



# UNIVERSIDAD CATÓLICA ANDRÉS BELLO FACULTAD DE INGENIERÍA ESCUELA DE INGENIERÍA INFORMÁTICA

# DISEÑO E IMPLANTACION DE UN DATA MART PARA EL AREA DE FACTURACION Y VENTAS DE LA EMPRESA MAKRO COMERCIALIZADORA, S.A. SEDE LA URBINA

Este Jurado; una vez realiza	ado el examen del presente trabajo ha evaluado su
contenido con el resultado: _	DIECISIETE PUNTOS (17)

JURADO EXAMINADOR

Nombre: Robato Cash Nombre: Licia Caldos Nombre: Susaul Garcis

REALIZADO POR:

Br. Régulo E. Guédez Flores

Br. Yaranais Zambrano Loyo

TUTOR:

Ing. Roberto Castañeda

FECHA:

Caracas, 29 de Noviembre de 2005



#### Universidad Católica Andrés Bello

Facultad de Ingeniería

Escuela de Ingeniería Informática

Diseño e implantación de un Data Mart para el área de facturación y ventas de la empresa Makro Comercializadora, S.A. Sede La Urbina.

Autores: Régulo Guédez y Yaranaís Zambrano

Tutor: Roberto Castañeda

Fecha: noviembre de 2005

#### RESUMEN

El presente trabajo especial de grado tiene como objetivo general diseñar, construir e implantar un Data Mart y una aplicación Web de Intranet para la construcción de reportes en el área de facturación y ventas de la empresa Makro Comercializadora, S.A. sede La Urbina. Entendiendo como Data Mart a una base de datos que almacena data histórica obtenida de los sistemas operacionales a través de procesos de extracción, transporte, transformación y carga para luego ofrecer información relevante para la toma de decisiones. Es la aplicación de Data Warehousing a una escala más pequeña. Para conseguir esto se siguieron tres pasos principales propuestos por la metodología Three-Stage Process. En primer lugar la etapa de planificación que se resume en la entrega de un documento de la definición del proyecto y un plan detallado del mismo basado en entrevistas y consultas teóricas preliminares. En la etapa de desarrollo se entrega un prototipo y un cronograma de la etapa de implantación, este prototipo está acompañado de los procesos que permiten obtener, transformar y cargar los datos correctos. Los resultados de la etapa de implantación son un Data Mart con data cargada y su documentación. El padre de esta herramienta de IN es Bill Imon, para él, aplicar Data Warehouse no es sólo obtener un producto, es un conjunto de componentes de hardware y software. Sin embargo, Ralph Kimball se enfoca principalmente en el diseño de la base de datos que almacenará la información para la toma de decisiones. En general, la aplicación de este tipo de herramientas permite recolectar y organizar la data producida por los procesos del día a día de la empresa y convertirla en información útil para evaluar el funcionamiento del negocio. Por la gran cantidad de data a manejar en producción debe hacerse su implementación en un servidor de la familia Sun Microsystem y apoyado por la herramienta MicroStrategy 8 para toma de decisiones.

**Descriptores**: Inteligencia de negocios, Data Warehouse, Data Mart, procesos de ETL (Extracción Transformación y Carga).

# TABLA DE CONTENIDO

DEDICATORIA	i
AGRADECIMIENTOS	ii
RESUMEN	
TABLA DE CONTENIDO	V
TABLA DE CUADROS Y FIGURAS	viii
INTRODUCCIÓN	1
CAPÍTULO I	4
I.1 EL PROBLEMA	4
Planteamiento del Problema	
I.2.1 Objetivo General I.2.2 Objetivos Específicos I.3 ALCANCES Y LIMITACIONES	6
I.3.1 Alcances. I.3.2 Limitaciones. I.4 JUSTIFICACIÓN.	7
CAPÍTULO II	10
II.1 MARCO REFERENCIAL	10
II.1.1 Inteligencia de Negocios II.1.2 Data Warehouse II.1.3 Data Mart II.1.4 Extract, Transform and Load (ETL) CAPÍTULO III	14 21
III.1 METODOLOGÍA	
III.1.1 Three-Stage Process  III.1.1.1 Etapa 1: Planificación  III.1.1.2 Etapa 2: Desarrollo  III.1.1.3 Etapa 3: Implementación  CAPÍTULO IV	29 30 32
IV.1 DESARROLLO	
IV.1.1 Etapa de Planificación	38

