

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 EXPLORACION 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



1080117202

TM

QA76

.9

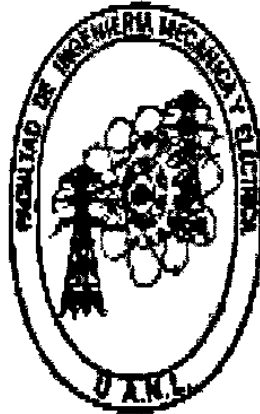
.D37

P6

2002



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”**

POR

ING. RODOLFO PORTILLO SALINAS

TESIS

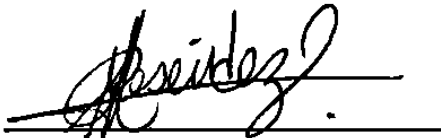
**EN OPCIÓN AL GRADO DE MAESTRO EN CIENCIAS DE LA
ADMINISTRACIÓN CON ESPECIALIDAD EN SISTEMAS**

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



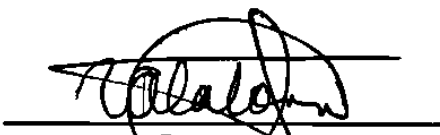
Asesor

M.C. Rosa María Reséndez Hinojosa



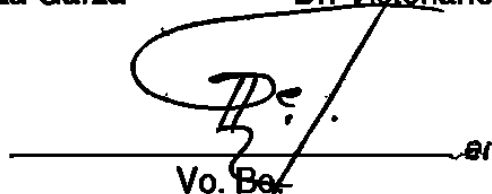
Coasesor

M.C. David Garza Garza



Coasesor

Dr. Victoriano F. Alatorre González



Vo. Bo.

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*

A mis hermanos

*Claudia, Luis y Jesús
Por apoyarme en todo momento y formar parte de mis logros*

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.

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 tecnológica.....	51
2.5.1.2 Herramientas Modeladoras.....	52
2.5.2 Herramientas de almacenamiento.....	52
2.5.3 Productos de acceso.....	59
2.6 El por qué una arquitectura FIC puede ayudar al problema de descentralización de información, inflexibilidad y optimización de un Data Warehouse.....	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