	8		
10	DISEÑO	24 days	Tue 5/22/01	Fri 6/22/01	0%	3		
11	Diseño de modelo general (Inmon)	3 days	Tue 5/22/01	Thu 5/24/01	0%			
13	Metodología Kimball	13 days	Fri 5/25/01	Tue 6/12/01	0%	11		
24	Desarrollo de modelo lógico-físico	2 days	Wed 6/13/01	Thu 6/14/01	0%	13		
25	Documento: Modelo ODS-DW	2 days	Wed 6/13/01	Thu 6/14/01	0%			
26	Realización de dimensionamiento	2 days	Fri 6/15/01	Mon 6/18/01	0%	24		
28	Diseñar la arquitectura a utilizar	2 days	Tue 6/19/01	Wed 6/20/01	0%	26		
30	Diseñar mapa de configuración de procesos	2 days	Thu 6/21/01	Fri 6/22/01	0%	28		
32	CONSTRUCCIÓN	18 days	Mon 6/25/01	Mon 7/16/01	0%	10		
33	Crear usuarios y aspectos de bases de datos(TS)	2 days	Mon 6/25/01	Tue 6/26/01	0%			
34	Crear estructura de la Base de datos ODS	2 days	Wed 6/27/01	Thu 6/28/01	0%	33		
35	Crear estructura de la Base de datos DW	2 days	Fri 6/29/01	Mon 7/2/01	0%	34		
36	Construcción de Interfases	10 days	Tue 7/3/01	Mon 7/16/01	0%	35		
37	PRUEBAS	9 days	Tue 7/17/01	Fri 7/27/01	0%	32		
38	Pruebas individuales de procesos	3 days	Tue 7/17/01	Thu 7/19/01	0%			
39	Pruebas individuales de datos	3 days	Fri 7/20/01	Tue 7/24/01	0%	38		
40	Pruebas integrales	3 days	Wed 7/25/01	Fri 7/27/01	0%	39		
41	IMPLEMENTACIÓN	10 days	Mon 7/30/01	Fri 8/10/01	0%	37		
43	FRONTEND	24 days	Mon 8/13/01	Thu 9/13/01	0%	2		
44	Apoyo en análisis de Back End para entender requerimiento	1 day	Mon 8/13/01	Mon 8/13/01	0%			
45	DISEÑO	8 days	Tue 8/14/01	Thu 8/23/01	0%	44		
46	Investigar los accesos de los reportes (quien los lee, cu	5 days	Tue 8/14/01	Mon 8/20/01	0%			
47	Diseño tentativo de reportes que se requerirán	3 days	Tue 8/21/01	Thu 8/23/01	0%	46		
48	CONSTRUCCIÓN	5 days	Fri 8/24/01	Thu 8/30/01	0%	45		
49	Construcción de reportes	5 days	Fri 8/24/01	Thu 8/30/01	0%			
50	PRUEBAS	5 days	Fri 8/31/01	Thu 9/6/01	0%	48		
51	Pruebas de reportes contra información	5 days	Fri 8/31/01	Thu 9/6/01	0%			
52	IMPLEMENTACION	5 days	Fri 9/7/01	Thu 9/13/01	0%	50		

ANEXO A2

Obtención de Requerimientos



[Logo de Proyecto]

UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

®

DIRECCIÓN GENERAL DE BIBLIOTECAS

Nombre de la compañía

Iteración:

Creado por:	<input type="text"/>	Fecha de creación:	<input type="text"/>
Contribuyó:	<input type="text"/>	Fecha de revisión:	<input type="text"/>
	<input type="text"/>	Fecha finalización:	<input type="text"/>

Versión:

RESUMEN EJECUTIVO

- **Nombre del proyecto**
- **Información del área de negocio**
- **Recomendaciones**

RESUMEN DE REQUERIMIENTOS

- **Requerimientos de información**

Indicadores	Dimensión	Dimensión	Dimensión	Dimensión

- **Necesidades de información histórica**

- **Requerimientos analíticos**

Creado por:	<input type="text"/>	Fecha de creación:	<input type="text"/>
Contribuyó:	<input type="text"/>	Fecha de revisión:	<input type="text"/>
	<input type="text"/>	Fecha finalización:	<input type="text"/>

-
- *Panorama del requerimiento formal*
 - *Potencial de impacto en el negocio*
 - *Análisis típicos de información*
 - *Tipos de accesos requeridos*
-
- ***Análisis preliminar de las fuentes de información***
 - *Fuente 1(Resumen)*
 - *Fuente 2(Resumen)*



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

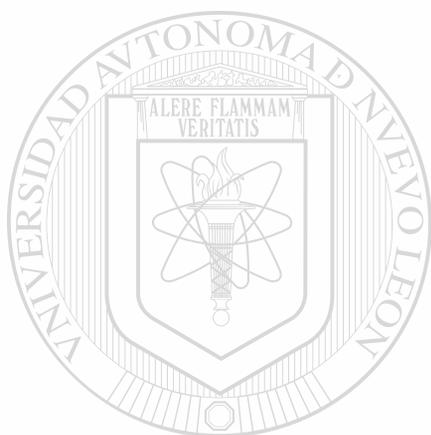
®

DIRECCIÓN GENERAL DE BIBLIOTECAS

Creado por:	<input type="text"/>	Fecha de creación:	<input type="text"/>
Contribuyó:	<input type="text"/>	Fecha de revisión:	<input type="text"/>
	<input type="text"/>	Fecha finalización:	<input type="text"/>

ANEXO A4

Documento para el modelado dimensional



[Logo del proyecto]

UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

Nombre de la compañía:

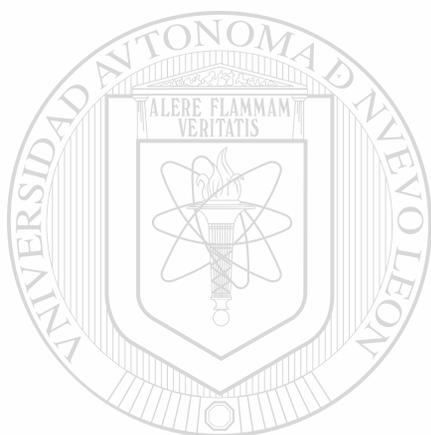
Iteración:

Version :

Creado por:	<input type="text"/>	Fecha de creación:	<input type="text"/>
Contribuyó:	<input type="text"/>	Fecha de revisión:	<input type="text"/>
	<input type="text"/>	Fecha finalización:	<input type="text"/>

Tabla de contenido

DIMENSIONES	X
DESCRIPCIONES DE DIMENSIONES	X
DETALLES DE DIMENSION	X
DETALLE DE DIMENSION XXX.....	X
DETALLE DE DIMENSION YYY.....	X
DETALLES DE HECHOS	X
DETALLES DE HECHOS ZZZ	X
DETALLES DE HECHOS AAA.....	X



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

®

DIRECCIÓN GENERAL DE BIBLIOTECAS

Creado por:	<input type="text"/>	Fecha de creación:	<input type="text"/>
Contribuyó:	<input type="text"/>	Fecha de revisión:	<input type="text"/>
	<input type="text"/>	Fecha finalización:	<input type="text"/>

Dimensiones de Data Mart

[Diagrama de las dimensiones]

Descripciones de las dimensiones

Nombre de la dimensión	Descripción de la dimensión

Detalles de dimensiones

Detalle de dimensión XXX

[Diagrama de detalle de dimensión]

Descripciones de atributos de las dimensiones:

Nombre del atributo	Descripción del atributo	Cardinalidad	Política de Cambios en dimensiones	Posibles valores

Creado por:	<input type="text"/>	Fecha de creación:	<input type="text"/>
Contribuyó:	<input type="text"/>	Fecha de revisión:	<input type="text"/>
	<input type="text"/>	Fecha finalización:	<input type="text"/>

Detalle de Hechos

Detalle de hechos ZZZ

[Diagrama de tabla de hechos ZZZ]
Incluye Granularidad, hechos y dimensiones

Descripciones de hechos base ZZZ

Nombre del hecho	Descripción	Regla de agregación

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN[®]
DIRECCIÓN GENERAL DE BIBLIOTECAS

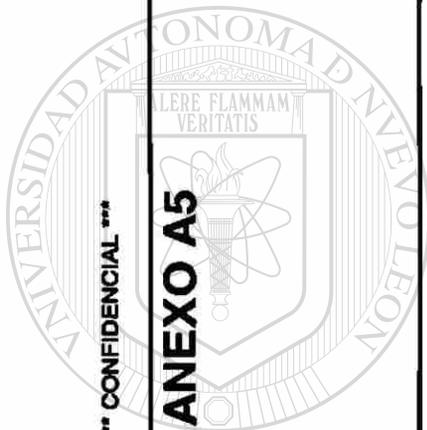
Creado por:	<input type="text"/>	Fecha de creación:	<input type="text"/>
Contribuyó:	<input type="text"/>	Fecha de revisión:	<input type="text"/>
	<input type="text"/>	Fecha finalización:	<input type="text"/>

