TESIS
IL210
B3

UNIVERSIDAD CATÓLICA ANDRÉS BELLO

FACULTAD DE INGENIERÍA

ESCUELA DE INGENIERÍA INFORMATICA

DISEÑO E IMPLEMENTACIÓN DE UN MODELO MULTIDIMENSIONAL DE DATA MART DEL ÁREA COMERCIAL PARA EMPRESAS DEL SECTOR ELÉCTRICO

TOMO I

REALIZADO POR — **JUSBELYS BARRIENTOS MENDOZA**, **KARINA HENRIQUES TAVARES**

PROFESOR GUIA — **DAVID UZCÁTEGUI**

FECHA — **OCTUBRE DE 2002**

UNIVERSIDAD CATOLICA ANDRES BELLO

FACULTAD DE INGENIERIA

ESCUELA DE INGENIERIA INFORMATICA

DISEÑO E IMPLEMENTACIÓN DE UN MODELO MULTIDIMENSIONAL DE DATA MART DEL ÁREA COMERCIAL PARA EMPRESAS DEL SECTOR ELÉCTRICO

TOMO I

Este Jurado; una vez realizado el examen del presente trabajo ha evaluado su contenido con el resultado: 19 (Diecinueve)

J U R A D O E X A M I N A D O R

Firma:
Nombre: LÚCIA CARDOSO

Firma:
Nombre: Susana García

Firma:
Nombre: David Uzcátegui

REALIZADO POR	**JUSBELYS BARRIENTOS MENDOZA** **KARINA HENRIQUES TAVARES**
PROFESOR GUIA	**DAVID UZCÁTEGUI**
FECHA	**OCTUBRE DE 2002**

DEDICATORIA

A Dios, por darme la fortaleza, serenidad y tesón para seguir siempre adelante y no desvanecer en los momentos difíciles.

A mis papás, Luis y Judith, gracias por estar siempre presentes, por ser un apoyo incondicional y un ejemplo a seguir, y por sobre todo ser mis mejores amigos. Ustedes me han brindado una y otra vez esa fuerza y esa energía para salir victoriosos ante los obstáculos de la vida. ¡Gracias por ser mis padres, los quiero!

A mis hermanos, Luis Manuel y Carlos Alberto, por ser mis compañeros en todo momento, con ustedes quiero compartir este triunfo.

A Karina Henriques, mi compañera de estudios y de tesis, con quien he compartido los mejores momentos de mi carrera universitaria y a quien hoy por hoy considero una de mis mejores amigas... ¡Lo logramos!

Jusbelys

A Dios por darme la vida y los conocimientos.

"Procura adquirir la sabiduría, no la abandones porque ella te protegerá; Ábrele el camino, pues ella te llenará de honores cuando la estreches en tus brazos...."
Proverbios 4:5-4:8

A mis padres, quienes me enseñaron desde pequeña a luchar por alcanzar mis metas. Mi triunfo es el de ellos. ¡Los Quiero!

A mis hermanos, por el cariño y apoyo otorgado durante mi carrera universitaria.

A Aurelio Muscelli, quien me brindó su amor, su cariño, su estímulo y su apoyo constante.

A Jusbelys Barrientos, por su apoyo incondicional en todo momento y por brindarme el mejor de los regalos: Su Amistad.

Karina

AGRADECIMIENTOS

Agradecemos a todas aquellas personas que de una u otra manera colaboraron con nosotras brindándonos su apoyo. En especial a:

- *EGT Consultores Independientes, y a nuestro tutor David Uzcátegui, por abrirnos sus puertas al presente estudio. Igualmente al Lic. Eduardo Mendible por su apoyo y ayuda.*
- *Al Profesor Nelson Rodríguez, quien confió en nosotras y nos dio la oportunidad de realizar este proyecto. Además de asesorarnos y guiarnos en el desarrollo del presente Trabajo, prestándonos su colaboración en todo momento.*
- *A la Profesora Lúcia Cardoso, quien pacientemente nos atendió y asesoró en los momentos difíciles en la construcción de esta Tesis.*

ÍNDICE

	Pág.
Dedicatoria	ii
Agradecimiento	iii
Índice de Tablas	vii
Índice de Figuras	viii
Resumen	ix
INTRODUCCIÓN	01

CAPÍTULOS

I.	SINOPSIS	03
	Planteamiento del Problema	03
	Objetivo General	07
	Objetivos Específicos	07
	Justificación	08
	Limitaciones y Alcances	10
II.	MARCO TEÓRICO REFERENCIAL	12
	1. Sistemas Inteligentes y de Soporte a la Decisión (DSS)	12
	1.1 Definición	13
	1.2 Ventajas de los Sistemas DSS	14
	2. Business Intelligence (BI)	15
	2.1 Definición	15
	3. Data Warehouse (DW)	17
	3.1 Definición	17
	3.2 Data Warehousing	18
	3.3 Características	19
	3.4 Componentes del Data Warehouse	20
	3.5 Arquitectura de Referencia	23
	3.6 A quién va dirigido el Data Warehouse	24
	3.7 Beneficios de contar con un Data Warehouse	24
	3.8 Valor del Data Warehouse	25
	4. Definición Data Mart (DM)	26
	5. Data Warehouse Vs. Data Mart	27
	6. Diferencias entre las Bases de Datos Operacional y el Data Warehouse	29
	7. Moselado Multidimensional (MMD)	31
	7.1 Definición	32
	7.2 Componentes del Modelo Multidimensional	34

7.2.1 Tabla Fact o de Hechos 34
7.2.2 Tablas Look-up o Dimensionales 35
7.3 Pasos Básicos del Modelado Multidimensional 36

III. MARCO METODOLÓGICO 37

Componentes de INTERATIONS 38
Fases Desarrolladas para la Implementación del Data Warehouse 39

IV DESARROLLO 45
1. Fase de Arranque 46
2. Fase de Análisis 48
3. Fase de Diseño 52
4. Fase de Construcción 54
5. Fase de Pruebas 60
6. Implementación 60

V RESULTADOS 61

Conclusiones 63
Recomendaciones 66

BIBLIOGRAFÍA 68
Glosario

TOMO II
APÉNDICES
A. Justificación de la Metodología Iterations 01
B. Línea de Acción de la Metodología Iterations 07
C. Fases de la Metodología Iterations 09
D. Requerimientos Comerciales del Data Mart 11
E. Lineamientos Generales para Seleccionar el Sistema Fuente Óptimo 20
F. Modelos Entidad Relación de las Áreas Comerciales 26
G. Data Warehouse Bus Architecture – Ralph Kimball 63
H. Factibilidad y Disponibilidad de la Información para Poblar el DM 65
I. Evaluación y Selección de la Herramienta para los Usuarios Finales 68
J. Recomendaciones de Infraestructura Informática 73
K. Diseño Físico del Data Mart y Diccionario de Datos 75

L. Corporate Information Factory (CIF) 104
M. Diseño de los Procesos de Extracción, Transformación y Carga 111
N. Cálculo de Volumen de Crecimiento de la Base de Datos del DM Comercial 129
O. Funcionalidades de Oracle Discoverer 151
P. Nombre y Descripción de los Programas de Extracción, Transformación y Carga 156
Q. Ejemplo del Código ETL Aplicado a las Tablas de Hechos y Dimensiones 162
R. Descripción de los Reportes Gerenciales 169
S. Prueba – Validaciones de Datos 245

ANEXOS

A. Misión y Visión de SENECA
B. Procesos del Negocio
C. Manual de Usuario
D. Manual Técnico
E. Herramientas Utilizadas

ÍNDICE DE TABLAS

Pág.

TABLA

1. Diferencias entre Base de Datos Operacional y el Data Warehouse 31

ÍNDICE DE FIGURAS

Pág.

FIGURA

1. Sistemas de Soporte de Decisión (DSS) 14
2. Business Intelligence 16
3. Herramientas de Extracción y Repositorio de Datos 20
4. Componentes del Data Warehouse 23
5. Panorama de la Arquitectura de Referencia del DW 32
6. Visualización del Modelo Dimensional mediante representación en cubo. 34
7. Ejemplo de un Esquema Estrella para una Base de Datos con Dimensiones Tiempo, Producto, Mercado y Cliente 37

UNIVERSIDAD CATÓLICA ANDRÉS BELLO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA INFORMÁTICA

DISEÑO E IMPLEMENTACIÓN DE UN MODELO MULTIDIMENSIONAL DE DATA MART DEL ÁREA COMERCIAL PARA EMPRESAS DEL SECTOR ELÉCTRICO

Autoras: Jusbelys Barrientos Mendoza
Karina Henriques Tavares
Tutor: David Uzcátegui
Fecha: Octubre de 2002

RESUMEN

Este Trabajo de Grado tiene por objetivo el diseño e implementación de un Modelo Multidimensional de Data Mart para el área Comercial de una empresa típica del Sector Eléctrico, con el propósito de apoyar la toma de decisiones estratégicas y tácticas de dicha Corporación. Este proyecto surge de la necesidad imperante de proporcionar a la gerencia media-alta de la empresa información que les permita realizar un análisis profundo de ella a fin de evaluar la efectividad y el desempeño de la gestión comercial y tomar los correctivos que permitan alcanzar los objetivos que los directivos se plantean. Al acercarse a la implementación de un Data Mart, resulta natural pensar en una metodología incremental para abordar su desarrollo. Por esta razón se optó trabajar con una adaptación de la metodología propuesta por Bill Inmon (1999), llamada ITERATIONS – Data Warehouse Development Methodology, la cual aplica un enfoque de desarrollo tipo espiral y es partidaria de la rápida generación de sistemas cada vez más funcionales con intervalos cortos entre versiones sucesivas. Los beneficios obtenidos por la utilización de este tipo de sistemas se basan en el acceso iterativo e inmediato a información estratégica del área de negocios a través de consultas ad-hoc, reportes, análisis y gráficos. Este acercamiento de la información permite al usuario final una toma de decisiones rápida y basada en datos objetivos.

Es necesario destacar que el presente trabajo abarcará sólo hasta la fase de implementación del Data Mart, entendiéndose por implementación a la programación real del software de aplicación, el cual incluye varias revisiones del diseño, creación de prototipos y desarrollo de consultas.

INTRODUCCIÓN

Tomar mejores decisiones de negocios rápidamente es la clave para el éxito en el mercado competitivo que se vive actualmente. Comprendiendo esto, las organizaciones están buscando mejorar sus sistemas de toma de decisiones ya que pueden ser rebasados por el volumen y la complejidad de los datos disponibles provenientes de sus sistemas de producción y transaccionales. El permitir que todos estos datos estén disponibles para toda la audiencia a lo largo de la empresa es hoy por hoy uno de los retos más significativos para los profesionales involucrados en las tecnologías de información.

El problema ante el que nos encontramos, se centra en que muchos de los datos de que disponen las empresas sobre sus clientes están "bloqueados" en cárceles de datos; es decir, bases de datos que han evolucionado como subsistemas independientes. Estas bases de datos son incapaces de ofrecer a la empresa una visión consolidada de quién es el cliente, o incluso de qué productos y servicios están interrelacionados en la base de clientes.

El conjunto de aplicaciones de procesamiento transaccional constituye el subsistema de información operativo o transaccional, en él se procesan de manera automática, grandes volúmenes de datos referentes a las actividades rutinarias, que se almacenan en las bases de datos operativas. De ellas se puede extraer información, fundamentalmente válida para las transacciones del día a día, sirven para apoyar y ejecutar las decisiones operativas que conducen las actividades básicas, pero no sirven para realizar análisis más avanzados, incluso de tipo estratégico, ya que no están

diseñadas para apoyar este tipo de tareas.

A partir de los datos almacenados en estas bases de datos operativas (las cuales suelen ser inconsistentes en la manera en que representan los datos), es posible extraer un cúmulo de conocimientos o informaciones que aporten un valor agregado a la gestión adecuada de la empresa, lo que constituirán los *Data Warehouse.*

El concepto de Data Warehouse se ha desarrollado a partir de la necesidad de disponer de un acceso sencillo e inmediato a determinada información de negocio estructurada y de calidad, que puede ser empleada para la toma de decisiones.

Con esta tecnología se convierten los datos operacionales de una organización en una herramienta competitiva, que permite a los usuarios finales examinar los datos de modo más estratégico, realizar análisis y detección de tendencias, seguimiento de tendencias, seguimiento de medidas críticas, producir informes con mayor rapidez, un acceso más fácil, más flexible y más intuitivo a la información que se necesite en cada momento.

En el presente trabajo, se propone el diseño y posterior implementación de un Data Mart para el área del sector Comercial de una empresa típica del sector eléctrico, el cual responde a la necesidad de contar con información que permita analizar el desarrollo, la situación actual, la evolución del mercado y el desempeño de la gestión comercial, para vincularla con tópicos relacionados del área financiera y técnica, con el propósito de apoyar la toma de decisiones estratégicas y tácticas de la Corporación.

CAPÍTULO I

SINOPSIS

Planteamiento del Problema

Actualmente nos encontramos sumergidos en la era de la información. Hoy día las nuevas necesidades de información de directores y ejecutivos de las distintas organizaciones exigen tomar decisiones basadas en calidad, requiriendo de ellas características como *confiabilidad* (que los datos representen la realidad), *oportunidad* (que la información se encuentre disponible cuando se necesite), *legibilidad* (que se encuentre presentada en un formato que facilite su aprovechamiento) y *globalidad* (que incluya todo lo que se necesite).

En estos últimos años la información ha pasado de ser el proceso de automatización de tareas rutinarias a ser un proceso de verdadera "*informatización*" de las empresas. Ahora la información se convierte en un aliado estratégico de gran importancia para la supervivencia de las empresas. Este ha sido el principal motivo que ha impulsado la difusión de los almacenes de datos, al contar las empresas con una enorme cantidad de datos mal gestionados.

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

Esto conlleva a afirmar que los aspectos relativos a la explotación, el control y la auditoría de los almacenes de datos cobran gran importancia tanto para los especialistas informáticos como para los directivos de las empresas. Todo esto, con base en un plan de seguridad específico, con el fin de garantizar la continuidad de la operación y la disponibilidad del sistema.

Atendiendo a lo planteado anteriormente, las empresas necesitan sistematizar la información para la toma de decisiones, principalmente aquellas en las que se identifican grandes volúmenes de datos, clientes, productos y transacciones. En este tipo de empresas entran las empresas de servicios eléctricos, tal es el caso del Sistema Eléctrico de Nueva Esparta y, específicamente su unidad de dirección *Comercial*, la cual se encarga de proveer un servicio comercial de calidad que sea percibido por los clientes, así como también de satisfacer las expectativas y necesidades de los mismos; esta área es la responsable de mejorar la gestión financiera. En esta organización, los desafíos de los distintos sectores económicos tienen, en general como puntos comunes clientes cada vez más exigentes, cambios cada vez más rápidos y una competencia cada vez más fuerte.

La empresa maneja muchos datos. Esta sobreabundancia tiene como consecuencia directa la necesidad de aprovechar esos valiosos activos. El problema consiste por una parte en seleccionar la información justa y útil y, por otra, referenciar esta información y almacenarla correctamente para ser capaz de recuperarla el día que se necesite.

De igual forma los gerentes comerciales requieren un acceso más rápido a los datos, más reportes y a mayor velocidad, análisis y reacciones oportunas para

administrar el negocio, controlar la calidad del servicio y aumentar los ingresos, incluso con costos adicionales en tecnología de la información, para crear y generar reportes oportunos y de esta manera que la información no pierda su novedad.

Otro factor que incide en la necesidad de implementar una herramienta de soporte de decisión (DSS) es la excesiva complejidad de los modelos relacionales que soportan los sistemas operacionales, (la base de datos operacional se llama OPEN), lo cual dificulta en extremo la extracción de información para llevar a cabo labores de gestión de negocios.

La razón por la cual la alta gerencia considera necesario desarrollar el Data Mart para sustituir el actual Sistema de Gestión Comercial (SGC) es que este último, a pesar de que fue diseñado especialmente para medir la gestión comercial, presenta ciertas dificultades, las cuales no permiten a la gerencia realizar análisis profundos y seguros de la gestión. Las fallas señaladas textualmente en las especificaciones de la empresa son las siguientes:

1. *"Carece de la información de ventas para los años 1998, 1999 e inicios del 2000 correspondiente a los suministros considerados como grandes clientes, toda vez que cuando se implementó el OPEN sólo fue convertida la facturación morosa del cliente para estos casos.*

2. *Rigidez en la forma como se presentan los datos, siendo imposible exportar éstos a otras herramientas que permitan mayor flexibilidad para hacer análisis complejos de la situación.*

3. *Errores en los programas que totalizan la información en tablas diseñadas*

para tal fin, cuyo manejo es difícil de determinar por lo que los programas son difíciles de corregir.

4. Las variables que se tomaron en algunos casos no son suficientes representación de la categorización de la información.".

A través de los años, esta compañía eléctrica ha acumulado grandes cantidades de datos corporativos de tipo operacional. Estos depósitos contienen una mina de oro en cuanto a información histórica relacionada, de la cual pudiera derivarse información con elementos tales como el desempeño corporativo, la competencia, los clientes, los productos, la experiencia y los presupuestos. Es por ello que sus trabajadores necesitan de un sistema basado en el acceso interactivo e inmediato a la información estratégica del área de negocios. Este acercamiento de información al usuario final permite la toma de decisiones rápida y basada en datos objetivos a partir de las bases de datos de la empresa.

Satisfacer a los clientes fue el enfoque para gerenciar la gestión comercial, para ello resulta fundamental implementar una nueva informática de decisión para obtener una mayor comprensión del valor de las informaciones disponibles, definir indicadores de negocio pertinentes para facilitar la toma de decisiones operativas y alcanzar la misión y visión de la empresa.