1 T T T T T T T T T T T T T T T T T T T	60
V.1 RESULTADOS	
CAPÍTULO VI	
VI.1 CONCLUSIONES	
VI.2 RECOMENDACIONES	
REFERENCIAS BIBLIOGRÁFICAS	
APÉNDICE A	
PLAN DE PROYECTO	
APÉNDICE B	79
DICCIONARIOS DE DATOS PARA HEAD OFFICE (HO) Y STO	ORE (ST) 79
APÉNDICE C	115
DISPONIBILIDAD DE DATOS	115
APÉNDICE D	128
DISEÑO DE LOS PROCESOS DE ETL	128
APÉNDICE E	144
METADATA	144
APÉNDICE F	187
CRECIMIENTO DEL DATA MART	187
APÉNDICE G	191
ELECTRÓNIC INVOICE ARCHIVE (EIA)	191
APÉNDICE H	209
SISTEMAS DE SOPORTE A LA TOMA DE DECISIONES	209
APÉNDICE I	219
ESCENARIO DE SERVIDORES PROPUESTO	219
ANEXO A	222
DESCRIPCIÓN DE LA EMPRESA MAKRO COMERCIALIZADO	
<u></u>	222
ANEXO B	
MOTODOLOGÍA EMPLEADA	
Stage 1: Planning	
Stage 2: Development	225

# TABLA DE CUADROS Y FIGURAS

# CUADROS

Cuadro	Página
Cuadro 1. Descripción de equipo de trabajo.	72
Cuadro 1. Descripción de servidor de trabajo.	73
Cuadro 3. Diccionario de datos para HO	75
Cuadro 4. Diccionario de datos para ST	95
Cuadro 5. Disponibilidad de datos para el Data	113
Mart	
Cuadro 6. Tablas de extracción de data	124
Cuadro 7. Carácter por nombre de tabla	137
Cuadro 8. Carácter por tipo de SERIAL ID	137
Cuadro 9. Diccionario de datos para el Data	156
Mart de facturación y ventas	
Cuadro 10. Contenido actual disponible para el	189
Data Mart	
Cuadro 11. Proyección de crecimiento a un año	191
Cuadro 12. Proyección de crecimiento a dos	192
años	
Cuadro 13. Métodos pertenecientes al modelo	195
del EIA	

# FIGURAS

Figura 1. Ambiente de Data Warehouse.	13
Figura 2. Tiempo Variante en Data Warehouse.	15
Figura 3. Data no volátil en Data Warehouse	16
Figura 4. Esquema estrella.	18
Figura 5. Esquema copo de nieve	19
Figura 6. Esquema constelación	20
Figura 7. Interrelación del Data Mart con el ambiente empresarial.	20
Figura 8. Data Marts por departamento.	
Figura 9. Relación de tablas de hechos y tablas de dimensiones	37
Figura 10. Arquitectura del Data Mart	42
Figura 11. Esquema Copo de Nieve para la tabla FA_INVOICE_LINE	46
Figura 12. Esquema Copo de Nieve para la tabla FA INVOICE MULTISAVE.	47
Figura 13. Esquema Copo de Nieve para la tabla FA_INVOICE_PAYMENT	48
Figura 14. Esquema Copo de Nieve para la tabla FA_INVOICE_VAT	
Figura 15. Ejemplo de archivo log	
Figura 16. Ejemplo de archivo .dsc.	59
Figura 17. Ambiente general de desarrollo.	60
Figura 18. Calendario de desarrollo	75
Figura 19. Modelo entidad - relación HO	96
Figura 20. Modelo entidad - relación ST.	
Figura 21. Estructura general de archivos para SQL*Loader	133
Figura 22. Dimensión Tiempo	144
Figura 23. Jerarquías de la dimensión tiempo	145
Figura 24. Dimensión Comprador	145
Figura 25. Jerarquías de la dimensión Comprador.	146
Figura 26. Dimensión de Artículo.	
Figura 27. Jerarquías de la dimensión Artículo.	147
Figura 28. Jerarquías de la dimensión Artículo.	147
Figura 29. Dimensión Proveedor.	
Figura 30. Jerarquías de la dimensión Proveedor.	148
Figura 31. Dimensión Tienda.	
Figura 32. Jerarquías de la dimensión Tienda.	
Figura 33. Dimensión Makro Mail.	
Figura 34. Jerarquías de la dimensión Makro Mail.	
Figura 35. Jerarquías de la dimensión Makro Mail	
Figura 36. Dimensión Cliente	
Figura 37. Jerarquías de la dimensión Cliente.	152