Compañía - Iteración

Hoja de hechos derivados

*** CONFIDENCIAL ***

ANEXO A5



HOJA DE HECHOS DERIVADOS

Bandera de cambio	Datamart(s)	Nombre	Descripción	Tipo	Regla de Agregación	Fórmula	Restricciones	Transformaciones

Creado por:

Contribuyó:

Fecha de creación:

Fecha de revisión:

Fecha finalización:

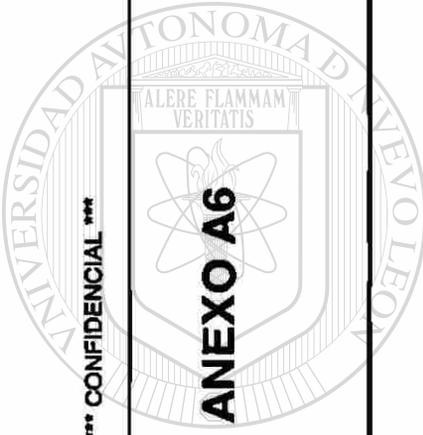
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
DIRECCIÓN GENERAL DE BIBLIOTECAS

UANL

Compañía - Iteración

Definición de fuentes

*** CONFIDENCIAL ***



DEFINICIÓN DE FUENTES

Fuente	Área de negocio responsable	Reponsable de informática	Plataforma	Localización	Descripción

CRITERIOS DE SELECCIÓN

- *Accesabilidad de los datos*
- *Exactitud de los datos*
- *Calendarización de proyectos*

Creado por:

Contribuyó:

Fecha de creación:

Fecha de revisión:

Fecha finalización:

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
DIRECCIÓN GENERAL DE BIBLIOTECAS

UNANL

**ANEXO A8
DIMENSIONAMIENTO**

GENERAL

TABLAS

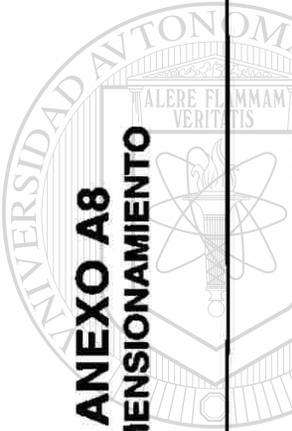
Tabla	Campo	Tipo Dato	Tamaño	Escala	Valor	Bytes	Total bytes registro	Num regs al periodo	Periodos en línea	Total registros históricos	Total espacio datos
TABLA	CAMPO1	DATE	7	0	NOT NULL	7					
TABLA	CAMPO2	NUMBER	10	0	NOT NULL	6					
TABLA	CAMPO3	VARCHAR2	20		NOT NULL	20					
Total											

INDICES

Indice	Tabla	Campo	Tipo Dato	Tamaño	Escala	Valor	Bytes	Total bytes registro	Num regs al periodo	Periodos en línea	Total registros históricos	Total espacio datos
INDICE	TABLA	CAMPO1	DATE	7		0	NOT NULL	7				
INDICE	TABLA	CAMPO2	NUMBER	10		0	NOT NULL	6				
INDICE	TABLA	CAMPO3	VARCHAR2	20		NOT NULL	20					
Total												

RESUMEN

Indice	Tabla	Cantidad de registros	Cant. Inicial	Next Extent	Tablespace
INDICE	CAMPO	1000	100000	5000	XXXX
Total					



UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
DIRECCIÓN GENERAL DE BIBLIOTECAS



**ANEXO A9
MAPA DE CONFIGURACIÓN**

GENERAL

Nivel 1	Nivel 2	Nivel 3	Nivel N	Objetivo	Origen	Destino
b_principal(Sec)	b_batch1(Con)	s_m1		Objetivo del proceso		
		s_m2				
	b_batch2	s_m3				
		s_m4				
	b_batch3	s_m5				
		s_m6				
	b_batch4	s_m7				
		s_m8				
		s_m9				

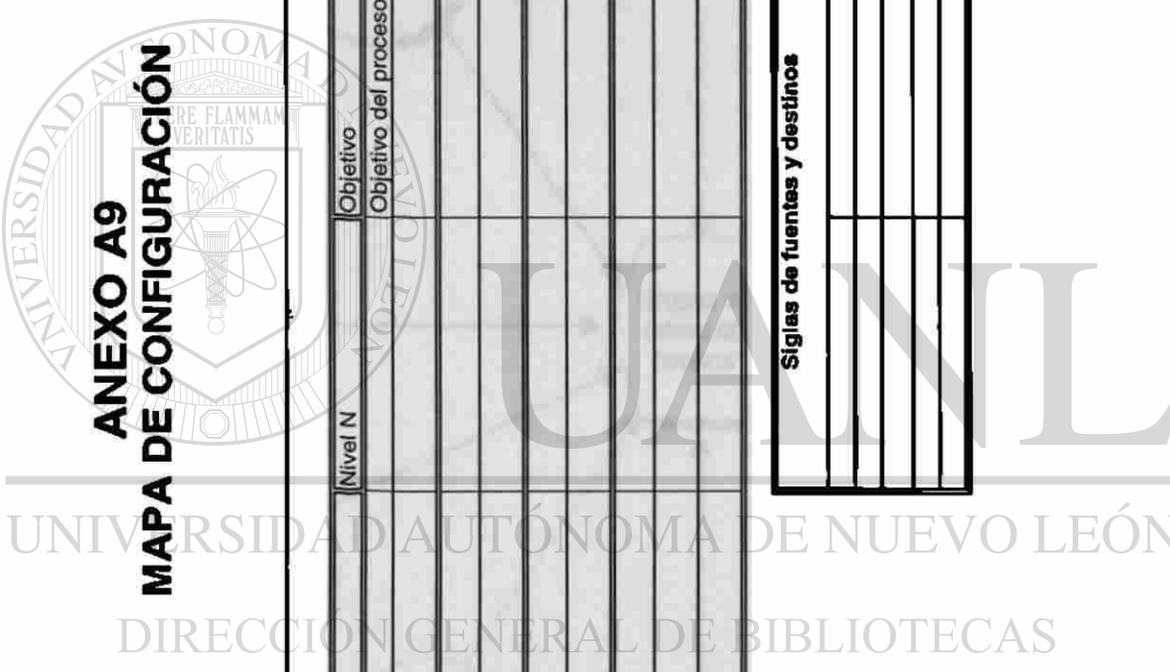
Siglas de fuentes y destinos

Siglas de proceso:

Con	Concurrente
S	Secuencial

Estándares:

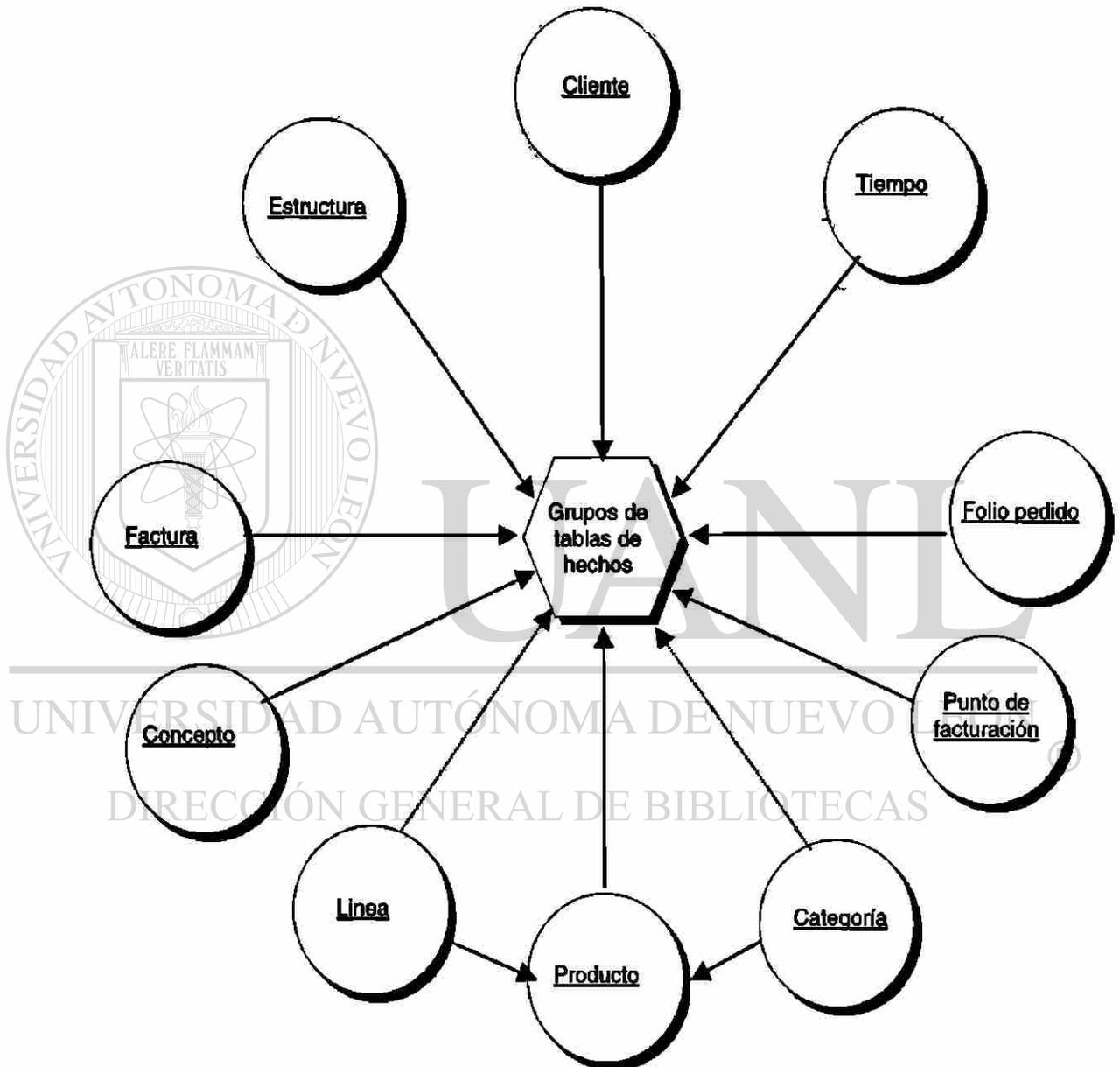
b_	batches
s_	sesiones



ANEXO B

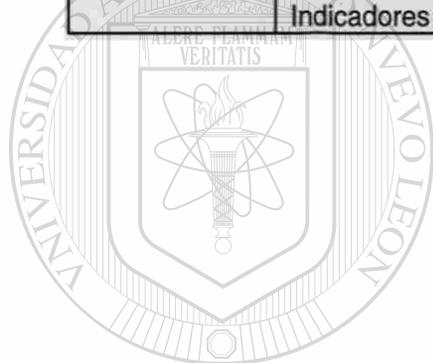
MODELADO DIMENSIONAL APLICADO

Dimensiones de Data Mart



Descripción de las dimensiones:

Nombre de la dimensión	Descripción de la dimensión
Estructura	Se refiere a la distribución Geográfica llevando implícitamente la distribución organizacional. Contiene todos los atributos relacionados con una jerarquía de Región, División, Distrito
Punto de Facturación	Se refiere al lugar en donde se realiza la facturación. Comúnmente denominado centro de distribución
Cliente	Contiene todos los atributos del cliente a quien se le realiza el pedido y la venta. Distingue los atributos de Canal, Grupo y Cadena.
Producto	Contiene todos los atributos de los productos de la compañía. Su jerarquía está representada por los conceptos línea, segmento, categoría, marca.
Línea	Se refiere a la línea de producto. (Galleta, Pastas, etcétera)
Categoría	Se refiere a la categoría de producto. (Chocolate, Rellenos, etcétera)
Tiempo	Contiene los atributos de tiempo cuando una actividad a ocurrido
Factura	Es el número de factura
Folio Pedido	Es el folio del pedido
Concepto	Dimensión Interna para modelar correctamente los diferentes tipos de Indicadores de negocio necesarios.



UANL

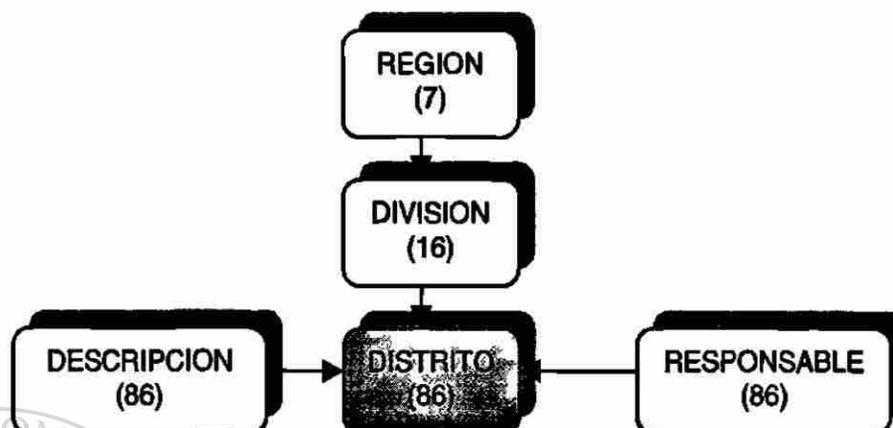
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

®

DIRECCIÓN GENERAL DE BIBLIOTECAS

Detalles de dimensiones

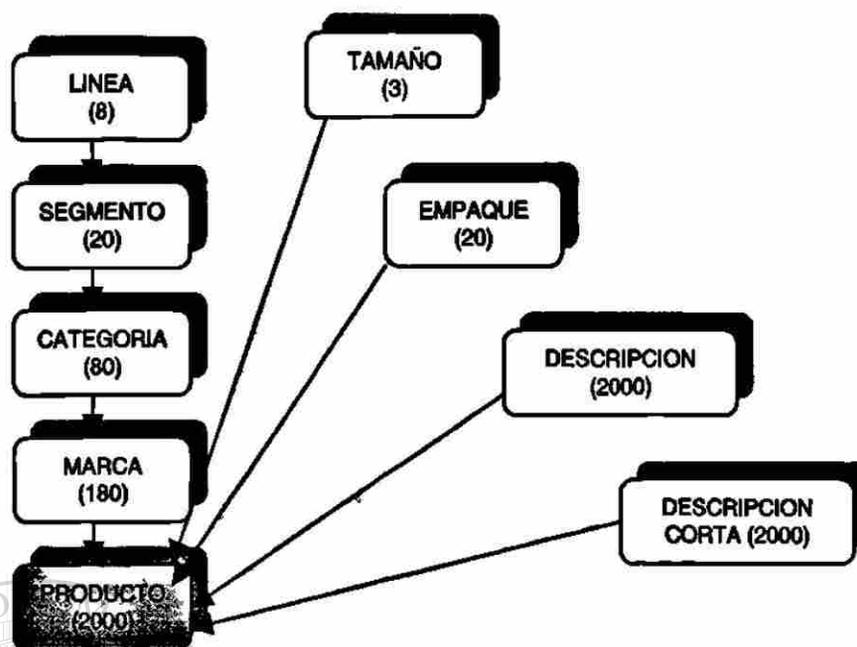
Detalle de dimensión: ESTRUCTURA



Descripciones de atributos de las dimensiones:

Nombre del atributo	Descripción del atributo	Cardinalidad	Política de Cambios en dimensiones	Posibles valores
Distrito	Zona de ventas perteneciente a una región y una división	86	No se actualiza	115
División	Es una división de la región.	16	Se sobrescribe	Monterrey
Región	Es la región en donde se realiza el movimiento o actividad tanto en demanda como pedido.	7	Se sobrescribe	Norte
Descripción	Descripción del distrito	86	Se sobrescribe	Monterrey - Guadalupe
Responsable	Es la persona responsable de su respectivo distrito	86	Se sobrescribe	Juan Perez