Es por ello que la empresa del sector eléctrico recurre a la ayuda de EGT Consultores Independientes, quienes son especialistas en la aplicación de tecnologías y de las más avanzadas técnicas gerenciales de transformación, capaz de aprovechar las oportunidades que ofrece el conocimiento, la inteligencia de negocios; con el fin de

transformar las organizaciones conforme a los cambios que actualmente conducen hacia una nueva economía, y de esa forma ayudar a los clientes a alcanzar la máxima ventaja competitiva del mercado.

De la problemática planteada surge la siguiente interrogante:

¿Cómo diseñar e implementar una nueva informática de decisión para que la empresa del sector eléctrico pueda obtener un mayor beneficio de las informaciones disponibles y definir indicadores de negocio que faciliten la toma de decisiones?

Para dar respuesta a la presente interrogante se proponen los siguientes objetivos.

Objetivo General

Desarrollar un modelo genérico que permita implementar Data Marts del área comercial para empresas del sector eléctrico. Caso de estudio: Sistema Eléctrico de Nueva Esparta (SENECA).

Objetivos Específicos

- Conceptualizar un modelo genérico funcional /operacional de Data Mart del área comercial para empresas del sector eléctrico.
- Estudiar, evaluar y elegir las herramientas (software) para los usuarios finales del Data Mart.
- Diseñar y desarrollar un Data Mart comercial para empresas del sector eléctrico.
- Desarrollar procedimientos genéricos para la definición de requerimientos de

hardware y comunicaciones.

Justificación

A lo largo del tiempo en el desarrollo de sistemas, se ha tratado de dar solución a esa necesidad intrínseca de los altos mandos de poseer la información de manera resumida o digerida, que le permita en un vistazo corto y rápido, darse cuenta de todo lo que pasa en los sistemas transaccionales de los niveles bajos de la empresa. Sin embargo, esto no es suficiente ya que actualmente se requiere conocer lo que sucede en todos y cada uno de los procesos tanto críticos como triviales que se dan lugar en una empresa, ya que es el conocimiento de éstos lo que da una visión clara de lo que ocurre en ella.

Buscando la administración de ese conocimiento tan importante en nuestros días, surge una de las opciones que se han implementado para cubrir esta necesidad de las empresas que es lo que se conoce desde hace tiempo como Data Warehouse (DW), el cual es el proceso de extraer datos de diferentes sistemas, funciones y lugares de la empresa; agregando, filtrando, resumiendo, estandarizando, transformando y clarificando los datos extraídos e integrándolos después en una base consolidada. Anthem Inc (citado en Treviño, 2002).

La empresa construye un sistema de soporte a las decisiones con el fin de mejorar su rendimiento. El Data Warehouse debe permitirle ser proactiva en su mercado, es decir, decidir y anticipar en función de la información disponible y capitalizar sobre sus experiencias.

La tecnología de la información, en sus distintos componentes, es fundamental para hacer factible la modernización de procesos de toma de decisión en el sector público. En vista de que el sector eléctrico es una de las entidades gubernamentales esenciales, no está inmune a estas presiones. Por ello debe responder de manera proactiva a estas demandas o enfrentar una ardua tarea en la recaudación de ingresos.

Responder a los clientes externos también tiene un gran sentido empresarial para quienes deben garantizar un suministro de electricidad al menor costo posible y con la calidad requerida por los usuarios.

El uso de la información como un arma estratégica, con soporte de herramientas informáticas, conteniendo aplicaciones analíticas, que ayudan a las organizaciones a maximizar su rendimiento de negocios, generando la eficiencia operativa, forma parte de la Inteligencia del Negocio.

Los sistemas de Data Marts han surgido como respuesta a la problemática de extraer información puntual a partir de datos atómicos almacenados en bases de datos de producción. Uno de los objetivos principales de este tipo de sistemas es servir como base de información para la toma de decisiones.

Los beneficios obtenidos por la utilización de este tipo de sistemas se basan en el acceso interactivo e inmediato a información estratégica de un área de negocios. Un Data Mart produce información correcta para los ejecutivos, gerentes y analistas de un área de negocios de la empresa. Estos usuarios pueden aprovecharlo para construir un rango de estrategias y planes de la empresa, así como también evaluar la eficiencia de la ejecución del plan, desde administrar mejor las relaciones con el cliente hasta modificar las estrategias de comercialización.

Dada la naturaleza del problema, se debe buscar los mecanismos adecuados de sustentación de los efectos de modernización logrados por un programa de asistencia técnica. Para eso es necesario implantar un modelo multidimensional de Data Marts del área comercial para empresas del sector eléctrico que garantice el equilibrio económico: confiabilidad, eficiencia, calidad, equidad, no-discriminación, solidaridad, y transparencia del servicio eléctrico.

Para construir aplicaciones informáticas lo suficientemente flexibles, que igualmente sean competitivas, es necesario reducir los costos y riesgos asociados a las etapas de diseño, por ello es necesario modificar los criterios de diseño; aquí entra la metodología llamada ITERATIONS, diseñada por W.H. Inmon, la cual tiene entre otros los siguientes beneficios:

- Rápida y precisa transferencia de tecnología
- Mejor comunicación entre los miembros del equipo del proyecto
- Simplifica la gerencia del Proyecto
- Reduce riesgos de fallas en proyectos

Limitaciones y Alcances

- Entre las áreas que abarca el sector eléctrico, se encuentran: distribución, mantenimiento, financiera, gestión técnica, comercial entre otras.

El Data Mart se desarrollará para el área comercial, abarcará los siguientes aspectos: Lectura de Medidores, Medidores, Atención al cliente (solicitudes, contratos y reclamos), Facturación, Cobranzas y Órdenes de Servicio.

- Las herramientas a utilizar ya están especificadas por la empresa del sector eléctrico, (Oracle 9i, Base de Datos y Componentes Oracle, Visio, entre otros)

- La metodología a utilizar está determinada por la firma de consultores EGT, la cual se trata de una adaptación de la metodología desarrollada por W.H. Inmon, llamada *ITERATIONS.*

CAPÍTULO II

MARCO TEÓRICO REFERENCIAL

En este capítulo se analiza el problema planteado dentro de un sistema de conocimientos existentes para así poder formularlo teóricamente.

Se hace uso de los ejes conceptuales que permean la propuesta como son: Sistemas Inteligentes y de Soporte a las Decisiones, Inteligencia de Negocios, conceptos y características del Data Warehouse, definición de Data Warehousing; componentes, usuarios, arquitectura, beneficios y valor del Data Warehouse, definición de Data Mart, comparación entre Data Mart y Data Warehouse, Modelado Multidimensional, tablas Fact y Dimensionales y pasos del Modelado Multidimensional.

La conjunción de estos elementos permite construir un sistema de saber generalizado y sistemático de la realidad que se investiga, para darle validez y fiabilidad a la propuesta. El análisis de estos ejes conceptuales permite sistematizar la realidad, así como orientar y crear las bases teóricas de la información.

1. Sistemas Inteligentes y de Soporte a la Decisión (DSS)

El auge de los sistemas de soporte a las decisiones ha causado gran impacto en las organizaciones que buscan un sistema que les ayude u oriente a tomar decisiones. El DSS incluye sistemas que usan cuentas y modelos financieros, modelos representativos y optimización de modelos.

1.1 Definición

Little (1970), define los DSS como un modelo basado en un conjunto de procedimientos para procesar datos y juicios que ayudan a los administradores a tomar decisiones. Bonczek (1980), define DSS como un sistema basado en computadoras con tres componentes clave: lenguaje de sistema (un mecanismo para proveer comunicación entre el usuario y otros componentes del DSS), un sistema de conocimiento (repositorio de conocimiento sobre problemas, datos o procesos) y el sistema procesador de problemas (la unión entre los otros dos componentes). Peribán (2001), define los DSS como un sistema que accede a un almacén de datos, crea una base de datos multidimensional, permitiéndole procesar analíticamente la información en línea (OLAP).

Más concretamente, las herramientas DSS se pueden definir como un sistema integrado de planificación y tratamiento de la información que incorpora la habilidad de interrogar al sistema de información de la empresa en una forma determinada, analizar la información obtenida y predecir, con base a modelos determinados, el impacto de las futuras decisiones antes de llevarlas a la práctica. Normalmente se trata de conjuntos integrados de programas (cálculos estadísticos, consultas de bases de datos, modelización, programación matemática, etc.) que pueden compartir los mismos datos e informaciones (tanto de fuentes internas como externas).

Figura Nº 1. Sistemas de Soporte de Decisión (DSS)
Fuente: Adaptación de BITAM Business Intelligence Tools

La identificación de los problemas, la formulación de las políticas y estrategias son las principales fases que debe tomar en cuenta para poder tener un DSS (Peribán, 2001)

1.2 Ventajas de los Sistemas DSS

- Analizan y procesan importantes volúmenes de información necesaria para la correcta toma de decisiones.
- Rápido acceso a la información.
- Algunos de estos sistemas ya vienen preparados para poner en marcha rápidamente sistemas integrados de medición de objetivos u otros indicadores claves, por áreas estratégicas o áreas de responsabilidad y su control a través de avisos que concentren la atención en los puntos débiles de la organización y permitan actuar directamente sobre ellos.
- Presentan, a través de una interfaz amigable y de fácil manejo, una visión clave de la organización según los factores críticos definidos en su creación.
- Permiten el análisis de datos "on line" y en tiempo real, a través de múltiples

perspectivas: geográfica, productos, distribución, etc; buscar patrones estadísticos significativos: relaciones entre datos, segmentación, etc, construir sistemas de información sinóptica y orientada a la alta dirección o la distribución electrónica de información a través de servidores de documentos.

- Se puede trabajar con información interna de la empresa procedente de distintas bases de datos, hojas de cálculo, informes de texto o cualquier otra fuente, y también con información externa de cualquier tipo (comparaciones sectoriales, evolución de la competencia, etc).

2. Business Intelligence (BI)

Además del soporte de decisión, entendido como algo que ayuda a la gente a tomar ciertas acciones de negocios, tenemos el término más reciente "inteligencia de negocios" (business intelligence). "Inteligencia de negocios o Inteligencia Comercial" es uno de esos apodos que abarcan un conjunto de productos y servicios para acceder a datos, analizarlos, y convertirlos en información (Dyché, 2001).

La Inteligencia de Negocios (Business Intelligence, BI) lucha por eliminar las suposiciones y la ignorancia en las empresas al potenciar las montañas de datos cuantitativos que las empresas reúnen cada día en una variedad de aplicaciones corporativas, Dresner (citado en Romero, 2002).

2.1 Definición

Es la respuesta que la tecnología ofrece para apoyar al tomador de decisiones en la generación y obtención de información relevante, para evaluar indicadores de negocio. Diseñando almacenes de datos independientes con la estructura necesaria

para analizar, desde cualquier punto de vista, la información y, además, estableciendo técnicas de análisis que sean una herramienta sencilla e intuitiva, obligada para que el usuario conozca su negocio directamente. BI se ha convertido en la principal Tecnología de Información que comunica en todos los niveles y apoya a la toma de decisiones, utilizando toda la información disponible dentro y fuera de la empresa (BITAM Business Intelligence Tools).

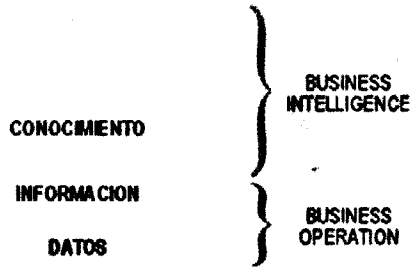

Figura Nº 2. Business Intelligence
Fuente: BITAM Business Intelligence Tools

La BI puede ayudar actualmente a las empresas en por lo menos cuatro áreas (Dresner, 2001):

- **Tener sentido del negocio:** es el proceso que trata de entender qué maneja el negocio (tendencias, modas, anomalías y comportamientos).

- **Medir rendimiento y proyectos:** una vez que se sabe qué medir, se puede usar la BI para poner expectativas a los empleados y ayudarles a seguir y gestionar su propio desempeño. El proceso de presupuestación es un ejemplo, estos se desarrollan con base en los datos del período anterior y usados para manejar pronósticos para el futuro.

- **Mejorar las relaciones:** al proporcionar información útil de la empresa y negocios a los clientes, empleados, proveedores, accionistas, y el público en general,

se incrementa el conocimiento y se fomenta a través de la cadena de información. Al introducir BI en estos grupos, se pueden identificar rápidamente problemas y manejarlos antes de que se conviertan en crisis. La BI también promueve la lealtad: un cliente informado, quien es parte del proceso, seguirá siendo un cliente.

- **Crear una oportunidad de beneficio:** todas las empresas que capturan información de su negocio, pueden vender esta información a alguien.

El mundo de los Decisión Suport Systems, en adelante DSS, ha creado varias arquitecturas de información. La más notable de las estructuras DSS es el Data Warehouse. El Data Warehouse contiene datos históricos que son comunes en toda la organización. La información puede estar sumarizada y/o detallada.

3. Data Warehouse (DW)

3.1 Definición

Un Data Warehouse o almacén de datos puede ser definido de muchas maneras:

- Una colección de piezas clave de información, usadas para manejar y dirigir los negocios para obtener la mayor rentabilidad (Anahory, Sam y Murray, 1997);
- Un conjunto integrado de bases de datos, con orientación temática, que están diseñados para el apoyo a la toma de decisiones, y donde cada unidad de datos es relevante en algún momento del tiempo. (Inmon,1996);
- Una colección integrada de información corporativa diseñada para la recuperación y el análisis en apoyo a los procesos de toma de decisiones (Cerezo, 2000).

En resumen, todos los autores concuerdan que un sistema de Data

Warehouse es un almacén de datos referentes a toda la organización, que genera información homogénea y fiable, en una estructura basada en la consulta y el tratamiento jerarquizado de la misma, y en un entorno diferenciado de los sistemas operacionales; todo con el fin de apoyar al proceso de toma de decisiones (Treviño, 2002).

El reto reside en transformar los enunciados estratégicos generales de la empresa en indagaciones empresariales precisas y después convertirlos en solicitudes y reportes del Data Warehouse (Harjinder y otros, 1996).

Entre las técnicas de análisis de datos están el entender las evaluaciones de información cuantitativa y derivar conclusiones basadas en hechos a partir de la información histórica. Incluyen la habilidad para descubrir patrones y tendencias, para extrapolar las tendencias basadas en los antecedentes a fin de entender e identificar las anomalías o cambios de paradigma, y para presentar recomendaciones de administración coherentes basados en el análisis de la información del Data Warehouse (Harjinder y otros, 1996).

3.2 Data Warehousing

Es el proceso de extraer y filtrar datos de las bases de datos de las operaciones comunes de la empresa, procedentes de los distintos subsistemas operacionales, para transformarlos, integrarlos, sumarizarlos y almacenarlos en un depósito o repositorio, para poder acceder a ellos cada vez que se necesite (Dale, 2002).

Por otro lado, el Data Warehousing o almacenamiento de datos, es el proceso que facilita la creación y explotación de un Data Warehouse, haciendo real la gestión

del conocimiento. Para conseguirlo es necesaria la aplicación de una metodología y la implantación de una arquitectura tecnológica (Treviño, 2002).

Una de las novedades que aporta el Data Warehousing como sistema de análisis de información, es la creación de la Meta Información (metadata). Se trata de un archivo al que se le considera como diccionario de estructura de datos que el administrador del sistema define con el objetivo de asistir en los procesos de consulta a la base de datos. La metadata se adaptará a las definiciones que el usuario utilizará posteriormente en sus interrogaciones al sistema. De esta manera se conseguirá minimizar los complejos procedimientos de definición de nombres de campos, jerarquías y relaciones entre necesidades de información a la que desea acceder cada compañía.

3.3 Características

Según Inmon (1996), reconocido como el padre del Data Warehousing, el Data Warehouse se caracteriza por ser:

Integrado: los datos almacenados en el Data Warehouse deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

Temático: sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales.

Histórico: el tiempo es parte implícita de la información contenida en un Data Warehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el Data Warehouse sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

No volátil: el almacén de información de un Data Warehouse existe para ser leído y no modificado. La información es por tanto permanente, significando la actualización del Data Warehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él, sin aplicar ningún tipo de acción sobre lo que ya existía.

Para mayor ampliación del tema se recomienda la lectura Data Warehousing (http://www.sqlmax.com/dataw1.asp).

3.4 Componentes del Data Warehouse

En la Figura N° 3, se presentan los componentes del Data Warehouse.

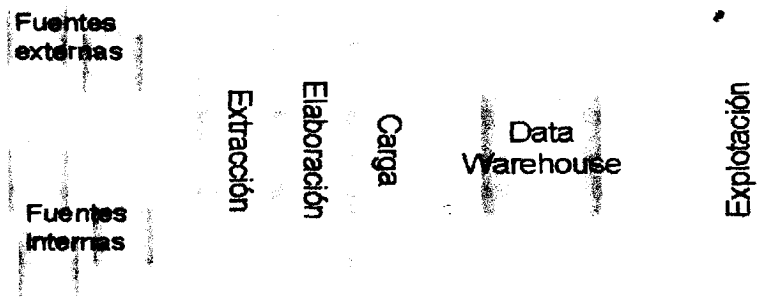

Figura N° 3 . Herramientas de Extracción y Repositorio de Datos
Fuente: http://sie.efpol.ua.es/websie/docdocencia/iti/tema1_iti.pdf

Datos propio, conocido como Modelado Multidimensional o dimensional, el cual busca ofrecer al usuario su visión respecto de la operación del negocio.

Tanto para la eficiencia operativa como para la planeación a futuro, se deben analizar muchos datos empresariales interrelacionados. Esta necesidad empresarial se aborda mediante el procesamiento analítico. En éste, el enfoque está en el análisis de los datos, específicamente en el análisis multidimensional.

7.1 Definición

El Modelado Dimensional es una técnica para modelar bases de datos simples y entendibles al usuario final. La idea fundamental es que el usuario visualice fácilmente la relación que existe entre los distintos componentes del modelo (Wolff, 1999).