Figura 38. Jerarquías de la dimensión Cliente.	152
Figura 39. Ambiente general de Desarrollo.	153
Figura 40. Mapa de navegación (Parte 1 de 2)	201
Figura 41. Mapa de navegación (Parte 2 de 2)	201
Figura 42. Pantalla inicial	202
Figura 43. Opciones del menú de consulta de facturas	203
Figura 44. Opciones del menú de consulta Estadísticas.	203
Figura 45. Consulta de factura por No. De factura	
Figura 46. Consulta de factura por Artículo.	204
Figura 47. Consulta de factura por Cliente	204
Figura 48. Consulta de Factura por No. Fiscal no vacío	
Figura 49. Resultado de cualquiera de las consultas de búsqueda de facturas	205
Figura 50. Ejemplo de detalle de factura generado por el EIA.	206
Figura 51. Detalle de cliente.	207
Figura 52. Consulta de TOP Artículos por Grupo de Clientes	207
Figura 53. Artículos más vendidos a un grupo de clientes	208
Figura 54. Consulta de TOP Clientes por Artículo.	
Figura 55. Escenario No. 1.	32.02
Figura 56. Escenario No. II.	220

# INTRODUCCIÓN

"La Inteligencia de Negocios o Business Intelligence (BI) se puede definir como el proceso de analizar los bienes o datos acumulados en la empresa y extraer una cierta inteligencia o conocimiento de ellos" Sánchez (s.f.). Este proceso representa un gran apoyo a la toma de decisiones de una empresa, para ello se han desarrollado diferentes herramientas y técnicas que permiten almacenar y analizar data histórica y actual.

En este contexto, se manejan los conceptos de datos, información y conocimiento que, a lo largo del tiempo, han ido cambiando y tomando importancia en el día a día de las compañías. El poseer y manejar correctamente los datos para producir información útil determina el éxito o fracaso de cualquier plan de negocios. En la actualidad el problema no radica únicamente en generar data ni dónde, cuándo y cómo almacenarla sino en tener herramientas adecuadas para poder recuperar, según sean las necesidades, toda la información almacenada.

Concepto como el de Data Warehousing, referido a las bases de datos diseñadas con el fin de almacenar grandes cantidades de datos, de sólo lectura, producto de los sistemas transaccionales con el objetivo de analizarlos en base a transformaciones y consolidaciones de data sumarizada con el manejo de dimensiones o variables, adquiriendo un dinamismo en la obtención de información para finalmente eliminar los reportes tradicionales de los sistemas de transacciones; se ubica como un escalón intermedio entre los sistemas operacionales y los de toma de decisiones. Para ello se ha desarrollado otro nuevo término, el de Data Mart, que no es más que aplicar Data Warehousing a un área específica del negocio.

En este trabajo especial de grado, se plasman los pasos cubiertos para diseñar, construir e implantar un Data Mart y una aplicación Web de Intranet para la construcción de reportes en el área de facturación y ventas de la empresa Makro Comercializadora, S.A. sede La Urbina basado en la metodología Three-Stage Process.

La documentación de este trabajo se divide en 6 capítulos. En el primer capítulo, se ubican el planteamiento general del problema: obtener, organizar y mostrar la información referente a facturación y ventas de la compañía. Para esto, es necesario conocer la información generada de los sistemas de transacción en cuanto a las facturas y las combinaciones de variables como clientes artículos, proveedores, etc. Esto, seguido de la definición de objetivos y la justificación del desarrollo del trabajo.

En el segundo capítulo, se reseña el marco referencial para el trabajo especial de grado en el que se desarrollan temas como la inteligencia de negocios, Data Warehouse, Data Mart (objetivos, arquitectura...) procesos de extracción, transformación y carga.

El tercer capítulo, resume la metodología aplicada para el desarrollo del trabajo. La metodología **Three-Stage Process** cubre tres etapas, la primera de **planificación** en la que se obtiene el plan de proyecto con las ideas generales de desarrollo; la segunda etapa, la de **desarrollo**, en la que se obtiene como producto principal el prototipo funcional del Data Mart acompañado de los procesos de extracción, transporte, transformación y carga y finalmente la etapa de **implementación** que produce el Data Mart terminado, la herramienta de consulta de usuario final y el metadata de la aplicación.