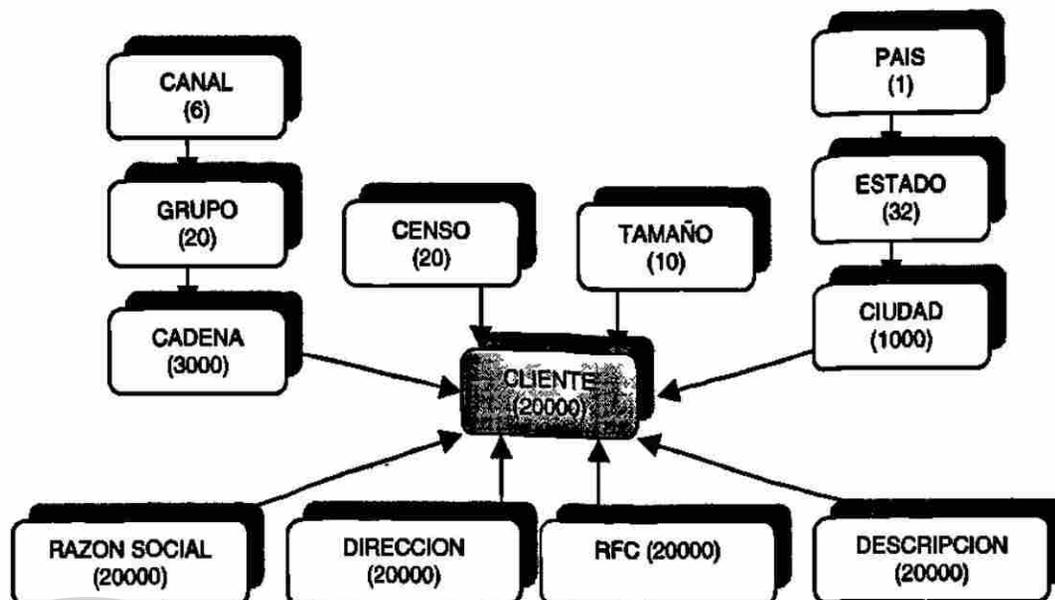
Detalle de dimensión: PRODUCTO



Descripciones de atributos de las dimensiones:

Nombre del atributo	Descripción del atributo	Cardinalidad	Política de Cambios en dimensiones	Posibles valores
Línea	Corresponde a la línea de producto.	8	Se sobrescribe	Galleta
Segmento	Corresponde al segmento de ventas donde aplica el producto	20	Se sobrescribe	Low End
Categoría	Corresponde a la categoría del producto	80	Se sobrescribe	Low End
Marca	Corresponde a la marca del producto	180	Se sobrescribe	Animalitos
Producto	Producto de gamesa	2000	No se actualiza	420200
Tamaño	Es el tamaño del producto	3	Se sobrescribe	Boxes
Empaque	Es el empaque del producto	20	Se sobrescribe	Super size
Descripción larga	Descripción del producto	2000	Se sobrescribe	Animalitos 20/200 grs.
Descripción corta	Descripción de 20 caracteres del producto	2000	Se sobrescribe	Animalitos 20/200 grs.

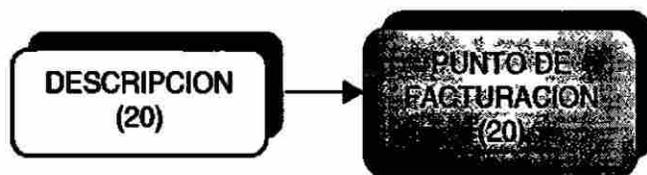
Detalle de dimensión: CLIENTE



Descripciones de atributos de las dimensiones:

Nombre del atributo	Descripción del atributo	Cardinalidad	Política de Cambios en dimensiones	Posibles valores
Cliente	Corresponde a la línea de producto.	20000	No se actualiza	52805
Canal	Corresponde al segmento de ventas donde aplica el producto	6	Se sobrescribe	Mayoreo
Grupo	Corresponde a la categoría del producto	20	Se sobrescribe	Sin grupo
Cadena	Corresponde a la marca del producto	3000	Se sobrescribe	Sin cadena
Pais	Producto de gamesa	1	No se actualiza	México
Estado	Es el tamaño del producto	32	Se sobrescribe	Nuevo León
Ciudad	Es el empaque del producto	1000	Se sobrescribe	Villaldama
Razon Social	Descripción del producto	20000	Se sobrescribe	Juan Pérez
Dirección		20000		Tepezcales #304 Col. Los Rodríguez
RFC		20000		FOPJ711201
Descripción	Descripción de 20 caracteres del producto	20000	Se sobrescribe	La tiendita

Detalle de dimensión: PUNTO DE FACTURACIÓN



Descripciones de atributos de las dimensiones.

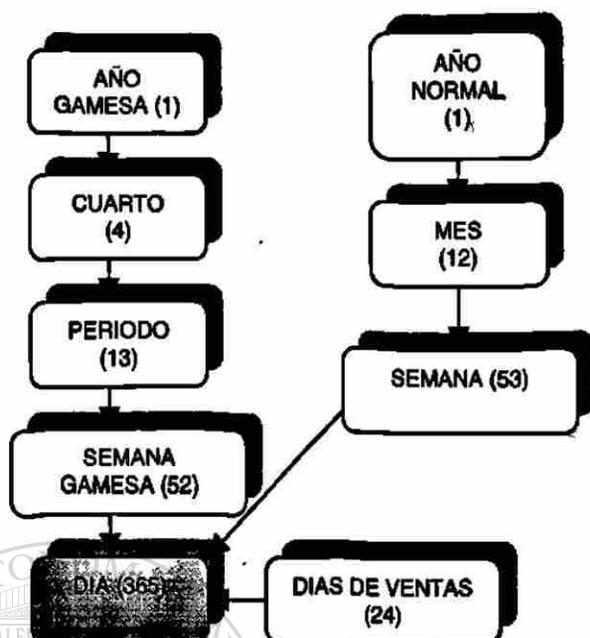
Nombre del atributo	Descripción del atributo	Cardinalidad	Política de Cambios en dimensiones	Posibles valores
Punto de Facturación	Lugar donde se realiza la facturación	20	No se actualiza	810
Descripción	Descripción del punto de facturación.	20	Se sobrescribe	CRSC Monterrey

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

DIRECCIÓN GENERAL DE BIBLIOTECAS

®

Detalle de dimensión: TIEMPO



Descripciones de atributos de las dimensiones:

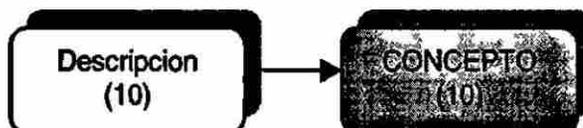
Nombre del atributo	Descripción del atributo	Cardinalidad	Política de Cambios en dimensiones	Posibles valores
Año Gamesa	Año correspondiente al ciclo de los periodos de Gamesa	1	No se actualiza	2001
Cuarto	División del año gamesa en 4	4	No se actualiza	200101,200102,200103,200104
Periodo	Corresponde a los periodos de cierre de la compañía y suelen ser 13.	13	No se actualiza	200101...200113
Semana Gamesa	Corresponde a la semana basado en el periodo gamesa.	52	No se actualiza	20010101
Año Normal	Año en curso del calendario normal	1	No se actualiza	2001
Mes	Mes del calendario normal	12	No se actualiza	1-12
Semana	Semana del mes normal	53	No se actualiza	1-53
Día	Día en que se realiza la actividad específica	365	No se actualiza	05-may-2001
Día de ventas	Agrupador de los días para observar la venta	28	No se actualiza	1-24

Detalle de dimensión: FOLIO DE PEDIDO (DEGENERADA)**Descripciones de atributos de las dimensiones:**

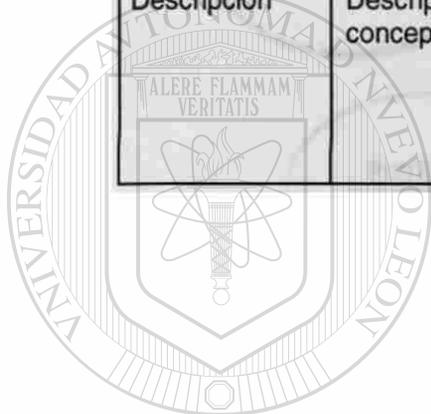
Nombre del atributo	Descripción del atributo	Cardinalidad	Política de Cambios en dimensiones	Posibles valores
Folio Pedido	Folio del pedido en cualquier estatus	10000000	No se actualiza	1000000

Detalle de dimensión: FACTURA (DEGENERADA)**Descripciones de atributos de las dimensiones:**

Nombre del atributo	Descripción del atributo	Cardinalidad	Política de Cambios en dimensiones	Posibles valores
Número Factura	Número de la factura	1000000	No se actualiza	1000000

Detalle de dimensión: CONCEPTO (Dimensión interna)**Descripciones de atributos de las dimensiones:**

Nombre del atributo	Descripción del atributo	Cardinalidad	Política de Cambios en dimensiones	Posibles valores
Concepto	Concepto interno del modelo para manejar los diferentes tipos de pedidos y para distinguir la facturación de la devolución	10	No se actualiza	1,2,3,4,5,6,7,8,9
Descripción	Descripción de los conceptos	10	Se sobrescribe	Facturación, Devolución, Demanda captada, Demanda Autorizada.