En la Figura N° 5, puede observarse cada punto dentro del cubo como una intersección de coordenadas definidas por los lados de éste (dimensiones).

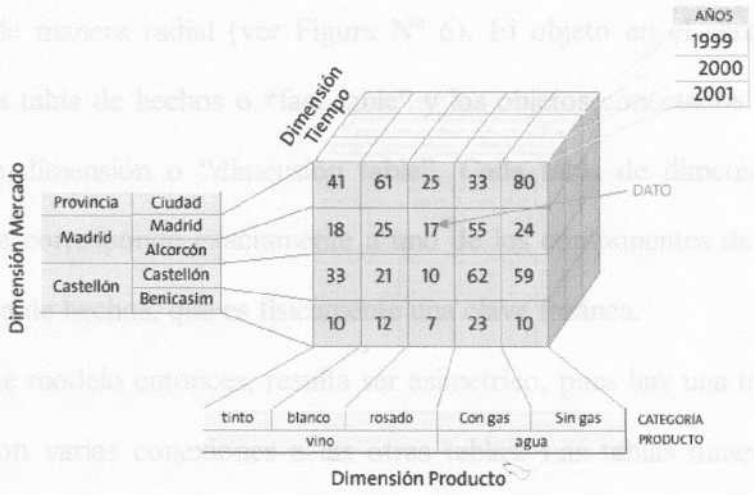

Figura N° 5 . Visualización del Modelo Dimensional mediante representación en Cubo
Fuente: Ferrando, S y otros (2001)

En el modelo multidimensional cada eje corresponde a una dimensión particular. Entonces la dimensionalidad de nuestra base estará dada por la cantidad de ejes (o dimensiones) que le asociemos.

Cuando una base puede ser visualizada como un cubo de tres o más dimensiones, es más fácil para el usuario organizar la información e imaginarse en ella cortando y rebanando el cubo a través de cada una de sus dimensiones, para buscar la información deseada.

El procesamiento analítico o análisis multidimensional se emplea para análisis de históricos complejos, con amplia manipulación (análisis de datos dinámicos), así como para planeación a futuro y pronósticos – el pasado como prólogo del futuro (Harjinder y otros, 1996).

Una de las técnicas más populares del modelado multidimensional se denomina **modelo estrella multidimensional**. Como su nombre lo sugiere, el esquema estrella es un paradigma de modelo que tiene un solo objeto en medio conectado con varios objetos de manera radial (ver Figura N° 6). El objeto en el centro de la estrella se denomina tabla de hechos o "fact table" y los objetos conectados a ella se denominan tablas de dimensión o "dimension table". Cada tabla de dimensión tiene una clave (key) que corresponde exactamente a uno de los componentes de la clave compuesta de la tabla de hechos, que es físicamente una clave foránea.

Este modelo entonces, resulta ser asimétrico, pues hay una tabla dominante en el centro con varias conexiones a las otras tablas. Las tablas dimensión tienen sólo la conexión a la tabla fact y ninguna más.

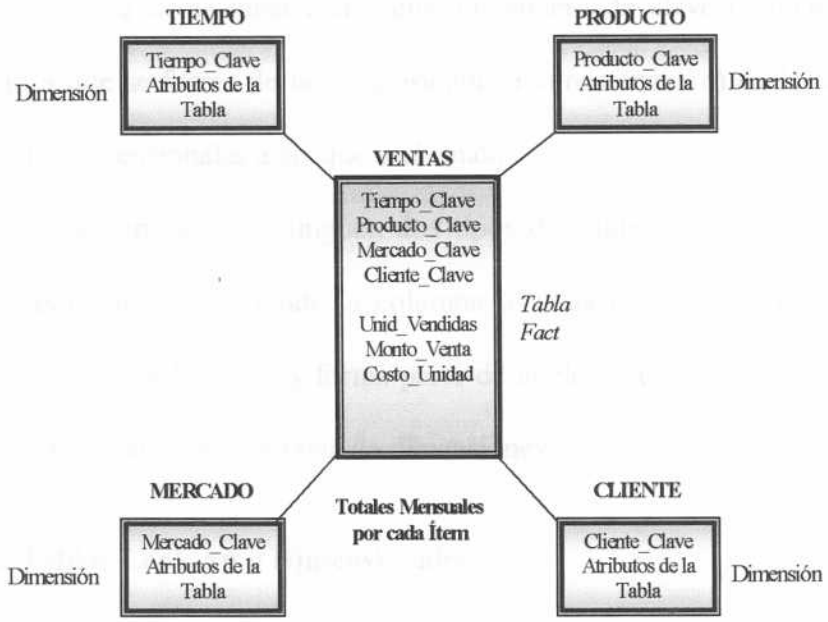

Figura Nº 6. Ejemplo de un Esquema Estrella para una Base de Datos con Dimensiones Tiempo, Producto, Mercado y Cliente
Fuente: Wolf Carmen, 1999

7.2 Componentes del Modelo Multidimensional

7.2.1 Tabla Fact o de Hechos.

Es la tabla central en un esquema dimensional. Es en ella donde se almacenan las mediciones numéricas del negocio. Estas medidas se hacen sobre el grano, o unidad básica de la tabla. El grano o la granularidad de la tabla queda determinada por el nivel de detalle que se almacenará en ella.

Cada medida es tomada de la intersección de las dimensiones que la definen. Idealmente está compuesta por valores numéricos, continuamente evaluados y aditivos. La razón de estas características es que así se facilita que los miles de registros que involucran una consulta sean comprimidos en unas pocas líneas en un set de respuesta.

técnicas difieren del enfoque utilizado para el diseño tradicional de bases de datos operacionales, es decir, el modelo lógico, la aplicación de normas y luego la desactivación selectiva de ellas para mejorar el desempeño (Wolff, 1999).

TABLA Nº 1. Diferencias entre Base las Bases de Datos Operacional y el Data Warehouse

Aspecto	Base de Datos Operacionales	Base de datos del Data Warehouse
Objetivo	Operaciones del día a día.	Toma de decisiones.
Granularidad	Datos detallados.	Datos depurados y sumarizados.
Orientación	Por aplicación o transacciones.	Temáticos, categorías amplias.
Nivel de Integración	Variable, no crítica para OLTP	Racionalización de datos.
Temporalidad	30-60 días (un año máximo)	1-5 años on-line (como norma)
Validez de Datos	Válidos para "ahora".	Válidos para "siempre"
Usuarios	Administrativos/ transacciones	Gestores/ Análisis
Volatilidad	Actualizable.	Sólo accesible una vez cargados.
Procesamiento	Repetitivo y conocido.	Puntual y no conocido.
Manejo	Un registro cada vez / pocos datos.	Manejo masivo de datos.
Disponibilidad	Alta.	Media
Normalización	Cuarta forma normal.	Redundancia.
Tecnología	Transaccional.	Acceso y gestión.

Fuente: Elaboración Propia

7. Modelado Multidimensional (MMD)

La tecnología Data Warehousing debido a su orientación analítica, impone un procesamiento y pensamiento distinto, la cual se sustenta por un modelado de Bases de

DM sea un subconjunto completo y natural del DW Empresarial.

El presente trabajo utilizará esta definición para la construcción de la arquitectura del DM, la cual será explicitada en el desarrollo del mismo.

3.6 A quién va dirigido el Data Warehouse

Dadas las características de un sistema de Data Warehousing, su aplicación puede tener variados fines, en una diversidad de industrias. No obstante, en términos generales, podemos decir que su aplicación más rica corresponde a entornos de empresas en los que se identifican grandes volúmenes de datos, asociados a: cantidad de clientes, variedad de productos y cantidad de transacciones (Dyché, 2001).

El Data Warehouse es utilizado en diversas industrias, en donde lo que se busca es conseguir más clientes, todas tratan de maximizar sus ganancias al tiempo que incrementan su cuota de mercado. En cierto sentido, todas las industrias están en el negocio minorista, todas tratan de inducir los clientes a comprar productos y servicios.

3.7 Beneficios de contar con un Data Warehouse

Gray (citado en Treviño, 2002) menciona que los beneficios que un Data Warehouse puede aportar son:

- Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.
- Facilita la aplicación de técnicas estadísticas de análisis y modelización para encontrar relaciones ocultas entre los datos del almacén obteniendo un valor añadido para el negocio a partir de dicha información.

- Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.

- Simplifica dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.

- Supone una optimización tecnológica y económica en entornos de centro de información, estadística o de generación de informes, con retornos de la inversión espectaculares.

3.8 Valor del Data Warehouse

El valor del Data Warehouse puede ser grande y creciente. Este valor se divide en tres categorías (Harjinder y otros, 1996):

Soporte de decisiones a un costo eficaz: los Data Warehouses producen reportes y consultas ad hoc a partir de los sistemas de producción. Además, los profesionales en tecnología de la información se liberan de dar soporte a los usuarios empresariales. La disponibilidad de datos integrados, limpios y consistentes incrementa la calidad y credibilidad de los reportes y consultas.

Los servicios de soporte de decisiones -procesamiento informático, procesamiento analítico y minería de datos- se emplean también para desarrollar recomendaciones factibles fundamentadas en datos sólidos, no sólo basada en la intuición. Una comprensión plena de las relaciones totales con el cliente provoca mejores servicios a éstos y, en consecuencia, clientes satisfechos. La integración logística ayuda a los gerentes de finanzas a manejar mejor los activos, lo que por lo general conduce a costos reducidos y una mejor recuperación sobre activos. Una mejor

comprensión de los patrones de compra de los clientes ayuda a tener el inventario adecuado en el lugar correcto y en el momento preciso, lo que a su vez mejora la recuperación en activos y en la inversión.

Rearquitectura de un sistema de aplicaciones: la separación de los sistemas de producción del sistema del Data Warehouse para soporte de decisiones ofrece al área de tecnología de la información una oportunidad única para depurar los sistemas de herencia, al tiempo que avanza la arquitectura del sistema de la empresa. La depuración puede aumentar la vida de los sistemas de producción, y posponer o eliminar la necesidad de una actualización, lo que da por resultado una mejor recuperación de la inversión y de activos. La necesidad de cargar el Data Warehouse con datos de calidad -limpios y consistentes- proporciona una retroalimentación para mejorar la calidad de los datos en el sistema de producción. Para algunas empresas, el Data Warehouse es precisamente el primer proyecto de arquitectura cliente/servidor que han estado esperando.

Reingeniería de procesos empresariales: los Data Warehouses también proporcionan a la empresa y a la organización, las medidas requeridas para medir la posición competitiva.

4. Definición Data Mart (DM)

El proceso de implantación de un sistema Data Warehouse, puede adaptarse de forma gradual o departamental creando soluciones específicas para cada área con el objetivo de conseguir resultados operativos a corto plazo. Esta solución departamental se denomina Data Mart.

Según define Meta Group, "Un Data Mart es una aplicación de Data Warehouse, construida rápidamente para soportar una línea de negocio simple". Los Data Marts, tienen las mismas características de integración, no volatilidad, orientación temática y de tiempo variante (histórico) que el Data Warehouse. Representan una estrategia del tipo "divide y vencerás" para ámbitos muy genéricos de un Data Warehouse.

Esta estrategia es particularmente apropiada cuando el Data Warehouse central crece muy rápidamente y los distintos departamentos requieren sólo una pequeña porción de los datos contenidos en él. La creación de estos Data Marts requiere algo más que una simple réplica de los datos: se necesitarán tanto la segmentación como algunos métodos adicionales de consolidación.

5. Data Warehouse Vs. Data Mart

Un DW puede ser concebido, como hemos visto, como el gran almacén de la información corporativa. El DM se define como un DW departamental, orientado a las necesidades de un grupo específico de usuarios y, por tanto, enfocado a un subconjunto de aspectos del negocio. Las actividades de planificación, diseño e implementación son las mismas que las encontradas en la construcción de DW. Igualmente sucede con las cuestiones referentes a metadatos, arquitectura, limpieza de datos, consistencia, integridad, accesibilidad y administración.

En esta arquitectura, el DW corporativo es el repositorio central encargado de asegurar la integridad de los datos. Este ámbito corporativo del DW obliga a que los conceptos de negocio que en él se integran -a través de las distintas iteraciones- sean conceptos validados a nivel de empresa, es decir, toda ella, o dicho de otra forma, todas las áreas de la empresa convergen en un mismo significado para los términos de

negocio y éste es el que se encuentra en el DW. Es condición indispensable, por tanto, que los procesos de validación sean exigentes para asegurar la calidad de la información que se encontrará en el DW ya que de aquí se difundirá al resto de la empresa.

Por el contrario los DMs, están organizados de acuerdo a los departamentos (por ejemplo, financiero, ventas o marketing) y, por consiguiente, para ser realmente útiles, además de los conceptos de negocio globales de la empresa, incorporan conceptos específicos que adquieren significado en su entorno restringido de aplicación. Estos conceptos específicos no necesariamente son aplicables al resto de la empresa, por esta misma razón, al disponer de información circunscrita a un pequeño ámbito, que no obligatoriamente se ha de incorporar al DW corporativo como "verdad" empresarial, las validaciones de la información pueden relajarse más que en el DW. La especialización de los DMs implica que en la mayoría de las ocasiones se incorporen fuentes externas específicas del área de negocio analizada junto con la información disponible procedente del DW.

En un entorno DM el nivel de detalle de la información suele ser menor que el correspondiente al DW, compuesto principalmente de datos agregados o sumarizados, por esta razón, y por el ámbito más reducido, los volúmenes de datos de un DM son sensiblemente inferiores al del DW del que proceden. En este ambiente se adaptan tecnologías multidimensionales y es donde se introduce la utilización de aplicaciones de análisis especializadas.

El DW requiere de ciclos de construcción mayores, debido precisamente a la necesidad de integración y validación a nivel corporativo. Sin embargo, en el DM los

ciclos de creación son menores y más flexibles -el ámbito está más acotado y hay que poner de acuerdo a menos personas-. Esta menor duración en la construcción de un DM constituye la principal tentación que en ocasiones ha llevado a asegurar que puede iniciarse un DW mediante la implementación de DMs. El proceso habitual aconseja comenzar por la construcción del DW corporativo y a partir de éste derivar aquellos DMs que se consideren necesarios.

6. Diferencias entre las Bases de Datos Operacional y el Data Warehouse

El Data Warehouse se organiza de manera diferente a las bases de datos operacionales. El modelo de Data Warehouse (y la administración de datos) difieren de sus equivalentes en las bases de datos operacionales del modo siguiente (Harjinder y otros, 1996):

- El modelo de datos en las bases de datos operacionales se concentra en eliminar la redundancia, coordinar las actualizaciones y soportar transacciones que realizan el mismo tipo de operaciones muchas veces al día sobre los mismos datos. En el Data Warehouse, el modelo de datos se orienta a manejar un amplio rango de consultas y recuperación de información en forma periódica. El usuario final del Data Warehouse realiza muy pocas actualizaciones, si acaso ocurren.

- El modelo de datos en la base de datos operacional contiene muchas normas, para manejar actualizaciones consistentes y mantener la integridad referencial. En el caso del Data Warehouse tiene muy pocas normas, para proporcionar acceso instantáneo sin necesidad de tener que realizar una gran cantidad de uniones.

- El modelo de datos en las bases de datos operacionales es sólo el producto de la

base de datos y de los desarrolladores de aplicaciones. A menudo es complejo y grande. Sin embargo, en el Data Warehouse, los usuarios finales deben entender el modelo de datos, pues la finalidad principal del Data Warehouse es que la visibilidad y el acceso a los datos sea fácil y conveniente.

- Los datos dentro de las bases de datos operacionales son sólo los que se necesitan para las operaciones actuales. Los datos innecesarios se archivan con rapidez para evitar una degradación en el desempeño de los sistemas operacionales. Los datos en el Data Warehouse son tanto operacionales como históricos. Los volúmenes de datos que contiene el Data Warehouse se ubican entre los cientos de megabytes a los cientos de gigabytes o más.

- Las bases de datos operacionales almacenan muy pocos datos derivados. La mayoría de las aplicaciones que operan con base de datos operacionales derivan los datos, siempre que se requiera, sobre la marcha. Por otra parte, un Data Warehouse almacena grandes cantidades de datos derivados para ahorrar el esfuerzo del cálculo repetitivo de las derivaciones. Esto se debe a que las derivaciones se hacen no sólo con los datos operacionales, sino también con los datos históricos.

- Las bases de datos operacionales contienen todos los datos que requiere una empresa para sustentar sus operaciones. Por su parte, un Data Warehouse sólo contiene datos que tiene valor a través del tiempo.

- Los datos operacionales están ligeramente resumidos, a menudo con fines de reporte. Un Data Warehouse precalcula y almacena datos muy resumidos.