En el cuarto capítulo, se expresa el desarrollo de todo el proceso en base a las tres etapas antes mencionadas especificando las actividades realizadas para cumplir con los objetivos planteados así como también los inconvenientes surgidos durante la ejecución acompañados de las

soluciones dadas por el equipo de trabajo. Este capítulo finaliza con las pruebas realizadas sobre la data cargada.

Para cerrar, se encuentran los resultados obtenidos al finalizar la ejecución del trabajo, siendo los principales la estructura de base de datos del Data Mart cargado con data real de desarrollo acompañado de su documentación (metadata) y la aplicación Web de consulta de la información guardada. Esto seguido de las conclusiones obtenidas sobre la elaboración del Data Mart y las recomendaciones para la implementación definitiva con data de producción así como también la infraestructura tecnológica en la que debe ser implementado el sistema final.

# CAPÍTULO I

#### I.1 EL PROBLEMA

#### Planteamiento del Problema

Las compañías de mayoristas comerciantes son empresas de propiedad independiente que se dedican exclusivamente a vender mercancía al mayoreo y que reciben la propiedad de los productos que le son ofrecidos. En este sentido, "Makro Venezuela es la mayor organización mayorista que opera en el mercado, vendiendo grandes volúmenes de productos alimenticios y no alimenticios, a sus clientes profesionales registrados. Su objetivo es abastecer a las empresas, comercios pequeños y medianos y al mercado profesional sin distinción de ramo" (Makro, s.f.).

La empresa Makro comercializadora, S.A. tiene una red de tiendas que supera los 20 establecimientos en Venezuela. Cada establecimiento cuenta con su propio repositorio de información referente a las ventas y facturación en el cual se almacenan, diariamente, los detalles (productos y cantidades vendidas, clientes...) de las ventas. Actualmente, en el día a día de las operaciones de la organización, se generan datos como producto secundario de todas las transacciones que se llevan a cabo. En general, ésta información se almacena, a través de sistemas transaccionales, en bases de datos relacionales que cumplen la función de almacenaje y administración de los mismos. Los datos por sí solos difícilmente pueden aportar algún valor a la organización. Su presencia en los sistemas transaccionales soluciona problemas puramente operacionales.

La forma en la que está almacenada la información sumada a la carencia de una tecnología que apoye la manipulación de los datos referentes a facturación y ventas no permite saber con alto grado de precisión y en corto tiempo, la situación de toda la empresa o de una sucursal en especial enfocado desde el punto de vista de los artículos y los clientes; puesto que en los actuales momentos, el proceso de toma de decisiones en la empresa habitualmente, consiste en tediosos pasos que incluyen la recopilación, clasificación e interpretación de información de ventas y distribución de productos.

Finalmente, todas estas condiciones limitan la toma de decisiones y agregan cierto grado de incertidumbre acerca de los análisis, predicciones y seguimientos que se puedan realizar a determinados rubros de facturación y ventas.

Tomando este y otros factores, tales como el almacenamiento de la información necesaria en los sistemas operacionales, es necesario realizar un proceso Batch para poder reunir la información de ventas de todas las tiendas y así la empresa pueda utilizar esta información para estudios y definición de estrategias de mercadeo, a la vez que se le da respuesta al aumento de los niveles de exigencia de los clientes quienes siguen viendo como en el mundo de la tecnología de la información el hecho de aportar un valor añadido es cada día más esencial.

En este escenario de Makro Comercializadora S.A., es en el que se plantea la necesidad de explorar el futuro a partir del conocimiento detallado del pasado y el presente, se quiere resolver el problema de cómo recolectar, almacenar, consultar, resumir, cruzar y desplegar tanta información referente a facturación de productos que llega de las diferentes sucursales ubicadas a lo largo del país.

#### I.2 OBJETIVOS

# I.2.1 Objetivo General

Diseñar, construir e implantar un Data Mart y una aplicación Web de Intranet para la construcción de reportes en el área de facturación y ventas de la empresa Makro Comercializadora, S.A. sede La Urbina.