UANL

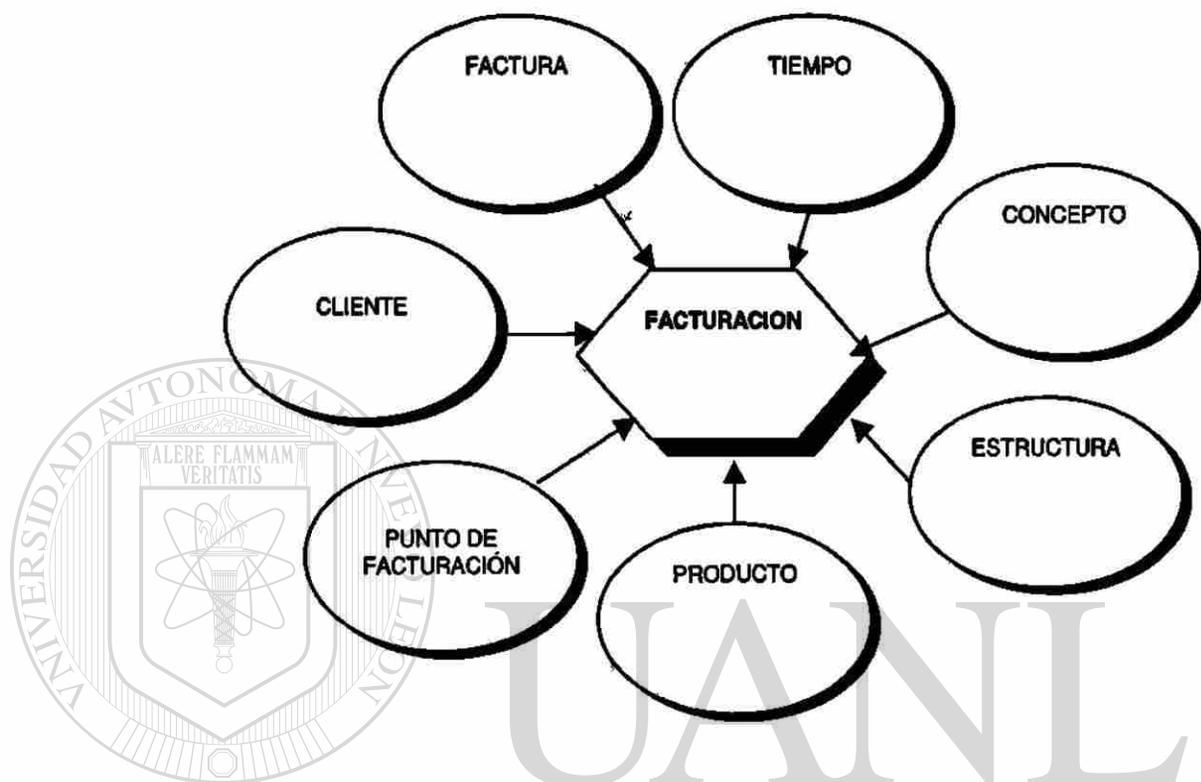
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

®

DIRECCIÓN GENERAL DE BIBLIOTECAS

Detalle de Hechos

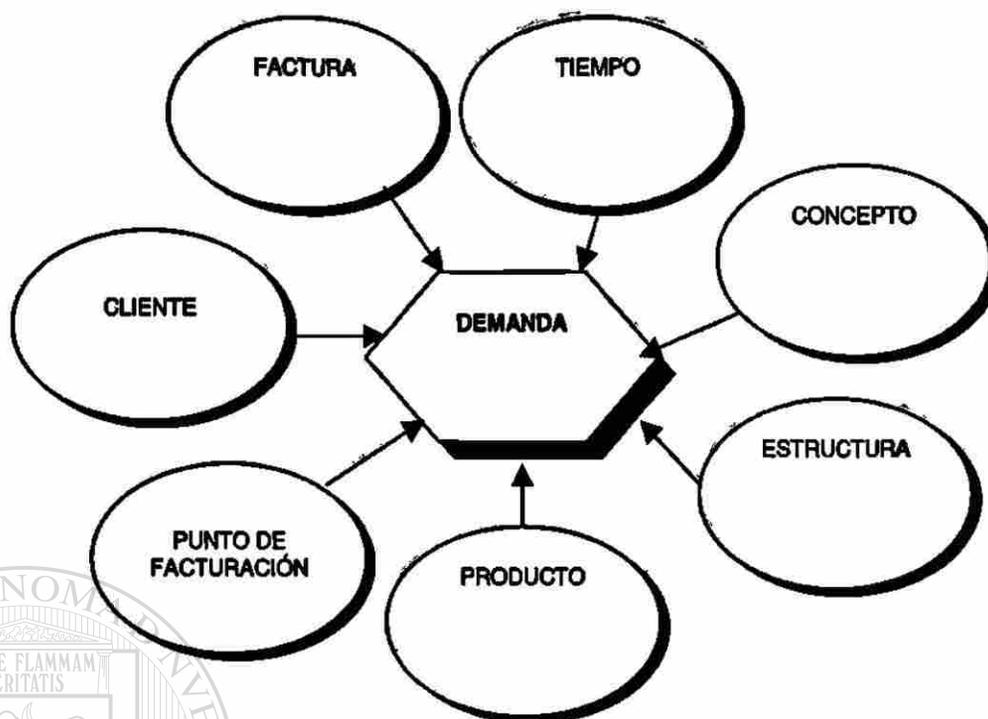
Detalle de tabla de hechos: FACTURACIÓN



Descripciones de hechos: FACTURACIÓN

Nombre del hecho	Descripción	Regla de agregación
Unidades Brutas	Unidades Vendidas + Producto Gratis	Sum
Unidades Netas	Unidades Vendidas	Sum
Kilogramos Brutos	Kgs. Vendidos + Producto Gratis	Sum
Kilogramos Netos	Kgs. Vendidos	Sum
Pesos Brutos	Venta en Pesos + Producto Gratis + descuentos	Sum
Pesos Brutos Intermedios	Venta en Pesos + descuentos	Sum
Pesos netos	Venta en Pesos	Sum
Pesos Especial	Descuentos especiales otorgados	Sum
Pesos Linea	Bonificaciones otorgadas	Sum
Pesos Promocional	Descuentos promocionales otorgados	Sum
Motivo de Rechazo	Si es una devolución, se almacena el motivo.	N/A

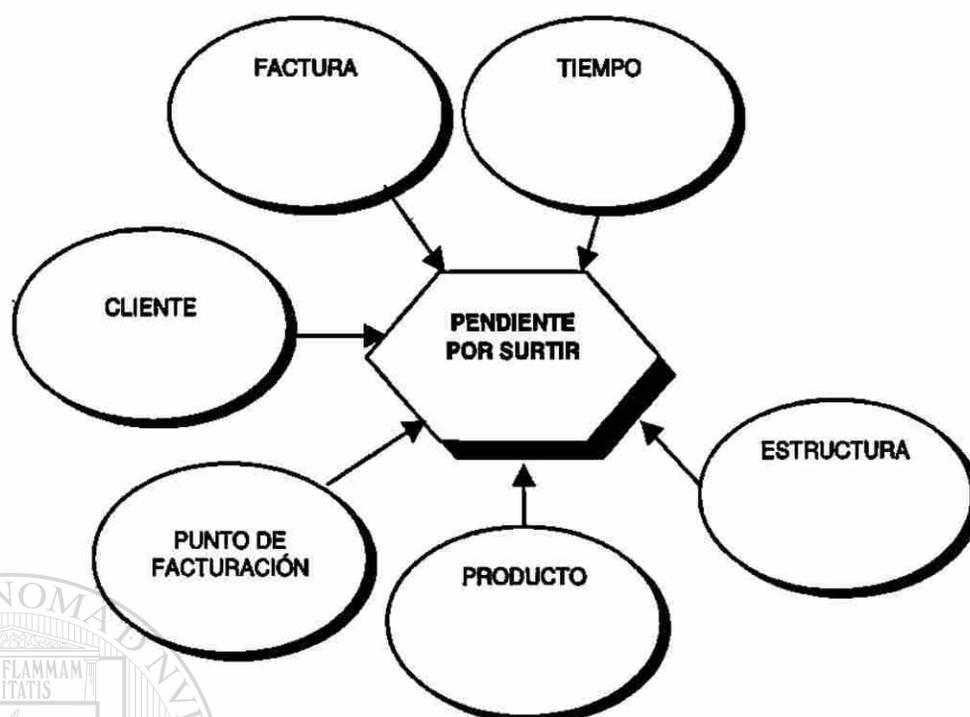
Detalle de tabla de hechos: DEMANDA



Descripciones de hechos: DEMANDA (Incluyen los diferentes tipos de demanda)