Debido a estas diferencias entre las bases de datos operacionales y el Data Warehouse, se han propuesto para este último técnicas de modelado diferente. Estas

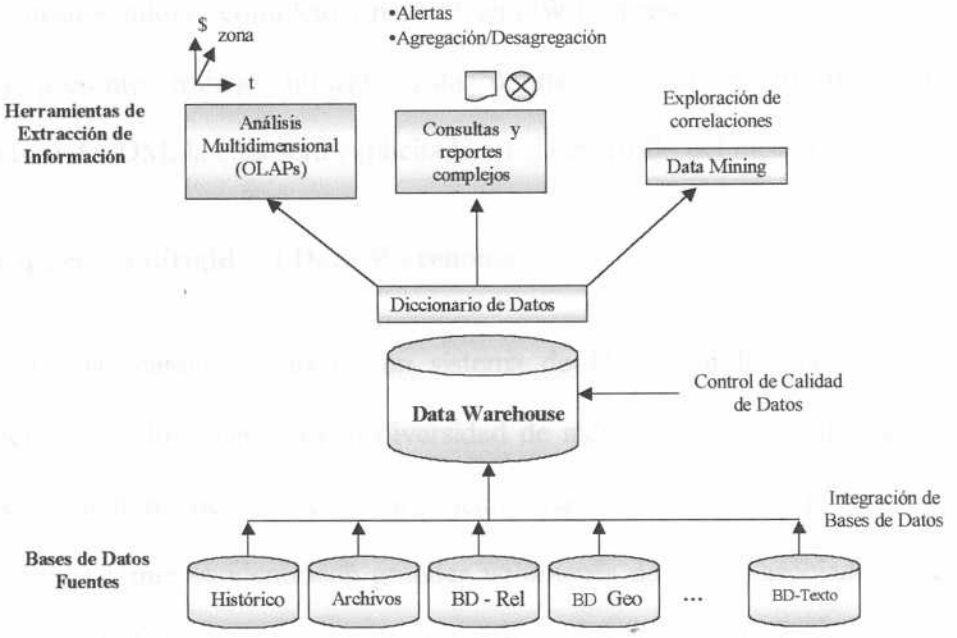

Figura Nº 4. Componentes del Data Warehouse
Fuente: http://sie.efpol.ua.es/websie/docdocencia/iti/tema1_iti.pdf

3.5 Arquitectura de Referencia

La arquitectura de referencia se describe primero desde un punto de vista abstracto y simplificado a alto nivel, del modo siguiente (Harjinder y otros, 1996):

- Un conjunto de datos extraídos de bases de datos operacionales.
- Un software que prepara los datos para que los accedan los usuarios.
- Un conjunto de aplicaciones y herramientas que ejecutan un conjunto de consultas y análisis complejos.

Kimball (2002) desarrolló una arquitectura del Data Warehouse llamada "Data Warehouse Bus Architecture", el cual es un compromiso entre el desarrollo de "todo a la vez" desde una perspectiva central ("Monolithic Approach") versus "construir en cada oportunidad Data Marts (DM) separados" sobre diferentes temas cada vez que sea percibido ("Stovepipe Approach"). Aplicando paso a paso este enfoque, se logra que el

Fuentes de datos: este componente es el que normalmente está presente en las organizaciones, y a partir del cual se realiza la captura de datos que se contemplará en el DW. Estas fuentes de datos pueden ser sistemas operacionales corporativos (*representan el entorno del que se obtienen la mayor parte de los datos significativos de la operativa diaria de la compañía*), sistemas operacionales departamentales y fuentes externas.

Extracción y transformación: este es el componente responsable de que la información pueda moverse, con las transformaciones que sean necesarias, desde las fuentes de datos que acabamos de mencionar, al DW. En este sentido hay que decir que por DW puede entenderse tanto el sistema completo como únicamente las bases de datos en las que se almacenan tanto la información extraída de los sistemas anteriores como los metadatos.

Servidor de datos: también podría denominarse componente de gestión. Los servicios que debe ofrecer incluyen un servicio de mantenimiento de datos y un servicio de distribución para exportar datos del DW a servidores de bases de datos descentralizadas y a otros sistemas de soporte de decisiones de usuario. El componente de gestión también ofrece servicios de seguridad (*respaldo, recuperación*) y monitorización. Generalmente estos servicios utilizan los medios suministrados por el software del sistema operativo y de bases de datos subyacente. El componente de **SGBD** (*Sistema de Gestión de Bases de Datos*) consiste en el software de base de datos que se utilice para mantener y extraer datos. Hay dos enfoques diferentes para el almacenamiento de la información: las bases de datos relacionales y las multidimensionales. Así, tendremos **gestores de bases de datos relacionales**

(*SGBDR*) o **gestores de bases de datos multidimensionales** (*SGBDM*).

Herramientas de acceso: sin las herramientas adecuadas de acceso y análisis, el DW se puede convertir en una amalgama de datos sin ninguna utilidad. Es necesario poseer técnicas y herramientas que capturen los datos importantes de manera rápida y puedan ser analizados desde diferentes puntos de vista. También deben transformar los datos capturados en información útil para el negocio. Actualmente a este tipo de herramientas se las conocen como Business Intelligence Tools (BIT) y están situadas conceptualmente sobre el DW. Cada usuario final debe seleccionar la herramienta que mejor se ajusta a sus necesidades y a su DW. Entre ellas se pueden citar las Consultas SQL (Structured Query Language), las Herramientas MDA (Multidimensional Analysis) y OLAP (On-Line Analytical Processing), entre otras.

Repositorio/Metadatos: los metadatos son básicamente datos acerca de los datos contenidos en el DW. Así, uno de los problemas con el que pueden encontrarse los usuarios de un DW es saber lo que hay en él y cómo pueden acceder a lo que requieren. El repositorio les ayuda a conseguirlo. Es sólo una de las utilidades del repositorio, pero éste tiene muchas funcionalidades: catalogar y describir la información disponible; especificar el propósito de la misma; indicar las relaciones entre los distintos datos; establecer quién es el propietario de la información; relacionar las estructuras técnicas de datos con la información de negocio; establecer las relaciones con los datos operacionales y las reglas de transformación; y limitar la validez de la información.

A continuación se presenta la Figura N° 4, donde se puede visualizar las herramientas de extracción de información y las bases de datos fuentes del DW.

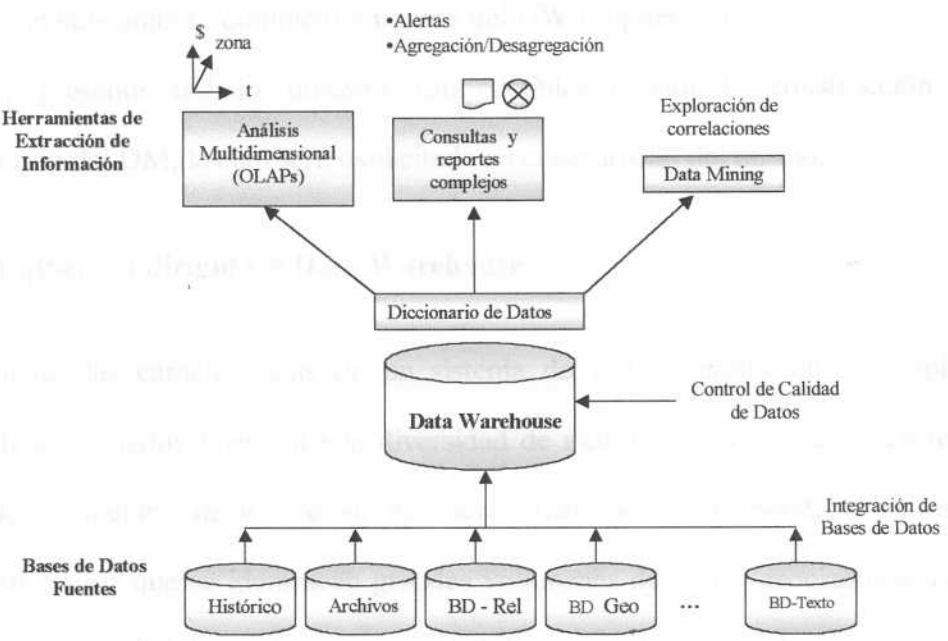

Figura Nº 4. Componentes del Data Warehouse
Fuente: http://sie.efpol.ua.es/websie/docdocencia/iti/tema1_iti.pdf

3.5 Arquitectura de Referencia

La arquitectura de referencia se describe primero desde un punto de vista abstracto y simplificado a alto nivel, del modo siguiente (Harjinder y otros, 1996):

- Un conjunto de datos extraídos de bases de datos operacionales.
- Un software que prepara los datos para que los accedan los usuarios.
- Un conjunto de aplicaciones y herramientas que ejecutan un conjunto de consultas y análisis complejos.

Kimball (2002) desarrolló una arquitectura del Data Warehouse llamada "Data Warehouse Bus Architecture", el cual es un compromiso entre el desarrollo de "todo a la vez" desde una perspectiva central ("Monolithic Approach") versus "construir en cada oportunidad Data Marts (DM) separados" sobre diferentes temas cada vez que sea percibido ("Stovepipe Approach"). Aplicando paso a paso este enfoque, se logra que el

DM sea un subconjunto completo y natural del DW Empresarial.

El presente trabajo utilizará esta definición para la construcción de la arquitectura del DM, la cual será explicitada en el desarrollo del mismo.

3.6 A quién va dirigido el Data Warehouse

Dadas las características de un sistema de Data Warehousing, su aplicación puede tener variados fines, en una diversidad de industrias. No obstante, en términos generales, podemos decir que su aplicación más rica corresponde a entornos de empresas en los que se identifican grandes volúmenes de datos, asociados a: cantidad de clientes, variedad de productos y cantidad de transacciones (Dyché, 2001).

El Data Warehouse es utilizado en diversas industrias, en donde lo que se busca es conseguir más clientes, todas tratan de maximizar sus ganancias al tiempo que incrementan su cuota de mercado. En cierto sentido, todas las industrias están en el negocio minorista, todas tratan de inducir los clientes a comprar productos y servicios.

3.7 Beneficios de contar con un Data Warehouse

Gray (citado en Treviño, 2002) menciona que los beneficios que un Data Warehouse puede aportar son:

- Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.
- Facilita la aplicación de técnicas estadísticas de análisis y modelización para encontrar relaciones ocultas entre los datos del almacén obteniendo un valor añadido para el negocio a partir de dicha información.

- Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.

- Simplifica dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.

- Supone una optimización tecnológica y económica en entornos de centro de información, estadística o de generación de informes, con retornos de la inversión espectaculares.

3.8 Valor del Data Warehouse

El valor del Data Warehouse puede ser grande y creciente. Este valor se divide en tres categorías (Harjinder y otros, 1996):

Soporte de decisiones a un costo eficaz: los Data Warehouses producen reportes y consultas ad hoc a partir de los sistemas de producción. Además, los profesionales en tecnología de la información se liberan de dar soporte a los usuarios empresariales. La disponibilidad de datos integrados, limpios y consistentes incrementa la calidad y credibilidad de los reportes y consultas.

Los servicios de soporte de decisiones -procesamiento informático, procesamiento analítico y minería de datos- se emplean también para desarrollar recomendaciones factibles fundamentadas en datos sólidos, no sólo basada en la intuición. Una comprensión plena de las relaciones totales con el cliente provoca mejores servicios a éstos y, en consecuencia, clientes satisfechos. La integración logística ayuda a los gerentes de finanzas a manejar mejor los activos, lo que por lo general conduce a costos reducidos y una mejor recuperación sobre activos. Una mejor

comprensión de los patrones de compra de los clientes ayuda a tener el inventario adecuado en el lugar correcto y en el momento preciso, lo que a su vez mejora la recuperación en activos y en la inversión.

Rearquitectura de un sistema de aplicaciones: la separación de los sistemas de producción del sistema del Data Warehouse para soporte de decisiones ofrece al área de tecnología de la información una oportunidad única para depurar los sistemas de herencia, al tiempo que avanza la arquitectura del sistema de la empresa. La depuración puede aumentar la vida de los sistemas de producción, y posponer o eliminar la necesidad de una actualización, lo que da por resultado una mejor recuperación de la inversión y de activos. La necesidad de cargar el Data Warehouse con datos de calidad -limpios y consistentes- proporciona una retroalimentación para mejorar la calidad de los datos en el sistema de producción. Para algunas empresas, el Data Warehouse es precisamente el primer proyecto de arquitectura cliente/servidor que han estado esperando.

Reingeniería de procesos empresariales: los Data Warehouses también proporcionan a la empresa y a la organización, las medidas requeridas para medir la posición competitiva.

4. Definición Data Mart (DM)

El proceso de implantación de un sistema Data Warehouse, puede adaptarse de forma gradual o departamental creando soluciones específicas para cada área con el objetivo de conseguir resultados operativos a corto plazo. Esta solución departamental se denomina Data Mart.

Según define Meta Group, "Un Data Mart es una aplicación de Data Warehouse, construida rápidamente para soportar una línea de negocio simple". Los Data Marts, tienen las mismas características de integración, no volatilidad, orientación temática y de tiempo variante (histórico) que el Data Warehouse. Representan una estrategia del tipo "divide y vencerás" para ámbitos muy genéricos de un Data Warehouse.

Esta estrategia es particularmente apropiada cuando el Data Warehouse central crece muy rápidamente y los distintos departamentos requieren sólo una pequeña porción de los datos contenidos en él. La creación de estos Data Marts requiere algo más que una simple réplica de los datos: se necesitarán tanto la segmentación como algunos métodos adicionales de consolidación.

5. Data Warehouse Vs. Data Mart

Un DW puede ser concebido, como hemos visto, como el gran almacén de la información corporativa. El DM se define como un DW departamental, orientado a las necesidades de un grupo específico de usuarios y, por tanto, enfocado a un subconjunto de aspectos del negocio. Las actividades de planificación, diseño e implementación son las mismas que las encontradas en la construcción de DW. Igualmente sucede con las cuestiones referentes a metadatos, arquitectura, limpieza de datos, consistencia, integridad, accesibilidad y administración.

En esta arquitectura, el DW corporativo es el repositorio central encargado de asegurar la integridad de los datos. Este ámbito corporativo del DW obliga a que los conceptos de negocio que en él se integran -a través de las distintas iteraciones- sean conceptos validados a nivel de empresa, es decir, toda ella, o dicho de otra forma, todas las áreas de la empresa convergen en un mismo significado para los términos de

negocio y éste es el que se encuentra en el DW. Es condición indispensable, por tanto, que los procesos de validación sean exigentes para asegurar la calidad de la información que se encontrará en el DW ya que de aquí se difundirá al resto de la empresa.

Por el contrario los DMs, están organizados de acuerdo a los departamentos (por ejemplo, financiero, ventas o marketing) y, por consiguiente, para ser realmente útiles, además de los conceptos de negocio globales de la empresa, incorporan conceptos específicos que adquieren significado en su entorno restringido de aplicación. Estos conceptos específicos no necesariamente son aplicables al resto de la empresa, por esta misma razón, al disponer de información circunscrita a un pequeño ámbito, que no obligatoriamente se ha de incorporar al DW corporativo como "verdad" empresarial, las validaciones de la información pueden relajarse más que en el DW. La especialización de los DMs implica que en la mayoría de las ocasiones se incorporen fuentes externas específicas del área de negocio analizada junto con la información disponible procedente del DW.

En un entorno DM el nivel de detalle de la información suele ser menor que el correspondiente al DW, compuesto principalmente de datos agregados o sumarizados, por esta razón, y por el ámbito más reducido, los volúmenes de datos de un DM son sensiblemente inferiores al del DW del que proceden. En este ambiente se adaptan tecnologías multidimensionales y es donde se introduce la utilización de aplicaciones de análisis especializadas.

El DW requiere de ciclos de construcción mayores, debido precisamente a la necesidad de integración y validación a nivel corporativo. Sin embargo, en el DM los

ciclos de creación son menores y más flexibles -el ámbito está más acotado y hay que poner de acuerdo a menos personas-. Esta menor duración en la construcción de un DM constituye la principal tentación que en ocasiones ha llevado a asegurar que puede iniciarse un DW mediante la implementación de DMs. El proceso habitual aconseja comenzar por la construcción del DW corporativo y a partir de éste derivar aquellos DMs que se consideren necesarios.

6. Diferencias entre las Bases de Datos Operacional y el Data Warehouse

El Data Warehouse se organiza de manera diferente a las bases de datos operacionales. El modelo de Data Warehouse (y la administración de datos) difieren de sus equivalentes en las bases de datos operacionales del modo siguiente (Harjinder y otros, 1996):

- El modelo de datos en las bases de datos operacionales se concentra en eliminar la redundancia, coordinar las actualizaciones y soportar transacciones que realizan el mismo tipo de operaciones muchas veces al día sobre los mismos datos. En el Data Warehouse, el modelo de datos se orienta a manejar un amplio rango de consultas y recuperación de información en forma periódica. El usuario final del Data Warehouse realiza muy pocas actualizaciones, si acaso ocurren.

- El modelo de datos en la base de datos operacional contiene muchas normas, para manejar actualizaciones consistentes y mantener la integridad referencial. En el caso del Data Warehouse tiene muy pocas normas, para proporcionar acceso instantáneo sin necesidad de tener que realizar una gran cantidad de uniones.

- El modelo de datos en las bases de datos operacionales es sólo el producto de la

base de datos y de los desarrolladores de aplicaciones. A menudo es complejo y grande. Sin embargo, en el Data Warehouse, los usuarios finales deben entender el modelo de datos, pues la finalidad principal del Data Warehouse es que la visibilidad y el acceso a los datos sea fácil y conveniente.

- Los datos dentro de las bases de datos operacionales son sólo los que se necesitan para las operaciones actuales. Los datos innecesarios se archivan con rapidez para evitar una degradación en el desempeño de los sistemas operacionales. Los datos en el Data Warehouse son tanto operacionales como históricos. Los volúmenes de datos que contiene el Data Warehouse se ubican entre los cientos de megabytes a los cientos de gigabytes o más.

- Las bases de datos operacionales almacenan muy pocos datos derivados. La mayoría de las aplicaciones que operan con base de datos operacionales derivan los datos, siempre que se requiera, sobre la marcha. Por otra parte, un Data Warehouse almacena grandes cantidades de datos derivados para ahorrar el esfuerzo del cálculo repetitivo de las derivaciones. Esto se debe a que las derivaciones se hacen no sólo con los datos operacionales, sino también con los datos históricos.

- Las bases de datos operacionales contienen todos los datos que requiere una empresa para sustentar sus operaciones. Por su parte, un Data Warehouse sólo contiene datos que tiene valor a través del tiempo.

- Los datos operacionales están ligeramente resumidos, a menudo con fines de reporte. Un Data Warehouse precalcula y almacena datos muy resumidos.