# I.2.2 Objetivos Específicos

- Extraer los datos de facturación y ventas de todas las tiendas a través de la implementación de los procesos de extracción, transformación y carga (ETL).
- Construir información a partir de los datos de facturación y ventas extraídos de las tiendas y los datos contenidos en la base de datos de clientes de la empresa Makro Comercializadora, S.A.
  - 3. Construir el Data Mart de facturación y ventas.
- Diseñar e implementar una aplicación Web para la Intranet de la empresa para la emisión de reportes.
- Analizar y evaluar herramientas de usuario final para la aplicación de la Inteligencia de Negocios.

#### 1.3 ALCANCES Y LIMITACIONES

#### I.3.1 Alcances

- Implementar los procesos de ETL (Extract, Transform and Load) para poder extraer los datos referentes a las ventas de las diferentes tiendas y poder transformarlos en información válida para el Data Mart.
- 2. Diseño y construcción de un Data Mart para la empresa Makro Comercializadora S.A. que contemple las dimensiones de Clientes, Artículos, Tiendas, Tiempo y Proveedores incluyendo sus respectivas jerarquías y agrupaciones. Esto aplicado a las tiendas existentes en Venezuela.
- 3. Diseño, construcción e implantación de una herramienta Web para la Intranet de la empresa que permita hacer consultas sobre la información almacenada en el Data Mart antes mencionado.

# I.3.2 Limitaciones

El almacenamiento de la información se realizará en base a herramientas ORACLE, ya que la empresa cuenta con las licencias necesarias.

# I.4 JUSTIFICACIÓN

Sin duda, uno de los recursos más importantes de cualquier organización, hoy en día, es la información. Gracias a ella se pueden tomar decisiones que permiten mejorar el rumbo de las empresas e instituciones. Las compañías han tenido que aprender a "escuchar" lo que dicen los datos que día a día generan. Sobre todo, si están interesadas en descubrir las nuevas necesidades de sus clientes, o bien, desean mejoras en sus procesos. Estos procesos deben estar basados en una investigación que de como resultado la determinación de las vías correctas para conocer y analizar el conjunto de datos que se producen y se manejan día a día.

Un Data Mart es, en esencia, un Data Warehouse a una menor escala ya que posee un conjunto de datos más reducido y se aplica a un área o departamento en particular. Por lo tanto, el objetivo que se persigue en la construcción e implantación de un Data Mart para el área de facturación y ventas de la empresa Makro Comercializadora S.A. es conocer en detalle a los clientes de la empresa, las tendencias del mercado, la forma cómo se posicionan los productos en el mismo, el impacto de las campañas de publicidad y mercadeo, la rentabilidad de cada bien o servicio, etc. En consecuencia, los Data Marts pueden generar tendencias inesperadas y almacenar datos históricos para ayudar en la toma de decisiones, reducir los costos, mejorar el servicio al cliente, y finalmente, generar ventajas comparativas y competitivas para la empresa.

En cuanto a los aportes que genera este trabajo especial de grado, en un primer aspecto, el aspecto teórico, se puede considerar la recopilación de información necesaria para el desarrollo del Data Mart como base a futuros desarrollos. En el aspecto práctico, se generará una implementación completa de los procesos necesarios para construir y alimentar al Data Mart y finalmente, en el aspecto metodológico, la inclusión de metodologías de trabajo para Data Warehousing en la compañía.

Adicionalmente el Data Mart de facturación y ventas constituye una pieza clave de un gran proyecto de Data Warehousing que se llevará a cabo en la empresa. Con el cual se persigue dar a la misma un gran nivel de competitividad con respecto empresas de la competencia y obtener un retorno de la inversión.

# CAPÍTULO II

#### II.1 MARCO REFERENCIAL

# II.1.1 Inteligencia de Negocios

El concepto de "Business Intelligence" (Inteligencia de negocio) fue creado por el grupo Gartner a finales de la década del ochenta. En general, se presenta la inteligencia de negocios como un proceso enfocado en el usuario final. A través de éste proceso los usuarios pueden acceder, explorar y analizar una serie de datos con el objeto de desarrollar conocimiento que les facilite la toma de decisiones. Las decisiones que resulten de este proceso serán mejores, ya que el usuario contará con más elementos de información que le permitan definir el curso de acción más apropiado a seguir en una situación determinada.

Por ejemplo, una factura generada por una orden de venta, es transmitida automáticamente al almacén y al departamento de distribución para su entrega, así como al departamento de producción para programar la siguiente corrida y finalmente al de finanzas y contabilidad para su registro en libros y procesamiento fiscal. Esto es lo que hace un sistema transaccional. Sin embargo, cuando el director de finanzas se pregunta ¿cuál ha sido la variación en ventas después de haberse ejecutado una campaña publicitaria millonaria?, la respuesta nunca llega por parte del sistema transaccional.