Nombre del hecho	Descripción	Regla de agregación
Unidades Brutas	Unidades de demanda + Producto Gratis	Sum
Unidades Netas	Unidades de demanda + Producto Gratis	Sum
Kilogramos Brutos	Kgs. de demanda + Producto Gratis	Sum
Kilogramos Netos	Kgs. de demanda	Sum
Pesos Brutos	Demanda en Pesos + Producto Gratis +descuentos	Sum
Pesos Brutos Intermedios	Demanda en Pesos + descuentos	Sum
Pesos netos	Demanda en Pesos	Sum
Pesos Especial	Descuentos especiales otorgados	Sum
Pesos Linea	Bonificaciones otorgadas	Sum
Pesos Promocional	Descuentos promocionales otorgados	Sum
Motivo de Rechazo	Si es una cancelación, se almacena el motivo.	N/A

Detalle de tabla de hechos: PENDIENTE POR SURTIR

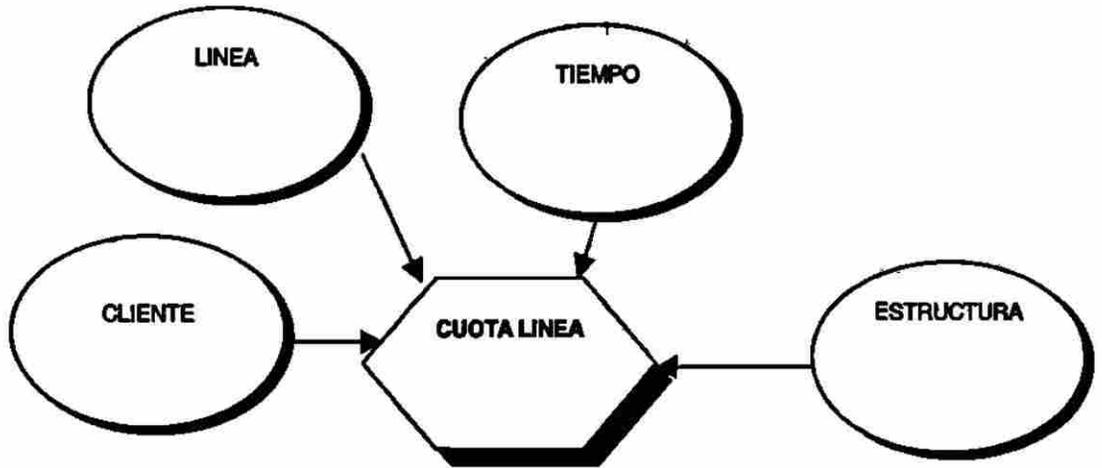


Descripciones de hechos: Demanda Pendiente Por surtir.

Nombre del hecho	Descripción	Regla de agregación
Unidades Brutas	Unidades PPS* + Producto Gratis	Sum
Unidades Netas	Unidades PPS* + Producto Gratis	Sum
Kilogramos Brutos	Kgs. PPS* + Producto Gratis	Sum
Kilogramos Netos	Kgs. PPS*	Sum (R)
Pesos Brutos	Pesos de PPS* + Producto Gratis + descuentos	Sum
Pesos Brutos Intermedios	Pesos de PPS* + descuentos	Sum
Pesos netos	Pesos de PPS*	Sum
Pesos Especial	Descuentos especiales otorgados	Sum
Pesos Linea	Bonificaciones otorgadas	Sum
Pesos Promocional	Descuentos promocionales otorgados	Sum
Fecha de Programación	Fecha en que se debe de entregar el pedido	N/A

***PPS: Pendiente por surtir**

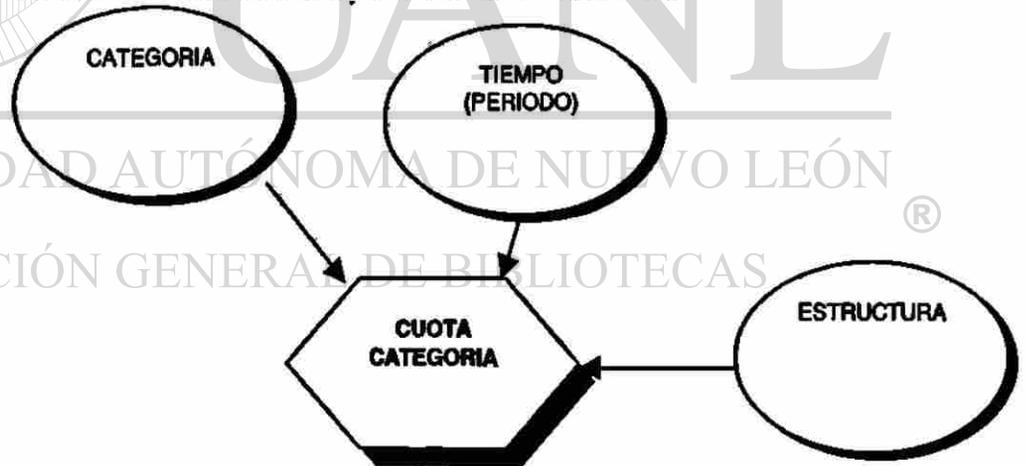
Detalle de tabla de hechos: CUOTA POR LINEA



Descripciones de hechos: Cuota por Línea.

Nombre del hecho	Descripción	Regla de agregación
Kilogramos	Kilogramos objetivo a facturar o vender.	Sum

Detalle de tabla de hechos: CUOTA POR CATEGORIA



Descripciones de hechos: Cuota por Categoría.

Nombre del hecho	Descripción	Regla de agregación
Kilogramos	Kilogramos objetivo a facturar o vender.	Sum

**ANEXO C
HOJA DE HECHOS DERIVADOS**

Bandera de cambio	Fecha	Nombre	Descripción	Tipo	Regla de Agregación	Fórmula	Restricción	Transformaciones
Ventas		Unidades Brutas	El número de unidades vendidas al consumidor incluyendo el producto gratis Sin considerar devoluciones	Calc	Sum	Sum(Unidades brutas) <concepto=1> - Sum(Unidades brutas) <concepto=2>	-	
Ventas		Unidades Netas	El número de unidades vendidas al consumidor sin incluir el producto gratis Sin considerar devoluciones	Calc	Sum	Sum(Unidades netas) <concepto=1> - Sum(Unidades netas) <concepto=2>	-	
Ventas		Unidades de PG	Son las unidades regaladas Sin considerar devoluciones	Calc	Sum	Sum(unidades brutas)-sum(unidades netas)	-	
Ventas		Kilogramos Brutos	El número de kilogramos vendidos al consumidor incluyendo el producto gratis Sin considerar devoluciones	Calc	Sum	Sum(Kilogramos brutos) <concepto=1> - Sum(kilogramos brutos) <concepto=2>	-	
Ventas		Kilogramos Netos	El número de unidades vendidas al consumidor sin incluir el producto gratis Sin considerar devoluciones	Calc	Sum	Sum(Kilogramos netos) <concepto=1> - Sum(Kilogramos netos) <concepto=2>	-	
Ventas		Kilos de PG	Son los kilogramos de producto gratis Sin considerar devoluciones	Calc	Sum	Sum(kilogramos brutos)-sum(kilogramos netos)	-	
Ventas		Pesos Brutos	Venta en pesos considerando producto gratis y descuentos	Calc	Sum	Sum(pesos brutos) <concepto=1> - Sum(pesos brutos) <concepto=2>	-	
Ventas		Pesos de PG	Son los pesos del producto gratis	Calc	Sum	Sum(pesos brutos) -sum (pesos brutos intermedios)	-	
Ventas		Pesos netos	Venta en pesos sin considerar descuentos y el producto gratis.	Calc	Sum	Sum(pesos netos) <concepto=1> - Sum(pesos netos) <concepto=2>	-	
Ventas		Pesos Especial	Descuentos especiales	Calc	Sum	Sum(pesos especial) <concepto=1> - sum(pesos especial) <concepto=2>	-	
Ventas		Pesos Línea	Descuentos en línea	Calc	Sum	Sum(pesos línea) <concepto=1> - sum(pesos línea) <concepto=2>	-	
Ventas		Pesos Promocional	Descuentos promocionales	Calc	Sum	Sum(pesos promocional) <concepto=1> - sum(pesos promocional) <concepto=2>	-	

Creado por:

Rodolfo Portillo Salinas

Contribuyó:

-

Fecha de creación:

15-Nov-2001

Fecha de revisión:

Fecha finalización:

20-Nov-2001

Compañía - Iteración Ventas

Hoja de hechos derivados

Bandera de cambio	Datamart(s)	Nombre	Descripción	Tipo	Regla de Agregación	Fórmula	Restricciones	Transformaciones
	Demanda y Demanda por surtir	Unidades Brutas	El número de unidades incluyendo el producto gratis	Calc	Sum	Sum(Unidades brutas)	<Concepto=3, Captada>	-
	Demanda y Demanda por surtir	Unidades Netas	El número de unidades sin incluir el producto gratis	Calc	Sum	Sum(Unidades netas)	<Concepto=4, Depurada>	-
	Demanda y Demanda por surtir	Unidades de PG	Son las unidades regaladas	Calc	Sum	Sum(unidades brutas)-sum(unidades netas)	<Concepto=5, Incumplida>	-
	Demanda y Demanda por surtir	Kilogramos Brutos	El número de kilogramos incluyendo el producto gratis	Calc	Sum	Sum(Kilogramos brutos)	<Concepto=6, Rechazada por credito>	-
	Demanda y Demanda por surtir	Kilogramos Netos	El número de unidades sin incluir el producto gratis	Calc	Sum	Sum(Kilogramos netos)	<Concepto=7, otras cancelaciones>	-
	Demanda y Demanda por surtir	Kilos de PG	Son los kilogramos de producto gratis	Calc	Sum	Sum(kilogramos brutos)-sum(kilogramos netos)		-
	Demanda y Demanda por surtir	Pesos Brutos	Pesos de la demanda considerando producto gratis y descuentos	Calc	Sum	Sum(pesos brutos)		-
	Demanda y Demanda por surtir	Pesos de PG	Son los pesos del producto gratis	Calc	Sum	Sum(pesos brutos) -sum (pesos brutos intermedios)	Sin concepto y del datamart de pendiente por surtir no aplican restricciones.	-
	Demanda y Demanda por surtir	Pesos netos	Pesos de la demanda sin considerar descuentos y el producto gratis.	Calc	Sum	Sum(pesos netos)		-
	Demanda y Demanda por surtir	Pesos Especial	Descuentos especiales	Calc	Sum	Sum(pesos especial)		-
	Demanda y Demanda por surtir	Pesos Linea	Descuentos en línea	Calc	Sum	Sum(pesos línea)		-
	Demanda y Demanda por surtir	Pesos Promocional	Descuentos promocionales	Calc	Sum	Sum(pesos promocional)		-
Bandera de cambio	Datamart(s)	Nombre	Descripción	Tipo	Regla de Agregación	Fórmula	Restricciones	Transformaciones
	Objetivos	Kilogramos	Son los kilogramos objetivos propuestos a vender.	Calc	Sum	Sum(kilogramos)		
	Objetivos Venta	% de cobertura	Es el porcentaje cumplido del objetivo contra la venta	Calc	ReCalc	(Sum(Pesos brutos)/sum(kilogramos))*100		

Siglas:
 Calc= Calculada
 Sum= Sumatoria
 ReCalc= Un cálculo sobre otro cálculo
 -condición, resultados = es una condición aplicada

Creado por:
 Rodolfo Portillo Salinas

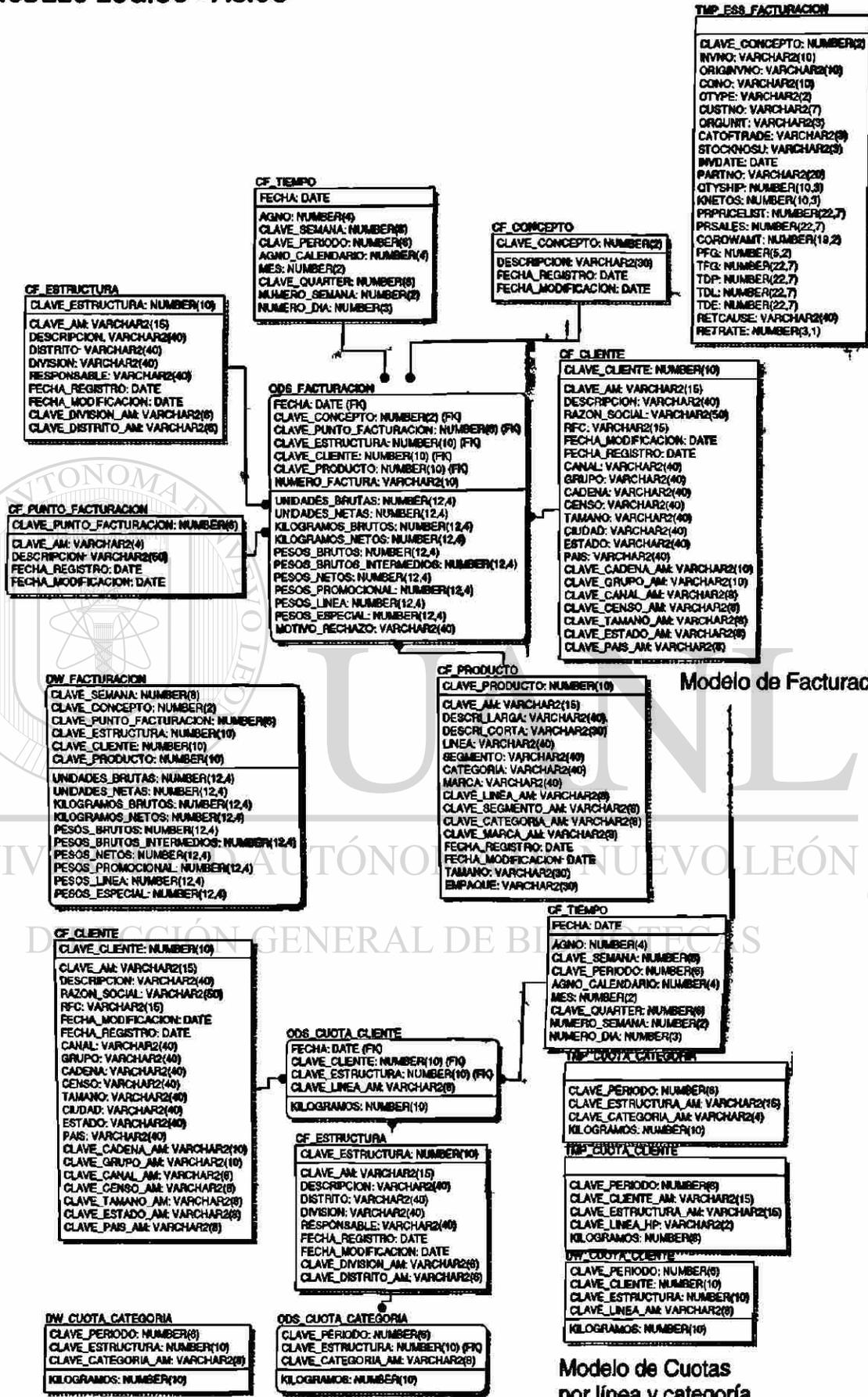
Contribuyó:
 -

Fecha de creación:
 15-Nov-2001

Fecha de revisión:
 20-Nov-2001

Fecha finalización:
 20-Nov-2001

ANEXO D MODELO LÓGICO - FÍSICO



ANEXO F GLOSARIO

copo

Término	Significado
Algoritmos Genéticos	Este concepto es un método de optimización combinatorial basada en el proceso de la evolución biológica. Se utiliza generalmente en minería de datos para optimización en la creación de agrupaciones.
Almacén de datos operativo	Es una estructura híbrida que tiene los elementos igualmente fuertes para tener procesamiento operacional y procesamiento de soporte a la toma de decisiones
Área de procesamiento	Es un componente de la arquitectura de la fábrica de información encargada del procesamiento de los datos.
Atributo	Se utiliza para describir una característica tangible de alguna cosa. En conjunto forman una dimensión.
Back End	Término en inglés utilizado para identificar todos los procesos y servicios relacionados con el almacenamiento de datos.
Base de datos	Es una colección organizada de datos.
Batch	Objeto utilizado en Powermart que agrupa una serie de sesiones para permitir mayor flexibilidad a la hora de ejecutar los procesos permitiendo ejecuciones secuenciales y concurrentes.
Bytes	Es un concepto de almacenamiento que está compuesto por 8 bits (unidades menores de almacenamientos).
Capa de integración y transformación	Es una arquitectura que alimenta al ODS o al DW, esta capa, a diferencia de estos componentes mencionados los cuales están hechos principalmente por datos, está constituida principalmente de programas.
Cardinalidad	Se refiere a la cantidad de apariencias de valores diferentes en un atributo
Data Mart	Un Data Mart es una colección de datos hechos a la medida de las necesidades de procesamiento para soporte a la toma de decisiones que ha sido configurada para satisfacer los requerimientos de "un departamento en específico".
Data Warehouse	Es una colección de información corporativa derivada directamente de los sistemas operacionales y de algunos orígenes de datos externos en donde su propósito específico es soportar la toma de decisiones en un negocio, no las operaciones del negocio
Designer	Es una de las 3 aplicaciones de Powermart en la cual se realizan todos los desarrollos de interfases sin considerar cuestiones de conectividades.