Debido a estas diferencias entre las bases de datos operacionales y el Data Warehouse, se han propuesto para este último técnicas de modelado diferente. Estas

técnicas difieren del enfoque utilizado para el diseño tradicional de bases de datos operacionales, es decir, el modelo lógico, la aplicación de normas y luego la desactivación selectiva de ellas para mejorar el desempeño (Wolff, 1999).

TABLA N° 1. Diferencias entre Base las Bases de Datos Operacional y el Data Warehouse

Aspecto	Base de Datos Operacionales	Base de datos del Data Warehouse
Objetivo	Operaciones del día a día.	Toma de decisiones.
Granularidad	Datos detallados.	Datos depurados y sumarizados.
Orientación	Por aplicación o transacciones.	Temáticos, categorías amplias.
Nivel de Integración	Variable, no crítica para OLTP	Racionalización de datos.
Temporalidad	30-60 días (un año máximo)	1-5 años on-line (como norma)
Validez de Datos	Válidos para "ahora".	Válidos para "siempre"
Usuarios	Administrativos/ transacciones	Gestores/ Análisis
Volatilidad	Actualizable.	Sólo accesible una vez cargados.
Procesamiento	Repetitivo y conocido.	Puntual y no conocido.
Manejo	Un registro cada vez / pocos datos.	Manejo masivo de datos.
Disponibilidad	Alta.	Media
Normalización	Cuarta forma normal.	Redundancia.
Tecnología	Transaccional.	Acceso y gestión.

Fuente: Elaboración Propia

7. Modelado Multidimensional (MMD)

La tecnología Data Warehousing debido a su orientación analítica, impone un procesamiento y pensamiento distinto, la cual se sustenta por un modelado de Bases de

Datos propio, conocido como Modelado Multidimensional o dimensional, el cual busca ofrecer al usuario su visión respecto de la operación del negocio.

Tanto para la eficiencia operativa como para la planeación a futuro, se deben analizar muchos datos empresariales interrelacionados. Esta necesidad empresarial se aborda mediante el procesamiento analítico. En éste, el enfoque está en el análisis de los datos, específicamente en el análisis multidimensional.

7.1 Definición

El Modelado Dimensional es una técnica para modelar bases de datos simples y entendibles al usuario final. La idea fundamental es que el usuario visualice fácilmente la relación que existe entre los distintos componentes del modelo (Wolff, 1999).

En la Figura N° 5, puede observarse cada punto dentro del cubo como una intersección de coordenadas definidas por los lados de éste (dimensiones).

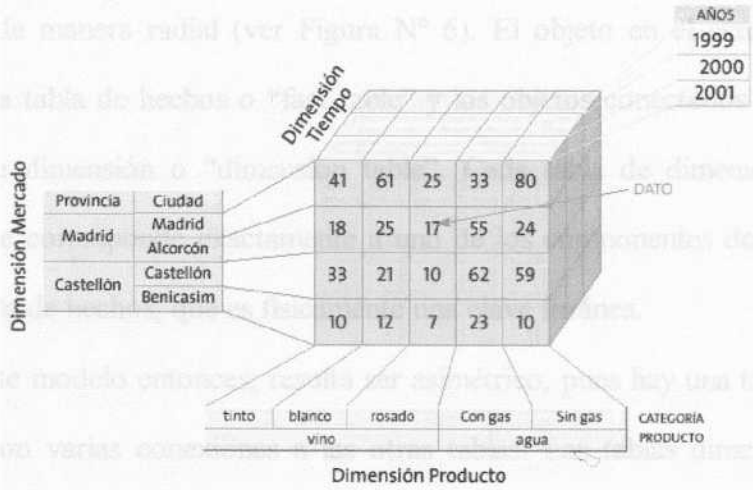

Figura N° 5 . Visualización del Modelo Dimensional mediante representación en Cubo
Fuente: Ferrando, S y otros (2001)

En el modelo multidimensional cada eje corresponde a una dimensión particular. Entonces la dimensionalidad de nuestra base estará dada por la cantidad de ejes (o dimensiones) que le asociemos.

Cuando una base puede ser visualizada como un cubo de tres o más dimensiones, es más fácil para el usuario organizar la información e imaginarse en ella cortando y rebanando el cubo a través de cada una de sus dimensiones, para buscar la información deseada.

El procesamiento analítico o análisis multidimensional se emplea para análisis de históricos complejos, con amplia manipulación (análisis de datos dinámicos), así como para planeación a futuro y pronósticos – el pasado como prólogo del futuro (Harjinder y otros, 1996).

Una de las técnicas más populares del modelado multidimensional se denomina **modelo estrella multidimensional**. Como su nombre lo sugiere, el esquema estrella es un paradigma de modelo que tiene un solo objeto en medio conectado con varios objetos de manera radial (ver Figura N° 6). El objeto en el centro de la estrella se denomina tabla de hechos o "fact table" y los objetos conectados a ella se denominan tablas de dimensión o "dimension table". Cada tabla de dimensión tiene una clave (key) que corresponde exactamente a uno de los componentes de la clave compuesta de la tabla de hechos, que es físicamente una clave foránea.

Este modelo entonces, resulta ser asimétrico, pues hay una tabla dominante en el centro con varias conexiones a las otras tablas. Las tablas dimensión tienen sólo la conexión a la tabla fact y ninguna más.

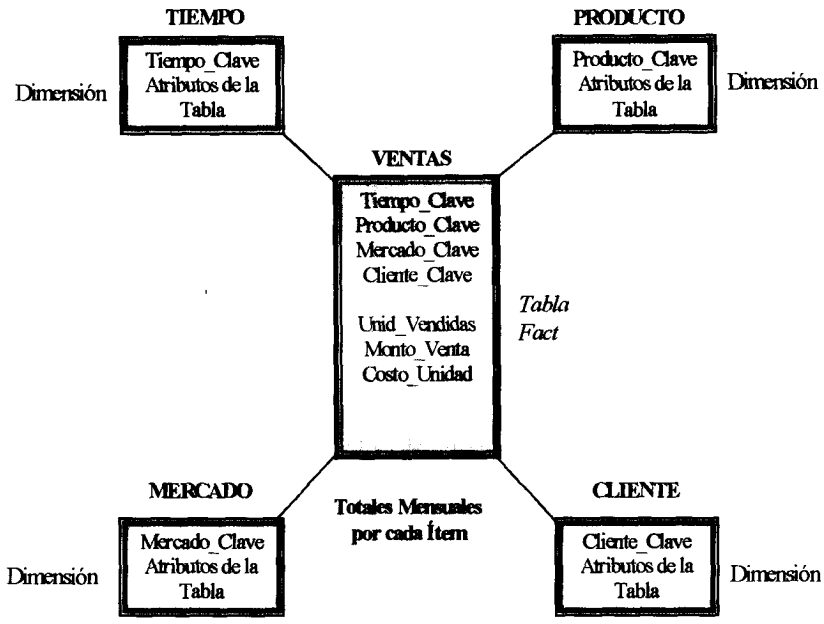

Figura N° 6. Ejemplo de un Esquema Estrella para una Base de Datos con Dimensiones Tiempo, Producto, Mercado y Cliente
Fuente: Wolf Carmen, 1999

7.2 Componentes del Modelo Multidimensional

7.2.1 Tabla Fact o de Hechos.

Es la tabla central en un esquema dimensional. Es en ella donde se almacenan las mediciones numéricas del negocio. Estas medidas se hacen sobre el grano, o unidad básica de la tabla. El grano o la granularidad de la tabla queda determinada por el nivel de detalle que se almacenará en ella.

Cada medida es tomada de la intersección de las dimensiones que la definen. Idealmente está compuesta por valores numéricos, continuamente evaluados y aditivos. La razón de estas características es que así se facilita que los miles de registros que involucran una consulta sean comprimidos en unas pocas líneas en un set de respuesta.

La clave de la tabla fact recibe el nombre de clave compuesta o concatenada debido a que se forma de la composición (o concatenación) de las claves primarias de las tablas dimensionales a las que está unida.

Así entonces, se distinguen dos tipos de columnas en una tabla fact: columnas fact y columnas key. Donde la columna fact es la que almacena alguna medida de negocio y una columna key forma parte de la clave compuesta de la tabla y equivale a una clave primaria de una tabla de dimensiones.

7.2.2 Tablas Look-up o Dimensionales

Estas tablas son las que se conectan a la tabla fact, es decir, las que la alimentan. Una tabla look-up almacena un conjunto de valores que están relacionados a una dimensión particular. Las tablas look-up no contienen hechos, en su lugar los valores en las tablas look-up son los elementos que determinan la estructura de las dimensiones. Así entonces, en ellas existe el detalle de los valores de la dimensión respectiva.

Una tabla look-up está compuesta de una clave primaria que identifica unívocamente una fila en la tabla junto con un conjunto de atributos, y dependiendo del diseño del modelo multidimensional puede existir una clave foránea que determina su relación con otra tabla look-up.

Para decidir si un campo de datos es un atributo o un hecho se analiza la variación de la medida a través del tiempo. Si varía continuamente implicaría tomarlo como un hecho, caso contrario será un atributo.

Los atributos dimensionales son un rol determinante en el modelo. Ellos son la

fuente de todas las necesidades que debieran cubrirse. Esto significa que la base de datos será tan buena como lo sean los atributos dimensionales.

7.3 Pasos básicos del Modelado Multidimensional

- Decidir cuáles serán los procesos de negocios a modelar, basándose en el conocimiento de éstos y de los datos disponibles.

- Decidir el grano de la tabla fact de cada proceso de negocio. En este punto se debe tener especial cuidado con la magnitud de la base de datos, con la información que se tiene y con las preguntas que se quiere responder. El grano decidirá las dimensiones del modelo. Cada dimensión debe tener el grano más pequeño que se pueda puesto que las preguntas que se realicen necesitan cortar la base en caminos precisos.

- Decidir las dimensiones a través del grano. Un grano bien elegido determina la dimensionalidad primaria de la tabla fact. Es posible usualmente agregar dimensiones adicionales al grano básico de la tabla fact, donde estas dimensiones adicionales toman un solo valor para cada combinación de las dimensiones primarias. Si se reconoce que una dimensión adicional deseada viola el grano por causar registros adicionales a los generados, entonces el grano debe ser revisado para acomodar esta dimensión adicional.

- Elegir las mediciones del negocio para la tabla fact. Se deben establecer los ítems que quedarán determinados por la clave compuesta de la tabla fact.

CAPÍTULO III

MARCO METODOLÓGICO

De acuerdo a la forma de implementación, se deben considerar y asociar metodologías congruentes con el desarrollo de incrementos dirigidos a grupos específicos en las organizaciones.

Al acercarse a la implementación de un DW, resulta natural pensar en una metodología incremental para abordar su desarrollo. Por esta razón se optó trabajar con una adaptación de la metodología propuesta por Bill Inmon (1999), llamada *ITERATIONS – Data Warehouse Development Methodology.*

La metodología Iterations aplica un enfoque de desarrollo tipo espiral y es partidaria de la rápida generación de sistemas cada vez más funcionales con intervalos cortos entre versiones sucesivas (ver Fig. Nº 7). El intervalo entre versiones se emplea para identificar e implementar funcionalidad adicional.

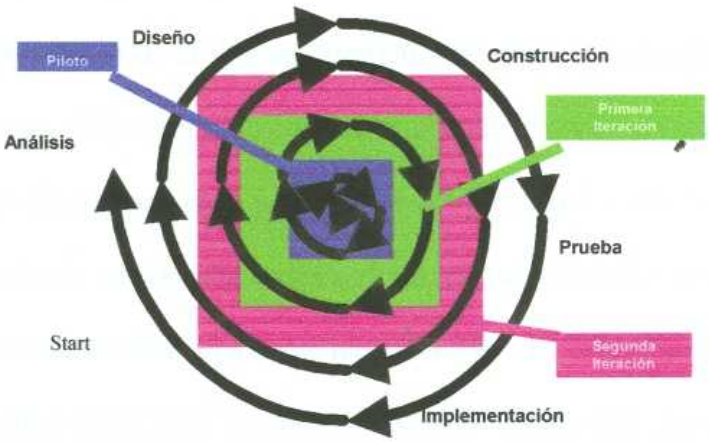

Figura Nº 7. Metodología Iterations
Fuente: Inmon, B (1999)

ITERATIONS fue diseñada con el fin de identificar y documentar cada una de las fases y pasos a seguir para la implementación efectiva de un DW.

En el Apéndice A, se expone de una manera exhaustiva el por qué de la utilización de esta metodología, así como los testimonios de varios clientes que mencionan los beneficios del uso de la metodología *Iterations*.

Componentes de ITERATIONS

- **Líneas de Acción (Tracks).** Otra forma de organizar la metodología es mediante *tracks*; esto facilita ver la metodología desde una perspectiva diferente – basado en un conjunto de actividades acordes al tipo de esfuerzo requerido. Los tracks representan conjuntos de módulos que deben desarrollarse en paralelo, este componente permite: agrupación de responsabilidades paralelas, división del proyecto en unidades de trabajo manejables que se desarrollan iterativamente e ilustración de la magnitud del DW asociado con el Proyecto (Ver Apéndice B).

- **Fases.** Las fases en ITERATIONS representan agrupaciones de módulos que se completan en forma concertada y a menudo muestran muchas interdependencias entre los módulos componentes de la fase en cuestión y la previa. La conclusión de una fase resulta normalmente en un punto de control en el cual se revisa y verifica la obtención exitosa de los productos previstos y el avance del proceso de DW. Este componente permite: agrupación de responsabilidades seriales, participación del recurso correcto en el tiempo preciso e ilustración del alcance de la iteración del DW (Ver Apéndice C).

- **Actividades.** Cada fase contiene actividades requeridas para completar el esfuerzo del DW. Estas actividades son el nivel más bajo de detalle dentro de la

metodología, y todas las formas y herramientas se ponen en correlación con una actividad específica.

Fases desarrolladas para la Implementación del Data Warehouse

1. Fase de Arranque (S - Startup). El objetivo de esta fase es discutir, validar y acordar el alcance del proyecto, así como el plan de actividades a desarrollar, que permitirá obtener, analizar y distribuir los productos generados por la solución.

1.1. Desarrollo de los objetivos empresariales. El primer paso para efectuar cualquier tarea importante y compleja consiste en desarrollar una lista de objetivos empresariales que el sistema debe satisfacer. Considerar el DW como un "producto" interno para una empresa es una simplificación, es un sistema de software cuyos objetivos son apoyar decisiones para una audiencia puramente interna.

1.2. Selección de un ámbito inicial de implementación. Después de definir un rumbo general y un conjunto de objetivos para el DW, se hace necesario derivar con rapidez un ámbito limitado para su implementación. El cual, a grandes rasgos, se divide en dos categorías principales:

- *Ámbito desde la perspectiva del usuario empresarial.* Deben formularse las siguientes preguntas: ¿Cuáles son los departamentos que necesitan utilizar el DW?, ¿Cuál es el rango de consultas empresariales a las que deben responder inicialmente el DW?

- *Ámbito desde la perspectiva de la tecnología.* Deben formularse las siguientes preguntas:¿Cuáles y cuántas son las fuentes de entrada de datos?, ¿Qué tan utilizables son los datos de las fuentes?

1.3 Enfoque de selección de una arquitectura. El desarrollo de esta fase se realiza mediante una serie de actividades:

1.3.1 Verificar la arquitectura técnica actual de la organización.

1.3.2 Arquitectura técnica propuesta para DW.

2. Fase de Análisis (A - Análisis). La fase de análisis del ciclo de desarrollo del Data Warehouse significa convertir los requerimientos acopiados en la fase anterior de requerimientos, en un conjunto de especificaciones que puedan apoyar el diseño. A continuación se mencionan las actividades que se deben realizar para cumplir el objetivo de esta fase.

2.1 Análisis de Requisitos Comerciales. El objetivo de un Data Warahouse es proporcionar un marco tecnológico y organizativo integrado, para la entrega de información a los profesionales de la empresa. Para lograr dicho objetivo, es necesario entender los requisitos informativos estratégicos de la organización. Durante la aplicación inicial de este módulo se capturan las metas corporativas, objetivos departamentales y requisitos del negocio para ser satisfechos por el Data Warehouse. Basado en las necesidades de los usuarios finales y sus prioridades, se identifican las preguntas de negocio significantes y las métricas que encaminan la iniciativa del Data Warehouse. Actividades:

2.1.1 Analizar la estrategia de la organización.

2.1.2 Entrevistas a los expertos de la materia.

2.1.3 Determinar los requisitos funcionales.

2.2 Análisis del Sistema Fuente. Durante este módulo, los sistemas operacionales y otras fuentes potenciales de datos son identificados y analizados,

según la información recabada en la fase de análisis de requerimientos del negocio. Actividades:

2.2.1 Recopilar información relacionada con los sistemas fuentes.

2.2.2 Evaluar las características de los sistemas fuentes.

2.3 Análisis de Datos. El análisis de datos involucra específicamente los datos relevantes y el uso de técnicas de modelado de datos para descubrir las relaciones entre ellos requeridas por el Data Warehouse para apoyar el análisis de negocios. Además del modelado, es importante determinar la viabilidad de integrar los datos de las fuentes identificadas. Actividades:

2.3.1 Revisar los modelos de datos operacionales de la empresa.

2.3.2 Crear el modelo lógico de datos del negocio. (Diagrama entidad-relación, que representa todas las entidades del área del negocio, y cómo ellas se relacionan a través del contexto de las reglas del negocio; así como los atributos de cada entidad).

2.3.3 Identificar la técnica de modelado de datos.

2.3.4 Transformar el modelo de datos lógico en el modelo de datos del Data Warehouse.

2.3.5 Determinar el origen y destino de los datos.

2.4 Diseño de la Arquitectura Técnica del Data Warehouse. Basado en la arquitectura técnica realizada durante la fase de Arranque, se desarrolla un plan detallado del ambiente técnico para el Data Warehouse. Se recomienda el software, hardware y componentes de red requeridos.