Los sistemas transaccionales están diseñados para funciones específicas a la operación no para el análisis. Sus modelos o estructuras de datos están compuestas por una infinidad de tablas, con numerosos campos en miles de

registros y muchas veces los datos son redundantes. Este tipo de estructura, la adecuada para sistemas de éste tipo, limita en gran medida la capacidad de análisis por parte de los usuarios. A esto hay que añadir que por soportar la operación, los sistemas transaccionales consumen muchos recursos durante las horas en que la operación es más intensiva. El querer utilizar éstos mismos sistemas con propósitos de análisis requeriría de una gran cantidad de recursos adicionales de hardware y software.

Hablar de Inteligencia de Negocios es hacer referencia a las técnicas de análisis de datos destinados a encontrar información útil para la toma de decisiones, incluido el conjunto del software que aporta las interfaces y funciones necesarias que apoyan dicho proceso. De esta forma, los sistemas analíticos o de toma de decisiones recuperan los datos de las operaciones diarias de la organización, generados por los sistemas transaccionales, y los procesan, con objeto de tenerlos disponibles para los tomadores de decisiones.

Se distinguen tres tipos tomadores de decisiones, los ejecutivos, quienes toman decisiones de forma analítica e intuitiva, y necesitan mucha información debido a que sus decisiones son de tipo estratégico. En segundo lugar, los administradores y analistas, responsables de decisiones de tipo táctico, quienes lo hacen de manera analítica y también demandan información considerable. Finalmente, los empleados, que toman decisiones rutinarias y cuentan con datos limitados a su dominio. Cada uno de ellos demanda niveles de información diferentes y distintas maneras de consulta.

Sin embargo, y como lo describe Argélida Gómez PC World Venezuela, "la solución de inteligencia de negocios (IN) no es un `pré á porter' sino `un traje a la medida del cliente'. Esto último significa que se trata de una solución diseñada de acuerdo a los requerimientos del negocio empresarial, es decir, que en su diseño se incluyen herramientas y métodos de análisis transaccionales estándares ajustados al tipo de negocio, a sus requerimientos, a sus objetivos estratégicos, tácticos y operativos. Es pues un sistema que se integra al negocio"

#### Funcionalidad De Los Sistemas Analíticos

Para Ramírez (2004), las principales funciones que los sistemas analíticos incluyen, son las siguientes:

- 1. Consultas ad-hoc. Se refiere a un tipo de consulta que permite a los expertos tomadores de decisiones no depender del área de sistemas para realizar una consulta o cruce de información. Con el fin de efectuar este tipo de consultas, el área de sistemas debe generar previamente una descripción en el sistema, a través de la definición de metadatos y de los conceptos de negocio que maneja el experto tomador de decisiones. Con ayuda de los metadatos, el informático hace "la traducción" entre los campos de la base de datos para formar los conceptos de negocios que maneja el experto tomador de decisiones.
- Sistema de semáforos. Mediante éste se puede visualizar el nivel de los indicadores claves del negocio para alertar sobre los problemas que puedan ocurrir en los procesos.
  - Cruces de información.
- Ordenamientos y agrupamientos de información por región, producto y usuario, entre otros conceptos.
  - Sumarización.
  - 6. Comparación de conceptos de negocio en diferentes periodos.
- 7. Generación de gráficos para visualizar de forma sencilla la información.

Las soluciones de inteligencia de negocios proporcionan grandes ventajas al facilitar el análisis de información de las organizaciones. Pero el verdadero reto, que no es más que lograr que estas herramientas de software funcionen de manera adecuada, implica el convencimiento de los directivos y gerentes en empezar a ver en la información y los problemas asociados como entes importantes y muy relacionados con otras áreas del negocio y que no se vean de forma aislada por una unidad funcional.