Diagrama copo de nieve	Es otra forma de realizar un modelado dimensional en la cual los atributos de las dimensiones pertenecen a otra dimensión por lo que el diagrama tiende a parecer un copo de nieve. Llamado en inglés snowflake.
Diagrama Estrella	Es una estructura de modelación compuesta por elementos denominados hechos y dimensiones los que se unen para formar una arquitectura que asemejan a una estrella.
Dimensión	Es un elemento del diagrama estrella que contiene el conjunto de atributos que afectan a un hecho.
Drills	O navegación. Se refiere a la forma en que se va accedendo la información a través del cubo dimensional, es decir, es la navegación a través de las dimensiones.
Ecosistema de información	Es un sistema con componentes diferentes, cada uno sirviendo a una comunidad directamente mientras se trabaja en consenso con los otros componentes para producir un ambiente cohesivo y balanceado de información.
ERP	Es un sistema de gestión de la información estructurado para satisfacer la demanda de soluciones de gestión empresarial, basado en el ofrecimiento de una solución completa que permite a las empresas evaluar, implementar y gestionar más fácilmente su negocio.
Excel	Es una hora electrónica de cálculo diseñada para realizar, primariamente, modelos aritméticos que ordinariamente se implementan en hojas tabulares (como estados de resultados, balances, reportes de calificaciones escolares).
Fábrica de información Corporativa	Una fábrica de información corporativa es la arquitectura física de lo que representa el ecosistema de información. Es aquella que proporciona la materia prima (los datos) para que sea transformada en información.
Front end	Término en inglés utilizado para identificar todos los procesos y servicios relacionados con la presentación de información a los usuarios.
Granularidad	La granularidad se refiere al nivel de detalle por unidad de dato en el DW. A mayor detalle menor granularidad, a menor detalle más granularidad.
Hecho	Es un elemento del diagrama estrella que contiene algo mensurable y que corresponde a un indicador del negocio. En pocas palabras es todo lo que puede medirse.
Impromptu	Es una herramienta ROLAP de la compañía de Cognos que permite la realización de análisis de información.
Internet	Se define como un conjunto de comunidades y como un conjunto de tecnologías, y su éxito se puede atribuir a la satisfacción de las necesidades básicas de la comunidad y a la utilización de ésta de un modo efectivo para impulsar la infraestructura.

Iteración	Es una forma conceptual de llamarle a cada desarrollo que va conformando la fábrica de información.
Lookup	O Búsqueda. Se refiere al objeto que permite realizar búsquedas sobre una llave o llaves de tablas para encontrar un atributo relacionado con ésta.
Mapeo	Desde el punto de vista Powermart es un objeto que está compuesto por todas las transformaciones necesarias en un proceso de carga. Desde el punto de vista de desarrollo es la asignación de un valor "A" a su equivalente valor "B".
Matriz de bus	Es una estrategia que permite visualizar las relaciones entre las dimensiones de los diferentes conceptos de negocio.
Metadata	Concepto que expresa los datos acerca de los datos. Es como un diccionario de datos pero va más allá incluyendo cuestiones de documentación a nivel proceso, desarrollo y conceptos de negocio.
Minería de datos	Es el proceso de encontrar tendencias y patrones en los datos. El objetivo de este proceso es el de ordenar a través de largas cantidades de datos y descubrir información nueva.
Modelado dimensional	Es una estrategia consistente en el diseño de los modelos de datos necesarios para soportar los análisis basados en los requerimientos analíticos de los usuarios con un enfoque de hechos y dimensiones.
MOLAP	Es una estrategia de procesamiento en línea en la cual se utiliza una arquitectura de cubos en los cuales se manejan dimensiones y comandos específicos para esta solución.
Normalización	Técnica empleada para resolver las inconsistencias en la actualización de cada Base de Datos, evitándose las redundancias, incoherencias, ambigüedades y pérdida de información
Paradigma	Es una percepción con la que vemos el mundo
Powermart	Es una herramienta especializada para las actividades de Extracción, Transformación y Carga que cumple con los servicios necesarios para satisfacer el área de Back End.
Redes neuronales	T. Kohonen en "An Introduction to Neural Computing". Neural Networks, Vol. 1, págs. 3-16, 1988 define como "Un conjunto de redes interconectadas masivamente en paralelo de elementos simples (usualmente adaptativos) y con organización jerárquica, las cuales intentan interactuar con los objetos del mundo real de mismo modo que lo hace el sistema nervioso biológico". Esta estrategia permite ser utilizada para obtener patrones sobre los datos en herramientas aplicadas de minería de datos.

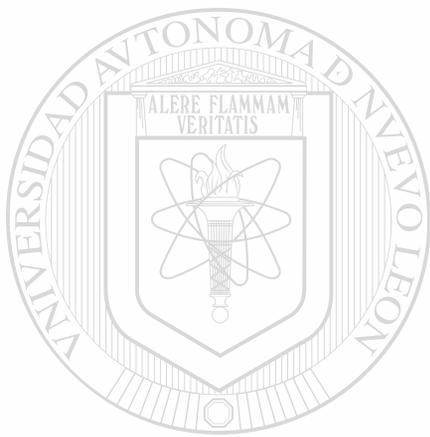
Referencias de integridad	En conceptos de bases de datos se refiere a la concordancia que debe guardarse entre una llave foránea de una tabla y una llave primaria de otra para poder formar una relación.
Repository Manager	Es una de las 3 aplicaciones de Powermart en la cual se administra y explota la metadata de los desarrollos y configuraciones.
ROLAP	Es una estrategia de procesamiento en línea en la cual se utiliza una arquitectura relacional para la simulación del ambiente dimensional.
Server Manager	Es una de las 3 aplicaciones de Powermart en la cual se configuran las conectividades y se realiza la calendarización y administración de los desarrollos realizados previamente en el Designer
Sesión	Es la forma en que se conoce en Powermart a la configuración de un mapeo.
Snapshot	Congelación de un instante. En el ambiente DSS se considera como una foto de la información al instante en que se extrae.
Sumarización	Es el procesamiento para llevar la información a un nivel de detalle menor de tal manera que la explotación sea óptima. Es decir es un acumulado de datos.
Transformación	Es la actividad que se realiza para obtener un resultado diferente sobre el original. Una transformación puede ser una operación aritmética, una comparación para dar un valor A o B, un filtrado, una sumarización.
Triangulación	Actividad que consiste en tomar los datos obtenidos por algún medio, reprocesarlos, enviarlos a otro proceso y que este vuelva a realizar alguna actividad con éstos.
Web	Es el término en inglés para representar la red mundial utilizada generalmente para la distribución de voz, video e imagen.

RESUMEN AUTOBIOGRÁFICO

Ing. Rodolfo Portillo Salinas

El grado que se desea obtener es Maestro en ciencias de la Administración con Especialidad en Sistemas, la tesis lleva por título "MEJORAMIENTO EN LOS PROCESOS DE DESARROLLO, CARGA, VALIDACIÓN Y EXPLOTACIÓN IMPLICADOS EN UN DATA WAREHOUSE MEDIANTE LA CREACIÓN DE UNA FÁBRICA DE INFORMACIÓN CORPORATIVA APLICADA AL ÁREA DE VENTAS DE LA EMPRESA GAMESA". Nací en la ciudad de Monterrey Nuevo León, el 31 de Octubre de 1976, mis padres son los señores Irma Alicia Salinas Borrego y Rodolfo Portillo Ruvalcaba. Me gradué de la Facultad de Ingeniería Mecánica y Eléctrica de la Universidad Autónoma de Nuevo León en la que obtuve el título de Ingeniero Administrador de Sistemas en 1998. Actualmente me encuentro desarrollando actividades en el área de soporte a la toma de decisiones en la empresa GAMESA S. DE R.L. de C.V.

Mi perfil está enfocado al análisis de conceptos de negocio para entenderlos y aplicar mis conocimientos tanto técnicos como conceptuales sobre aquellos que se puedan considerar por las empresas como importantes y que proporcionen un valor agregado al negocio o conceptos que sean considerados áreas de oportunidad.



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



DIRECCIÓN GENERAL DE BIBLIOTECAS



100