2.4.1 Evaluar y seleccionar las herramientas para los usuarios finales.

2.4.2 Seleccionar los componentes de la arquitectura técnica de hardware y software.

2.4.3 Desarrollar un diseño de la arquitectura técnica.

3. Fase de Diseño (D - Design). En la fase de diseño, los modelos lógicos desarrollados en la fase de análisis se convierten en modelos físicos. Los procesos identificados en la fase de análisis para conectar las fuentes de datos con el Data Warehouse y el Data Warehouse con las herramientas de estación de trabajo del usuario final, se convierten en diseños para programas que realizarán las tareas requeridas por los procesos. Para la realización de esta fase es necesario cumplir con las siguientes actividades:

3.1 Desarrollo de modelos físicos. Desarrollar los modelos físicos de datos para las bases de datos de almacenamiento del Data Warehouse.

3.2 Desarrollar la correspondencia de los modelos físicos de datos de las fuentes de datos con los modelos físicos del Data Warehouse. Esta correspondencia ayuda a los procesos de extracción y refinamiento a efectuar sus funciones dentro del Data Warehouse.

3.3 Diseño de la interfaz de la aplicación. Diseñar las soluciones para acceder a los datos del DW, el cual reúne la información del negocio para los usuarios finales y definir el diseño de los reportes gerenciales de acuerdo a la información identificada en la fase de análisis.

3.4 Diseñar programas, procesos y métodos para la extracción, transformación, depuración y carga de datos. El propósito es examinar el diseño físico de la base de datos para proceder a la extracción de los elementos de datos específicos. La limpieza o depuración de los datos en el sistema de información es uno de los aspectos más importantes, consiste en eliminar parte de los datos operacionales,

tales como información de transacciones de bajo nivel. Además se diseñan las rutinas de carga (transporte) de los datos.

3.5 Diseño Operacional de la base de datos. Estimar el volumen de datos esperado, con los que va a ser poblada la aplicación del DW. También, planear los procesos operacionales para el ambiente de DW, equipo de desarrollo y la comunidad de usuarios finales.

4. Fase de Construcción (C - Construction). La fase de construcción es responsable de implementar físicamente los diseños desarrollados durante la fase de diseño. La construcción de un Data Warehouse lleva implícitos los siguientes pasos: extracción de los datos del sistema operacional, depuración y transformación de los datos, carga de los datos y explotación del Data Warehouse mediante diversas técnicas y aplicando herramientas especializadas.

4.1 Generación de la base de datos de desarrollo. Después de que la base de datos lógica se ha diseñado, es necesario convertirlo en un esquema físico.

4.2 Prueba de construcción de los datos. Seleccionar y cargar los datos prueba a ser usados en la unidad, sistema inicial y la prueba de integración del acceso a los datos y los procesos de adquisición de los datos.

4.3 Construcción de procesos para el acceso a los datos. Crear programas, procesos y métodos necesarios para facilitar el acceso de los datos en el Data Warehouse para los usuarios finales.

4.4 Construcción de programas, procesos y métodos para la extracción, transformación y carga de datos.

4.5 Construcción de los reportes gerenciales

4.6 Elaboración de Manuales (Manual de Usuario y Manual Técnico)

5. Fase de Pruebas (T - Test). Verificar los programas, métodos y scripts de la implementación utilizando grandes cargas de datos. Actividades:

5.1 Prueba del Sistema de DW

5.1.2 Pruebas de integración de grandes cantidades de data.

5.1.3 Verificar data en pruebas de carga

5.2 Carga completa y actualización del DW

6. Fase de Implementación (I - Implementation). La implementación se establece como la construcción de la aplicación. La actividad sólo lleva a la práctica el sistema que se modeló en la fase de diseño. Esta fase incluye las actividades de codificación e integración de los diferentes módulos constitutivos del sistema.

CAPÍTULO IV

DESARROLLO

El desarrollo del proyecto del DM se llevó a cabo en dos iteraciones. La primera, tuvo como propósito que el usuario pudiese comenzar a acostumbrarse a los conceptos de Inteligencia de Negocios mediante el manejo de la información contenida en un primer alcance del Data Mart comercial, para generar una retroalimentación que ayudara a pulir el modelo originalmente diseñado. En el momento en que el usuario dispone de una herramienta más potente para analizar la información, así como nuevas fuentes de datos integradas, es cuando comienza también a imaginar nuevas preguntas al sistema, que antes eran simplemente irrealizables.

La metodología utilizada, ITERATIONS, apoya esta filosofía, la cual consiste en desarrollar, en la primera iteración, una versión preliminar del DM Comercial y, en la segunda iteración, a medida que la gerencia adquiera conocimiento y experiencia en las técnicas de DW y en la operación y uso de la información suministrada por la primera versión del DM, se desarrollarán versiones más amplias.

El ámbito de la primera iteración del DM quedó circunscrito a las siguientes áreas: lectura, facturación (conceptos y anomalías) y cuentas por cobrar.

Dicha iteración estuvo conformada por un grupo de trabajo de EGT Consultores, quienes se encargaron de recopilar la información (requisitos comerciales) y diseñar el primer esbozo del DM de las áreas antes mencionadas. Es importante acotar que las autoras del presente trabajo participaron en la implementación de esta primera versión del proyecto.

Posteriormente se comenzó a desarrollar la segunda iteración del sistema, la cual se basó en su crecimiento global, tanto en lo que se refiere a funcionalidad añadida como al propio crecimiento del volumen de información.

Los nuevos módulos incorporados a esta etapa contempla las áreas de: medidores, órdenes de servicios y atención al cliente (solicitudes de servicios, contratos y reclamos).

Esta iteración estuvo a cargo de las autoras de este trabajo de grado, quienes asumieron la responsabilidad y el control de esta etapa del proyecto del DM.

A continuación se mencionan y explican cada uno de los pasos necesarios para el diseño e implementación del DM Comercial, con base en las actividades de la metodología ITERATIONS.

1. Fase de Arranque

1.1 Desarrollo de los objetivos empresariales

Para cumplir esta actividad fue necesario entender qué maneja el negocio (tendencias, modas, anomalías y comportamientos) de la compañía en estudio. Así como también estudiar la misión, el mercado (clientes) y el área de la estructura organizativa que nos compete, la Dirección Comercial (Esto se puede visualizar en el Anexo A).

En esta fase se definieron los objetivos y el alcance del proyecto del DM Comercial, llegando de común acuerdo con la alta gerencia que el enfoque de desarrollo del DM estaría basado en iteraciones y fundamentado en una metodología especializada en el análisis, diseño, tecnología y gerencia de DW, como lo es el caso de ITERATIONS, propuesta por Bill Inmon. (En el Apéndice A se especifica el por

qué de la utilización de esta metodología).

1.2 Selección de un ámbito inicial de implementación

- Ámbito a partir de la perspectiva del usuario empresarial

Se estudiaron las posibles áreas de la empresa que estuviesen relacionadas con el proyecto, y se concluyó que el departamento involucrado en el desarrollo del DM es la *Dirección Comercial*, que es el responsable de proveer un servicio comercial de calidad que sea percibido por los clientes. Esta dirección está apoyada por la Gerencia de Atención al Cliente, la Gerencia de Mercadeo Corporativo y la Coordinación de Crédito y Cobranza. Los usuarios principales del Sistema son: el Gerente General, el Gerente Comercial, el Director de Finanzas, el Gerente de Administración y el Gerente de Sistemas, quien a su vez tiene a cargo la operación y mantenimiento del DM.

Las consultas a las que debe responder el DM Comercial corresponde a las áreas de: Órdenes de Servicios, Facturación, Lectura, Cuentas por Cobrar, Reclamos, Contratos, Solicitudes de Servicios y Medidores.

- Ámbito con base en consideraciones tecnológicas

En el desarrollo de esta actividad se estudiaron las consideraciones tecnológicas del DM, en donde la base de datos debe mantener información sobre diversas áreas de mercadeo y de gestión comercial de la empresa, para facilitar el análisis y apoyar la toma de decisiones de la gerencia media y alta. La información se extrajo del sistema operacional SGC (la cual es la única fuente del DM).

Desde el punto de vista de la plataforma tecnológica del DM, a esta altura del proyecto, se estimó en el uso de un manejador de base de datos, que puede ser relacional o multidimensional y en la utilización de una herramienta front-end, sobre

un sistema operativo MS Windows 2000.

1.3 Enfoque de selección de una arquitectura

1.3.1 Verificar la arquitectura técnica actual de la organización

1.3.2 Arquitectura técnica propuesta para DW

Las actividades 1.3.1 y 1.3.2 se realizaron en forma paralela, ya que los detalles de evaluación de la plataforma de hardware han sido simplificados debido a que la organización decidió utilizar para los servicios de operación del DM, a saber un computador Servidor IBM RS/6000 Enterprise Server. Asimismo quedó simplificada la elección de la plataforma de software, ya que la estrategia de software de la empresa se basó totalmente en productos Oracle para garantizar la integridad, homogeneidad de la operación, tanto del ambiente transaccional como del DM.

2. Fase de Análisis

2.1 Análisis de requerimientos comerciales

2.1.1. Analizar la estrategia de la organización

Para el estudio y análisis de los procesos de negocio se obtuvo información referente a las funcionalidades, normas y procedimientos de las áreas que componen el DM Comercial. En el Anexo B, se presenta de forma gráfica los flujos de información de las distintas áreas del DM.

2.1.2 Entrevistas a los expertos de la materia

2.1.3 Determinar los requisitos funcionales

Las actividades 2.1.2 y 2.1.3 se realizaron conjuntamente, en ellas se revisaron y documentaron los requerimientos de información de la alta gerencia, en el área comercial, los cuales se fundamentan en el conjunto de preguntas que se espera

del modelo de datos operacional del sistema SGC, en función de los requerimientos de información y de los procesos de negocio relacionados, los cuales fueron explicados en las actividades 2.1.1, 2.1.2 y 2.1.3. La selección del sistema fuente óptimo estuvo basado en ciertos lineamientos; los cuales se encuentran especificados en el Apéndice E.

La revisión del código fuente utilizado para crear, actualizar y eliminar los datos fuentes, fue de gran utilidad para entender las relaciones e implicaciones de este sistema, además del apoyo presentado por el personal encargado de la aplicación SGC. Los puntos mencionados sentaron las bases para el diseño lógico de los modelos Entidad Relación (ER) de las áreas del DM Comercial; para su construcción se utilizó una herramienta tipo CASE (Computer Arded Software Engineering) llamada Visio 2000. En el Apéndice F se muestran los modelos de datos por cada área, y su respectiva documentación (Diccionario de Datos).

2.3.3 Identificar la técnica de modelado de datos

El modelado dimensional parte del principio de que el objetivo principal de un sistema de decisión es el análisis del rendimiento. Este rendimiento puede materializarse a través de un conjunto de indicadores, los cuales se analizarán a través de dimensiones (Franco, 1997).

Luego de estudiar las diversas técnicas para el modelado del DM (esquema estrella, esquema copo de nieve o esquema mixto), la técnica de modelación utilizada es la conocida como "*Enfoque de Esquema Estrella*", ya que este modelo sirve como una herramienta analítica para planear el DM y parte del principio de que los análisis de los indicadores son los que interesan al usuario;

éste es un punto fundamental porque caracteriza las dos mayores fuerzas de este tipo de modelo: la legibilidad y el rendimiento. Con la elección de este modelo es más fácil controlar y optimizar los accesos a las tablas de hechos porque se accede a ellas tras haber efectuado selecciones sobre las tablas de dimensiones, logrando tiempos de respuestas proporcionales al resultado esperado.

2.3.4 Transformar el modelo de datos lógico en el modelo de datos del DW

El DM Comercial fue desarrollado siguiendo el enfoque de "Data Warehouse Bus Architecture" propuesto por Ralph Kimball. Tomaremos la definición tal cual como fue explicitado en el Marco Teórico. Algunas premisas de esta arquitectura están mencionadas y explicadas en el Apéndice G.

2.3.5 Determinar el origen y destino de los datos

Esta actividad consiste en tener acceso a la metadata exacta y entendible, esto es posible luego del análisis de datos donde se tiene como resultado de la revisión de las características de la información existente y su utilización; es aquí donde se atiende a la factibilidad de su incorporación inmediata a la base de datos del DM, así como la evaluación de la calidad y la cantidad de la información disponible (fuentes de datos, compatibilidad, periocidad, soporte, etc).

En el Apéndice H se muestra un análisis exhaustivo sobre la factibilidad de información necesaria para poblar el DM y de esta manera lograr responder a las preguntas que permitan asegurar el control de la gestión comercial.

2.4 Diseño de la arquitectura técnica del DW

2.4.1 Evaluación y selección de la herramienta para usuarios finales

Esta actividad consistió en estudiar, evaluar y seleccionar una herramienta que se

adaptara a las necesidades de los usuarios, con la finalidad de conseguir aquella que sea capaz de ofrecer una interfaz sencilla al usuario, que manejase reportes y gráficos orientados a la toma de decisiones; tomando en cuenta que dicha selección estaría limitada por el producto de software Oracle (requerimiento de la compañía). El resultado de esta actividad se puede observar en el Apéndice I.

2.4.2 Seleccionar los componentes de la arquitectura técnica de hardware y software

2.4.3 Desarrollar un diseño de la arquitectura técnica

Las actividades 2.4.2 y 2.4.3 se realizaron en conjunto, se hizo la evaluación de la infraestructura informática y las recomendaciones para su actualización y/o adecuación, en función de las necesidades específicas de operación del DM. (Ver Apéndice J).

3. Fase de Diseño

3.1 Desarrollo de modelos físicos

En esta actividad el modelo lógico se convierte en el modelo físico de datos para las bases de datos de almacenamiento del DM y mercado de datos, para ello se utilizó la herramienta Visio 2000. Además, se definen los formatos, tipos y tamaños de los elementos del DM. En el diseño, cada área de estudio consta de una tabla de hechos y varias tablas de dimensiones; la tabla de hechos contiene las mediciones básicas de las variables del negocio y las tablas de dimensión contienen atributos del negocio que se emplearán como criterios de búsqueda SQL.

Los diseños físicos y sus diccionarios de datos se muestran en el Apéndice K.

3.2 Desarrollar la correspondencia de los modelos físicos de datos de las fuentes de datos con los modelos físicos del DW

Esta correspondencia ayuda a los procesos de extracción, refinamiento y reingeniería a efectuar sus funciones dentro del DM, los cuales se convierten en diseños para programas que realizarán las consultas requeridas por la alta gerencia, esto implica los procesos que son internos a las fuentes de datos y se relacionan con depuraciones o extracciones de información y procesos que conectan la fuentes de datos con el DM. En el Apéndice L, se puede observar la estructura lógica bajo la cual se desarrolló el DM Comercial, definida por Bill Inmon como *Corporate Information Factory* (CIF).

Esta actividad determinó la estructura que deben tener los procedimientos de extracción, transformación y carga (ETL).

En el Apéndice M se puede visualizar de forma detallada los diseños de los ETL de las tablas hechos y dimensiones por cada área que conforma el DM Comercial.

3.3 Diseño de la interfaz de la aplicación

Esta actividad consiste en el diseño del módulo de consulta, el cual se desarrollará exclusivamente con base y de acuerdo con la funcionalidad que provee el producto de software Discoverer de Oracle. El formato de dichos reportes vienen limitados por las funciones de interfaz predefinidas de esta herramienta. Los tipos de reportes estudiados fueron: el Table, el Crosstab, el Page Detail Table y el Page Detail Crosstab; su selección estuvo sujeta al tipo de consulta requerida por la gerencia (cálculos, filtros, etc), así como también por la calidad y la cantidad de pivoteos.

3.4 Diseñar programas, procesos y métodos para la extracción, transformación, depuración y carga

Desde el punto de vista del CIF, referido en la sección 3.2 de este capítulo, basado en el estándar Oracle como plataforma de software para soportar el componente arquitectural CIT, no requirió de una herramienta de extracción, transformación y carga (ETL); las mismas se orientan hacia la extracción de datos estructurados en ambientes operacionales basados en plataformas tecnológicas heterogéneas.

Las funciones de las herramientas de ETL se realizaron utilizando las propias facilidades presentes en los manejadores de base de datos Oracle, uno ya instalado en el ambiente operacional y el otro en donde reside el DM.

3.5 Diseño operacional de la base de datos

Se realizó el análisis de los volúmenes de datos que se manejan en el sistema SGC, para comenzar a establecer el tamaño de la base de datos del DM Comercial, en función del diseño lógico propuesto. Para el cálculo del crecimiento de la base de datos fue necesario definir muestras representativas agrupando datos similares, utilizando técnicas estadísticas. La explicación exhaustiva y detallada de estos cálculos se puede apreciar en el Apéndice N.

4. Fase de Construcción

4.1 Generación de la base de datos

Esta actividad consistió en la implementación del diseño físico de la base de datos, para tal fin se recurrió a la herramienta Enterprise Manager de Oracle, la cual se encargó de generar los scripts en el lenguaje PL/SQL de la creación de las tablas

dimensiones y hechos, así como también la definición de claves primarias, foráneas y las dependencias (constraints) entre ellas.

Luego de la implementación de la base de datos (llamada *"Prueba")*, fue necesario importar las tablas al producto de software Discoverer, esto se hizo gracias a la herramienta Oracle Discoverer Administration, con la finalidad de acceder a los datos que se encuentran en *Prueba* y posteriormente realizar las consultas con base a los requerimientos de información.

4.2 Prueba de construcción de los datos

Con base a la información recopilada de las distintas tablas del ambiente operacional (las cuales se encontraban en formato .txt), se procedió a poblar las tablas de hechos y dimensiones con datos de prueba, aplicando y cumpliendo con las reglas de integridad de una base de datos.

4.3 Construcción de procesos para el acceso a los datos

La herramienta de ayuda a la decisión debe proporcionar una interfaz de consulta sofisticada que oculte la complejidad del lenguaje SQL. Además, debe ser capaz de adaptar los datos al contexto del usuario. Las necesidades de gerentes, analistas empresariales y de muchos especialistas en información se solventan, por lo regular, en un ambiente de *consulta administrada*, en el cual las herramientas de procesamiento informático ocultan la complejidad del DM mediante una "gruesa capa semántica" de términos y reglas empresariales.

Por las razones expuestas es que se decidió trabajar con Oracle Discoverer, este producto de software permite crear en forma sencilla e intuitiva, consultas ad hoc, reportes, análisis y gráficos que pueden ser publicados en una Intranet bajo ambiente

Web. Para conocer las funcionalidades de Oracle Discoverer véase el Apéndice O.

4.4 Construcción de programas, procesos y métodos para la extracción, transformación y carga.

Esta actividad comenzó con la exportación de la estructura de la base de datos del sistema SGC y se aplicó el proceso "*Reverse Engineering*" para proceder a cargar las definiciones del repositorio de la base de datos y evaluar la disponibilidad de los elementos de datos del sistema SGC que cumplan con los requerimientos de información de la alta gerencia.

Se generaron programas cuya extensión es .sql para cada tabla de hechos y dimensiones; su estructura está formada por el encabezado y cuerpo de paquetes, en ellos se hacen las llamadas necesarias mediante procedimientos o funciones para llevar a cabo el proceso de extracción, transformación y carga.

En el Apéndice P se presenta, para cada tabla de hechos y dimensiones, un cuadro donde se especifica el nombre y la descripción de los procedimientos utilizados para su respectiva extracción, transformación y carga de datos.

La fase de extracción debe hacer frente a la homogeneidad de las fuentes de datos a consultar, esto pudo realizarse utilizando las facilidades presentes en los manejadores de base de datos Oracle, uno instalado en el ambiente operacional y otro en la ubicación del DM.

Para llevar a cabo la migración de datos se creó un link con la base de datos operacional, a la cual se le hacen las llamadas y consultas necesarias para poblar el DM. Este link se llama DM_PRUEBA y el código se muestra a continuación:

```
CREATE PUBLIC DATABASE LINK DM_PRUEBA
CONNECT TO DBACOM IDENTIFIED BY DBACOM
USING 'DBACOM';
```

Los programas de ETL se realizaron para cada tabla del DM en el lenguaje PL/SQL debido a la homogeneidad de las bases de datos, usando para ello la herramienta SQL*Plus de Oracle. En estos programas se especifica la instrucción de SQL correspondiente a la información que se desea extraer (utilizando el comando SELECT).

El proceso de transformación y depuración consistió en el formateo y la preparación de los datos fuente para su análisis, para ello se asociaron campos fuente con campos destino (mapping de datos); estas asociaciones fueron simples, como lo es el caso de no realizar ningún tipo de transformación al campo fuente; o un poco más compleja cuando aumenta la cardinalidad de la asociación (ejemplo: concatenación de datos).

Para las tablas de dimensiones, en general, la transformación estuvo restringida por la definición del código en función de las reglas del negocio. Ejemplo: la tabla *TIPO DE USO*, cuyo código debe comenzar con 'TF'.

Con el fin de asegurar la calidad de los datos, se desarrollaron funciones en PL/SQL, que reciben como parámetros de entrada el código que se desea insertar en la tabla de hechos y que está referenciado con una dimensión.

El proceso de carga de los datos consistió en transferir la información y poblar las tablas del DM, esto se logró gracias al comando INSERT de SQL.

Antes de ejecutar los programas ETL es necesario compilarlos para verificar que su código esté correcto y que haga la referencia a las tablas del SGC (mediante el link

DM_PRUEBA) y que los atributos o elementos tengan el nombre y formato tal cual como se encuentran en la base de datos del DM. Esto se hace utilizando el siguiente comando:

*START NOMBRE_ARCHIVO, en SQL*Plus*

Luego de que estos paquetes se compilaron satisfactoriamente, se procedieron a ejecutarlos con el uso del comando EXECUTE, logrando de esta forma la carga de los datos de las tablas pertenecientes al DM.

Si la carga de datos no se realizó correctamente, el programa envía un mensaje de "*ERROR en la actualización HECHOS*" al sistema.

Es importante mencionar que la frecuencia de actualización de las tablas de hechos se hará diariamente en las noches, ya que para entonces la actividad operativa es relativamente baja. Si por algún motivo o circunstancia no se pudiese realizar la actualización del DM tal como se tiene previsto, se creó para este fin una tabla por cada área de negocio donde guarda la última fecha de actualización, la cual es llamada en los procedimientos de ETL.

En el Apéndice Q se puede visualizar el ejemplo del código ETL aplicado a las tablas de hechos y dimensiones del DM.

4.5 Construcción de los reportes gerenciales

Para satisfacer las necesidades del rango completo de usuarios del procesamiento informático, el ambiente de éste debe manejar un amplio rango de tipos de consultas y reportes. Dicho ambiente demanda ciertas opciones de desempeño, desde la respuesta instantánea a las solicitudes ejecutivas y gerenciales hasta una espera razonable para las consultas ad-hoc de los analistas. En la arquitectura del DM y del ambiente de

procesamiento informático se consideró la tríada de actividades frecuentes (FAQ, FAR y FAD), se prepararon con anticipación consultas de acceso frecuente (FAQ – Frecuently Accessed Queries) y reportes solicitados con frecuencia (FAR – Frecuently Asked Reports) para los ejecutivos y gerentes, los cuales se depositaron en el DM para una recuperación instantánea de tablas, reportes matriciales, cuadros y gráficas. Los gerentes y analistas empresariales necesitan de las FAQ y FAR junto con los datos de acceso frecuente (FAD – Frecuently Accessed Data).

El requerimiento por parte de la gerencia de la empresa incluye: la generación de reportes y gráficos ad hoc, predefinidos y creados dinámicamente por los usuarios del DM Comercial. Estos reportes y gráficos permiten: 1) Controlar el mercado, su situación y evolución, 2) Controlar el ciclo comercial, incluyendo plazos y calidad de los procesos del ciclo, 3) Controlar la calidad del servicio, los plazos de resolución de actividades gestionadas por los clientes, y 4) Controlar la efectividad de la facturación, el cobro, la rentabilidad de los clientes, etc.

En Apéndice R, se explican detalladamente los reportes generados por cada área de trabajo.

4.6 Elaboración de Manuales

Esta actividad consistió en realizar el manual de usuario y el manual técnico (o de administración del sistema), los cuales contienen la información necesaria para una posible adecuación y/o actualización del sistema. Esto se hace para continuar con la operatividad funcional del DM (Ver Anexos C y D).

5. Fase de Pruebas

Una vez que la herramienta ha descubierto modelos y relaciones entre los datos, se debe permitir a los usuarios probar las hipótesis entrando nuevos juegos de datos y observando en qué medida corresponden a los modelos descubiertos, es por ello que es necesaria esta fase de pruebas al sistema.

5.1 Prueba del sistema DW (Validación de la información)

Esta actividad consistió en generar consultas SQL que permitan extraer directamente informaciones de la base de datos operacional y posteriormente verificar su correspondencia con los queries hechos a la base de datos del DM. La explicación de ello y el formato aplicado a esta fase se puede observar en el Apéndice S.

5.2 Carga completa y actualización del DW

Se procedió a limpiar o borrar los datos que contenían la base de datos del DM (la cual estaba poblada con datos de prueba), luego se cargó la información real ejecutando los programas ETL definidos y explicados en la sección 4.4 de este capítulo.

6. Implementación

La implementación de los mecanismos de gestión comercial requiere de la implementación de la base de datos, de los procesos de ETL del sistema operacional, la carga de los datos validados en el DM y de la herramienta de explotación de los datos almacenados en el DM.

CAPÍTULO V

RESULTADOS

Cada una de las iteraciones realizadas en el desarrollo e implementación del DM Comercial debe proporcionar resultados (productos) tangibles, los cuales se presentan a continuación:

- Se obtuvo toda la información pertinente para el desarrollo del Data Mart. Los requerimientos de información identificados proporcionaron las bases para realizar el diseño y la modelización del Data Mart.

- Se realizó el modelo de datos fuente, ya que el mismo constituiría el origen de la extracción de información necesaria para poblar el DM Comercial.

- El desarrollo del modelo de datos lógico del DM, correspondiente a los requerimientos o especificaciones de la gerencia. Estos modelos proveen una estructura para identificar, nombrar, describir y asociar los componentes de la base de datos, cuyo tipo de modelado es dimensional y la técnica utilizada se basó en el esquema estrella. También se generaron los diccionarios de datos para dar a conocer la metadata del DM al usuario final. El propósito es familiarizar al usuario con la visión analítica que propone el sistema y con los conceptos de negocio implementados.

- La selección de las herramientas de acceso a los datos estuvo en función de las necesidades de los usuarios finales, capacidades técnicas empresariales y las fuentes de datos existentes. Para ello se estudiaron dos herramientas de Oracle, Oracle Express y Oracle Discoverer, concluyendo que la primera opción es mejor, ya que la misma

cuenta con un motor de cálculo avanzado que permite procesamiento analítico en línea (OLAP). Sin embargo, la alta gerencia optó por trabajar con la herramienta Oracle Discoverer por ser ésta menos costosa.

- Definición de la arquitectura lógica bajo la cual se desarrolló el ecosistema de información. Esta se basó en los conceptos de *Corporate Information Factory,* introducidos por Bill Inmon y Claudia Imhoff.

- Implementación física del modelo lógico de datos del DM, es decir, la creación de la base de datos (la cual se llama "*Prueba*").

- Diseño de procesos de conversión que los datos origen, provenientes del sistema operacional SGC, sufrirán hasta ser incorporados al DM Comercial.

- Generación de programas para la extracción, transformación y carga de datos, cuya extensión es .sql, para cada tabla de hechos y dimensiones.

- Cargar o poblar la base de datos del DM utilizando los programas de extracción, limpieza, conversión e integración de los datos fuente; los cuales fueron desarrollados en el punto anterior.

- Diseño e implementación de actividades de pruebas al sistema, a través de grandes cargas de datos, para asegurar la credibilidad de los resultados arrojados por el DM Comercial.

- Paralelamente a la carga de datos obtenidos de los operacionales, se diseñaron los metadatos para las herramientas front-end. Este diseño se basó en el modelo lógico de datos creados, así como un análisis del tipo de informes que se ejecutan en el sistema, para facilitar al usuario la navegación y creación de consultas libres, así como la eficiencia de los reportes gerenciales (predefinidos o libres).

CONCLUSIONES

Una vez finalizado el Trabajo Especial de Grado, se tienen, como conclusiones generales referidas al diseño e implementación de un DW las siguientes:

- Es importante construir al principio del diseño del DM las funciones predefinidas y periódicas del negocio, pero lo fundamental es permitir a largo plazo a los diferentes tipos de usuarios acceder, mediante la herramienta adoptada, a todos los datos del DM (datos agregado y datos detallados); ya que el valor añadido viene por una parte de los datos detallados y por otra, de la capacidad del usuario para utilizar eficaz y eficientemente las herramientas adaptadas a sus necesidades.

- Es necesario conducir el cambio según la experiencia de la empresa en el ámbito de la toma de decisiones, y de las tecnologías y herramientas utilizadas, por ello es de gran utilidad implementar un plan de formación. También es muy importante que los miembros del equipo sean favorables al cambio, debido a las especificidades de los proyectos de decisión.

- Es importante que la herramienta de apoyo a la toma de decisiones sea capaz de abrirse, de transmitir datos y ofrecer funcionalidades. Lo ideal es entonces poder difundir los resultados no en una forma fija, sino en la forma del análisis que ha permitido obtenerlos, teniendo el destinatario la posibilidad de recuperar ciertos puntos.

- En la fase de recolección de requerimientos de un DM hay que ser bien selectivos con los usuarios finales, hay que tener cuidado en no elegir aquéllos que estén orientados a la tecnología para que participen en la definición de motores

comerciales. Un DM no es un problema técnico, sino la integración de un conjunto de iniciativas "orientadas a temas" en un entorno de tecnología. La palabra tema utilizada sistemáticamente es ilustrativa de la necesaria orientación al negocio.

- Los problemas no terminan una vez finalizado e implementado el DM, éste es el sistema de apoyo a la toma de decisiones de la empresa y su ciclo de vida está relacionado con las evoluciones de la organización. La evolución regular del número de usuarios y del tipo de accesos realizados necesita de constantes actualizaciones y optimizaciones.

Para garantizar el éxito del Proyecto de desarrollo del DM Comercial es necesario tener en cuenta una serie de consideraciones:

- La implantación de un DM y su operación, impondrá una nueva visión de los procesos de calidad de datos tanto desde el punto de vista técnico como del negocio.

- En la medida en que se avance en el desarrollo del DM será necesario contar con nuevas herramientas de software para explotar la mina de datos disponible de las operaciones del negocio de los últimos tres años.

- La operación de un DM obligará a la empresa a mantener una plataforma tecnológica especializada en Inteligencia de Negocios, diferente al ambiente actual de operaciones.

- En la medida en que se avance en el uso de estas herramientas será necesario darle mayor importancia al área de tecnología de información, para que esté más enfocada hacia el soporte de las funciones centrales del negocio.

- La falta de experiencia del personal de la organización, tanto en el área técnica como en el área gerencial, en el uso de las tecnologías de Data Warehouse, obliga a complementar el desarrollo del Proyecto con una serie de cursos de orientación y adiestramiento al personal involucrado, con el fin de garantizar el proceso de cambio hacia un ambiente más dinámico y pro-activo en el uso de la información para el análisis del negocio y la toma de decisiones.

- El diseño del Sistema SGC no se apega a los criterios más elementales del modelo relacional, lo que hace difícil los procesos de selección de las tablas, esto agrega gran complejidad al diseño e implementación de los procesos de ETL (Extracción, Transformación y Carga) del DM.

RECOMENDACIONES

En función de las conclusiones obtenidas, se recomiendan las siguientes actividades con el fin obtener un mayor beneficio de la aplicación del DM Comercial:

- En algún momento, será necesario implementar un ambiente integrado de comunicaciones, basado en tecnología Web, que soporte la infraestructura de la Arquitectura del CIF de la empresa y facilite las comunicaciones internas y externas. El Portal de la organización conferirá acceso ubicuo, puntual, extenso, completo, seguro y personalizado a la información de la empresa. Especialmente, permitirá el manejo de la información no estructurada y la integración de los procesos de E-Business, y de las relaciones del tipo: Bussines-to-Business y Business -to-Consumer.

- Las herramientas de Minería de datos son una opción en la medida en que crezca el volumen de datos del DW, especialmente con las próximas iteraciones o etapas más avanzadas de su ciclo de vida.

- Revisar y mejorar los procesos y el uso del sistema SGC para aminorar sus problemas de diseño y garantizar una operación alejada de malas prácticas que deterioran la calidad de los datos.

- Diseñar y dictar una serie de cursos gerenciales y técnicos orientados a mejorar el perfil del personal en el uso de herramientas de análisis de negocio.

- Prepararse para la adquisición de nuevas herramientas de software orientadas hacia el análisis del negocio, como son: OLAP, minería de Datos, Portal, etc.

- Para mejorar en forma significativa el tiempo de respuesta del DM, se debería

utilizar una tecnología de indización de mapa de bits, ya que reduce en gran medida la cantidad de operaciones de lectura a los datos y es capaz de acelerar la ejecución de consultas complejas. Además de permitir que más usuarios tengan acceso simultáneo al DM, la indización de mapa de bits también facilita la capacidad para que los usuarios propongan una serie de consultas a fin de analizar datos.

- Con el crecimiento acelerado del volumen del DM es necesario segmentar (particionar) las tablas en varios discos. La segmentación también debe proporcionar la capacidad de ejecutar respaldos y restauraciones a nivel de segmento.

- Realizar actualizaciones, cuando sea requerido, de los reportes gerenciales, para que los mismos se ajusten y se adapten a las necesidades y requerimientos del negocio, los cuales son cambiantes con el transcurrir del tiempo.

- Es conveniente diseñar políticas de respaldo y recuperación para resguardar los componentes del sistema de apoyo a la toma de decisiones (base de datos del DM, reportes gerenciales, programas de ETL), siguiendo los estándares de seguridad y respaldo de los sistemas de la empresa. Al mismo tiempo es necesario desarrollar un plan de integración y mantenimiento del sistema del DM, con la finalidad de que este continúe con su funcionalidad operativa.

BIBLIOGRAFÍA

Anahory, S., Murray, D. (1997). **Data Warehousing in the Real World: A Practical Guide for Building Decision Support Systems.** Published July 1997 by Addison Wesley Professional Copyright 1997.

Armostrong, M. – Smith y Armstrong, M. – Smith (2001). **Manual de Oracle Discoverer.** Edición Oracle Press Oficial Osborne. España: MacGraw Hill

BONCZEK, R.; HOLPSAPPLE, C. y WHINSTON, A.: "**Future Decisions for Developing Support Systems**". Decision Sciences

Business Intelligence. Disponible: http://www.bitam.com/DefinicionBI.htm. Copyright © 2001-2002 BITAM. (Consulta: 15 de junio de 2002).

Cerezo, C. (2000). "**Data Warehouse: Poderosa herramienta de soporte de decisiones**". Tomado de la base de datos de Infolatina a través de la Intranet del ITESM-CEM: http://biblioteca.itesm.mx/cgi-bin/salta? Disponible: http://zeus.infolatina.com.mx/universidad/index.cgi (Consulta: 18 de julio de 2002)

Dale, T. (2002). **Estrategias para la implementación de un Data Warehouse.** Disponible: http://www.infonegocio.com.pe/areas/com_electronico/especiales/20020617_03_01_02.shtml

Data Mart / Data Warehouse. Disponible: http://www.kpp.cl/datamart%20datawarehouse.htm . (Consulta: 2002, septiembre 14)

Data Warehouse e Inteligencia de Negocios. Disponible: http://www.ideasa.net. Consulta: 2002, septiembre 25)

Data Warehouse. Disponible: http://www.dlsi.ua.es/~jtrujillo/doctoradotema1bnpro.pdf. (Consulta: 2002, mayo 13)

Data Warehousing. Disponible: http://www.sqlmax.com/dataw1.asp. (Consulta: 17 de agosto de 2002)

Díaz Caprino, J. L. **Data Warehouse: funcionalidad y servicios.** http://www.tecnomarkets.com/servicios/tecnonews/research62.htm

Dresner, H. (2001). **Business Intelligence in 2002: A Coming of Age.** Disponible: http://www4.gartner.com/DisplayDocument?doc_cd=103282. (Consulta: 05 de mayo de 2002)

Dyche, J. (2001). **E-Data, Transformando datos en información con Data Warehousing**. (1ra. Edición). Prentice Hall.

Elizondo G., M. (2002). **Sistemas inteligentes y de soporte a la decisión**. [Artículo en Línea]. Disponible: http://www.netmedia.info/netmedia/articulos.php?id_sec=32&id_art=3678. Consulta: 01 de julio de 2002)

Ferrando Girón, S. J., Carreira Perpiñán, I. J. (2001). **Soluciones Data Warehouse bajo el Nuevo modelo de organización e-business: la necesidad de integrar la información.** Disponible: www.tid.es/presencia/publicaciones/ comsid/esp/22/08.pdf. Septiembre 2001. (Consulta: 12 de julio de 2002)

Giovinazzo, W. (2000). **Object – Oriented Data Warehouse Design. Building a Star Schema**. New Jersey (USA): Prentice Hall.

Gismondi, N. (2000). **Data Warehousing.** Disponible: http://www.cp.com.uy/45/data45.htm. (Consulta: 14 de mayo de 2002)

Gray, P., Watson, H. (1997). **Decision Support in the Data Warehouse (Data Warehousing Institute Series from Prentice Hall Ptr).** Paperback / Published 1997

Harjinder S. Gill, Prakash C. Rao (1996). **Data Warehousing. La integración de información para la mejor toma de decisiones.** México: Prentice Hall.

Imhoff, C., et al. (1997).**Corporate Information Factory.**

Implementando un Data Warehouse. Disponible: http://www.inf.udec.cl/revista/edicion5/cwolff.htm. (Consulta: 2002, abril 28)

Implementando un DataWarehouse. Disponible: http://mmhernandez.galeon.com/articulo1.htm. (Consulta: 2002, junio 18)

Inmon, W. (1999). **A Data Warehouse Development Methodology.** Disponible: http://www.billinmon.com/library/methods/methods_frame.asp. (Consulta: 2002, agosto 28)

Iterations™ -The Data Warehouse Development Methodology. Disponible: www.ascentialsoftware.com. (Consulta: 2002, abril 24)

Jean – Michel Franco y EDS – Institut Prométhéus. (1997). **El Data Warehouse – El Data Mining**. (Primera Edición). España : Ediciones Gestión 2000

Jucan, G. (2002). **Designing a Corporate Information Factory using the Zachman Architecture Framework.** Disponible: http://www.inconcept.com/JCM/January2002/jucan.html. (Consulta: 17 de junio de 2002)

Kimball, R. (1996). **The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses**.John Wiley & SONS, INC.

Kimball, R., Ross, M. (2002). **The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling** (Second Edition)

Kimball, R., Ross, M., Reeves, L., Warren Thomthwaite John Wiley & SONS, INC. (1998).**The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses.**

Los componentes del Data Warehouse. Disponible:http://sie.efpol.ua.es/websie/docdocencia/iti/tema1_iti.pdf. (Consulta: 16 de agosto de 2002)

Molina R., G. **Sistemas de Data Warehousing.** Disponible: http://www.logicnet.com.mx/gamolina/Data%20Warehousing.html. (Consulta: 2002, agosto 18)

Oracle9i. Data Warehousing Guide, Release 2 (9.2). Copyright © 1996, Oracle Corporation All rights reserved. Paul Lane

Peribán Villa, G. (2001). **DSS, Impacto en el Proceso de Toma de Decisiones.** http://www.claveempresarial.com/soluciones/notas/nota010326c.shtml.(Consulta: 06 / Marzo / 2001)

Procesos en el Diseño de un Data Warehouse Disponible: http://inf.umayor.cl/inf/DATAWAREHOUSE.htm. (Consulta: 2002, junio 16)

SENECA. Disponible: http://www.seneca.com.ve. (Consulta: 04 de abril de 2002)

SISTEMA DE SOPORTE A LA TOMA DE DECISIONES (DSS). Disponible: http://www.geocities.com/SiliconValley/Pines/7894/introduccion/dss.html

Treviño G., E. (2002). **Sistema de soporte a la decisión y sistemas inteligentes**. Disponible: http://.netmedia.info/netmedia/artículos.php?idsec=32&id art=3680 (Consulta: 01 de julio de 2002-10-12

Vega Romero, A. (2002). **Data Warehouse.** [Artículo en línea]. Disponible: http://www.netmedia.info/netmedia/articulos.php?id_sec=32&id_art=3677 (Consulta: 04 Marzo de 2002)

W. H. Inmon, R. H. Terdeman, Claudia Imhoff .(2000).Exploration Warehousing: Turning Business Information into Busines Opportunity

W.H. INMON: **Building the Data Warehouse**. Wiley Computer Publishing, 1996. 2° Edición

Wolf, C. (1999). **Modelamiento Multidimensional.** [Libro en línea]. Disponible: http://www.inf.udec.cl/revista/edicion4/cwolff.htm. (Consulta: 04 de abril de 2002)

GLOSARIO

GLOSARIO

Análisis: operación consistente en estudiar y descomponer los datos a fin de extraer de ellos los elementos esenciales y obtener un esquema de conjunto.

Análisis multidimensional de datos: análisis simultáneo de múltiples dimensiones de datos.

Arquitectura Cliente/Servidor: sistema tecnológico basado en la distribución de funciones. El servidor realiza funciones compartidas, gestión de comunicaciones, facilitar servicios a las bases de datos, etc. El cliente realiza funciones individuales, tales como facilitar interfaces, navegación por pantallas, funciones de ayuda, etc.

Arquitectura de referencia (Data Warehouse): clasificación general de los componentes arquitectónicos del Data Warehouse.

Atributo: elemento de información, que describe una característica de una entidad.

Backup/Restore: el acto de copiar archivos y bases de datos para protegerlos en caso de fallas o catástrofes y regresarlos a su estado normal en una fecha posterior.

Base de datos (BD): conjunto de datos no redundantes, almacenados en un soporte informático, organizado de forma independiente de su utilización y accesible simultáneamente por distintos usuarios y aplicaciones. La diferencia de una BD respecto a otro sistema de almacenamiento de datos es que éstos se almacenan en la BD de forma que cumplen tres requisitos básicos: no redundancia, independencia y concurrencia.

Base de datos histórica: base de datos que provee una perspectiva histórica sobre los datos.

Base de datos multidimensional: una base de datos diseñada alrededor de un conjunto de dimensiones; se usa el análisis multidimensional.

Base de datos operacional: una base de datos que apoya sistemas de software que están soportando normalmente a las operaciones empresariales de una organización. También llamada base de datos OLTP.

Base de datos relacional: sistema de almacenamiento de datos basado en un conjunto de tablas unidas mediante relaciones.

Bloque de Data Warehouse: el componente de la arquitectura del Data Warehouse que representa actividades relacionadas directamente con operaciones del Data Warehouse.

Bloque de fuentes de datos: el componente de la arquitectura de referencia del Data Warehouse que representa las actividades que tratan con las bases de datos operacionales y las fuentes externas que aportan datos para cargarlos en el Data Warehouse.

Capa de administración de datos: el componente de la arquitectura de referencia del Data Warehouse que maneja las opciones de administración de datos.

Capa de administración de metadatos: el componente del Data Warehouse relacionado con los aspectos de definición de los datos que se capturan y guardan en el Data Warehouse. Todas las herramientas y procesos que deben interactuar con el Data

Warehouse usan los aspectos de definición.

Capa de infraestructura: componente del Data Warehouse relacionado con las plataformas y el soporte computacional y el hardware y software de ambiente adquiridos en la industria que se requieren para apoyar el Data Warehouse.

Capa de transporte: el componente del Data Warehouse relacionado con el transporte de datos entre componentes de la solución de Data Warehouse.

CASE: siglas de Computer Arded Software Engineering (ingeniería de software apoyada con computadora). Una metodología de desarrollo de software soportada por herramientas basadas en computación para el apoyo a tareas de análisis, diseño y desarrollo.

Consulta: una petición formal y claramente especificada de información del Data Warehouse o del mercado de datos planteada por un usuario o una herramienta operada por un usuario.

Consulta ad-hoc: pregunta que no está previamente determinada y que es necesario establecer. Una pregunta que consiste en una instrucción SQL construida dinámicamente.

Data Mart: una implementación de Data Warehouse con un ámbito de datos y funciones de Data Warehouse más pequeño y restringido, que sirve a un departamento único o una parte de la organización. Una organización generalmente tiene varios Data Marts.

Data Warehouse: una implementación de una base de datos informativa usada para

almacenar datos desde una base de datos operacional de registro. Es típicamente una base de datos que permite a los usuarios explotar en un almacén extenso de datos operacionales y responder a las tendencias de las organizaciones y facilitar pronósticos y esfuerzos planificados.

Data Warehousing: proceso de implementación de un proyecto de Data Warehouse.

Dato: unidad básica de información.

Datos Derivados: datos producidos por cálculos o procesos que aplica el Data Warehouse a los datos provenientes de las fuentes. Los datos derivados sólo residen en el Data Warehouse y éste los usa para precalcular varios valores que se requieren para responder a frecuentes solicitudes. La alternativa de guardar datos derivados es calcular el valor justo cuando se ejecuta la solicitud, lo cual provoca una disminución significativa en el desempeño.

Diccionario de Datos: descripción lógica de los datos para el usuario. Reúne la información sobre los datos almacenados en la BD (descripciones, significado, estructuras, consideraciones de seguridad, edición y uso de las aplicaciones, etc.).

Dimensión: entidad independiente dentro del modelo multidimensional de una organización, que sirve como llave de búsqueda (actuando como índice), o como mecanismo de selección de datos.

DSS (Sistema de Soporte de Decisiones): sistema de aplicaciones automatizadas que asiste a la organización en la toma de decisiones mediante un análisis estratégico de la información histórica.

ETT (Extracción, Transformación y Transporte de datos): pasos por los que atraviesan los datos para ir desde el sistema OLTP (o la fuente de datos utilizada) a la bodega dimensional. Extracción, se refiere al mecanismo por medio del cual los datos son leídos desde su fuente original. Transformación (también conocida como limpieza) es la etapa por la que puede atravesar una base de datos para estandarizar los datos de las distintas fuentes, normalizando y fijando una estructura para los datos. Finalmente está el Transporte, que consiste básicamente en llevar los datos leídos y estandarizados a la bodega dimensional (puede ser remota o localmente). Generalmente, para un Data Mart no es necesario atravesar por todos estos pasos, pues al ser información localizada, sus datos suelen estar naturalmente estandarizados (hay una sola fuente).

Extracción de datos: modalidad de descubrimiento del análisis de datos, o analizar datos de detalle para revelar relaciones, patrones y asociaciones insospechados o desconocidos.

Extracción: la actividad relacionada con transferir datos de bases de datos operacionales (fuentes de datos) al Data Warehouse.

Front-end: formato de pantalla. Dícese de la presentación que se le hace al usuario al usuario de la pantalla de trabajo.

FAD: siglas de Frequently Accessed Data (datos frecuentemente accesados).

FAQ: siglas de Frequently Accessed Query (consulta frecuente accesada).

FAR: siglas de Frequently Asked Reports (reportes frecuentemente accesados).

Fuentes de datos de producción: bases de datos operacionales que actúan como

fuentes de datos para el Data Warehouse. Los datos que aportan son en línea y se guardan en la base de datos.

Granularidad: término que se usa en los Data Warehouse para expresar el nivel de detalle. A más alto nivel de granularidad más bajo nivel de detalle (más alto nivel de abstracción).

Hecho: dato numérico que sirve de base para la definición de los indicadores en un modelo multidimensional.

Implementación: corresponde a la programación real del software de aplicación y puede incluir revisiones del diseño, creación de prototipos y desarrollo de consultas.

Indizado: una técnica de uso frecuente para optimizar el desempeño de la base de datos mejorando el método de acceso para buscar y recobrar registros de la base de datos.

Información histórica: información que se recopiló en el pasado y que por lo general se archivó y se retiró de las bases de datos operacionales debido a los volúmenes de datos.

Integridad: condición de seguridad que garantiza que la información es modificada, incluyendo su creación y borrado, sólo por el personal autorizado.

Inteligencia de negocio: la utilización de información proveniente de varios sistemas operacionales, Data Marts y/o Data Warehouse con el objeto de mejorar los procesos de negocio, optimizar el servicio al cliente, acelerar los procesos de ventas e incrementar la rentabilidad de la empresa.

Jerarquía: forma de agrupar todos o sólo algunos de los elementos de una dimensión con relaciones padre-hijo. Casi siempre, pero no obligatoriamente, implican que el padre se calcula como la suma de sus hijos. Una dimensión puede tener cero, una o varias jerarquías.

Limpieza y refinamiento de datos: la actividad de eliminar errores en las fuentes de datos antes de intentar cargarlas en el Data Warehouse o en el mercado de datos (Data Mart).

Manejador de base de datos: un DBMS es una colección de numerosas rutinas de software interrelacionadas, cada una de las cuales es responsable de alguna tarea específica tales como: almacenar físicamente, garantizar consistencia, garantizar integridad, atomicidad transaccional, y manejar vistas a la información.

Metadata: información que describe un dato. En un contexto de Data Warehouse, cualifica un dato precisando por ejemplo su semántica, las reglas de gestión asociadas, su fuente, su formato etc.

Método de desarrollo en espiral: un método iterativo de desarrollo de software en el que la funcionalidad del software se produce en etapas y las mejoras se identifican en realidad desplegando sucesivamente el software con funcionalidad creciente pero bajo control estricto.

Métrica: indicador. Entre los indicadores pertinentes, algunos distinguen los hechos, almacenados físicamente en la base de decisión, de las métricas, derivadas de estos hechos.

Modelado de datos: actividad de representar las categorías de datos y relaciones entre ellas como una abstracción en forma de diagrama.

Modelo lógico de datos: una representación formal y abstracta de las categorías de datos y las relaciones entre ellos, en forma de diagrama. La representación es independiente del modo en que las clases y sus relaciones se implementan físicamente en la base de datos.

Modelo dimensional (o multidimensional): técnica de modelado consistente en modelar una base de decisión a partir de la identificación de los hechos a analizar y las dimensiones de análisis asociadas.

Modelo estrella: como el nombre sugiere, el esquema de estrella es un paradigma de modelado que tiene en medio un objeto único conectado a varios objetos en forma radial: El esquema estrella refleja la visión del usuario en una consulta empresarial: hechos tales como ventas, compensaciones, pagos o facturas calificadas por una o más dimensiones (por mes, producto, región geográfica). El objeto en el centro de la estrella se llama tabla de hechos y lo que se conectan desde la periferia se llaman tablas dimensión.

Muestra: conjunto de datos sacado, aleatoriamente o no, del Data Warehouse que permite efectuar pruebas e investigaciones sobre volúmenes reducidos.

OLAP (On Line Analytical Processing): caracteriza la arquitectura necesaria para implementar un sistema de información de decisión. Se opone a OLTP. El término OLAP designa a menudo las herramientas de análisis basadas en base de datos multidimensionales.

OLTP (On Line Transactionnel Processing): tipo de entorno de tratamiento de la información en el que debe darse una respuesta en un tiempo aceptable y consistente.

Pivotear: a partir de un contexto matricial de filas y columnas, mover un ítem desde un eje fila a un eje columna o viceversa. Lo anterior permite visualizar dinámicamente las relaciones entre los datos.

Procesamiento informático: modo de verificación del análisis de datos: para acceder y analizar los datos en forma bi o tridimensional y luego presentar los resultados en forma de reportes o resultados de consulta (lo que también se conoce como consulta o [query] y reporte [reporting]).

Query: Ver *Consulta.*

Redundancia: repetición de los mismos datos en varios lugares.

Reporte: documento que se produce en respuesta a una consulta.

Repositorio de datos: una base de datos diseñada para almacenar metadatos y acceso de usuarios y administradores.

Segmentar: esta es la actividad de dividir un dato en dos o más elementos pequeños. Se usa para separar una tabla única de la fuente en dos o más tablas dentro del Data Warehouse. También se aplica para dividir las filas de una tabla en dos o más tablas con la misma estructura, pero de manera que cada una contenga un subconjunto de filas.

SGBD: acrónimo de sistema de Gestión de Bases de Datos. Dícese de la serie de aplicaciones que nos permiten la creación, la búsqueda y la organización de los datos

en una base de datos.

SQL (Structure Query Language): lenguaje Estructurado de Consulta, usado con casi todas las herramientas de programación para gestionar bases de datos y extraer información de ellas.

Sumarización: actividad de incremento de la granularidad de la información en una base de datos. La sumarización reduce el nivel de detalle, y es muy útil para presentar los datos para apoyar al proceso de Toma de Decisiones.

Tabla de dimensión: un componente del esquema de estrella (ver también *Modelo Estrella*).

Tabla de hechos o fact: un componente del esquema estrella (ver también *Modelo Estrella*).

Verificar: comprobaciones de la calidad de los datos definidas por reglas empresariales, el modelo de Data Warehouse, etc; la congruencia y la integridad dominan la verificación.