Entre los conceptos familiares a la inteligencia de negocios, Crespín (2004), director de Sabe de Datos, hace referencia a los siguientes elementos:

- Flujo de información. Proviene de las operaciones y nutre la base de datos analítica, o la base de datos informacional o el Data Mart o el Datawerahouse, dependiendo del alcance.
- 2. <u>Datamining</u> es minería de datos y está basado en herramientas de TI que procesan de manera automática patrones comunes o asociaciones entre los datos. La minería de datos tiene muchas posibles aplicaciones y hay muchos algoritmos diferentes (...)
- <u>Datawerahouse</u> es todo el proceso y la metodología para lograr crear repositorios de datos que contengan los datos necesarios para poder crear los indicadores globales del negocio.
- 4. <u>Data Mart</u> es como un datawerahouse, pero de un área específica del negocio.

La figura No. 1 representa el ambiente general de un Data Warehouse.

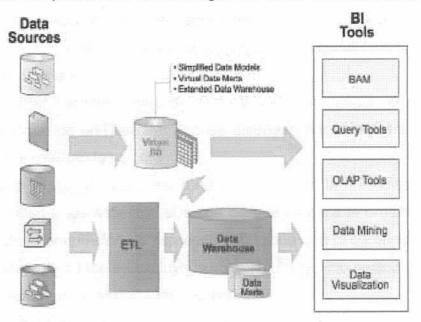


Figura 1. Ambiente de Data Warehouse.
Tomado de: http://www.tdwi.org/Publications/WhatWorks/display.aspx?id=7305

### II.1.2 Data Warehouse

Data Warehouse es un concepto definido desde 1981 por Bill Inmon, considerado el padre del tema.

Un Data Warehouse no es un producto, es un conjunto de componentes de hardware y software que pueden usarse para analizar de una manera prospectiva, los grandes volúmenes de datos que las compañías acumulan día a día a través de los sistemas transaccionales, apuntando hacia la configuración de un Sistema de Información Gerencial que permita la toma de decisiones basadas en información oportuna y veraz. En el Data Warehouse se encuentran replicados los datos operacionales, los cuales necesitan ser convertidos en información que ayude a la empresa a hacer mejores elecciones, tomar mejores decisiones y así obtener una ventaja empresaria.

Sin embargo, hay que tener claridad que mientras los Sistemas Transaccionales (OLTP, On Line Transactions Processing), están diseñados para el manejo eficiente de información de transacciones y soportar la operación del día a día de la empresa, los Sistemas de Información Gerencial apuntan hacia otro tipo de manejo, más estadístico, exploración de tendencias, comparaciones y análisis, procesos para los cuales los sistemas OLTP no están diseñados.

En cuanto a las características del Data Warehouse, según el manual para la construcción de un Data Warehouse diseñado por la Universidad de Antioquia (1997), resaltan las siguientes:

- 1. <u>Orientado al tema</u>: Esto se refleja en cuanto la información es clasificada en base a los aspectos que son de interés para la organización. Se diferencia de la orientación de procesos y funciones en el contenido de la data a nivel detallado donde se excluye la información que no será usada por el proceso de sistemas de soporte de decisiones.
- 2. <u>Integración</u>: Se muestra de diferentes formas, en convenciones de nombres consistentes, en la medida uniforme de variables, en la codificación de estructuras consistentes, en atributos físicos de los datos consistentes,

fuentes múltiples y otros. Los puntos de integración afectan casi todos los aspectos de diseño porque los resultados deben ser los mismos independientemente del modelo que se utilice. La data debe ser almacenada en un modelo global aceptable.

- 3. <u>De tiempo variante</u>: Toda la información almacenada es requerida en algún momento. El tiempo variante se muestra de varias maneras:
  - a. La información representa los datos sobre un horizonte largo de tiempo
  - b. Cada estructura clave en el DW contiene, de manera implícita o explícita, algún elemento de tiempo (días, semanas, meses...)
  - c. La información, una vez registrada correctamente, no puede ser actualizada.

Como se muestra en la figura 2, todos los datos se relacionan con algún momento en el tiempo.

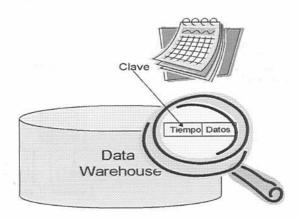


Figura 2. Tiempo Variante en Data Warehouse. Fuente propia.

4. No volátil: La información sólo es útil cuando es estable. La actualización de data solo se hace en el ambiente operacional pero la

manipulación base que ocurre en el DW es sólo de carga inicial y de acceso. Se debe considerar lo siguiente:

- a. Los datos se filtran cuando pasan desde el ambiente operacional al de depósito
- b. El horizonte de tiempo de los datos es muy diferente de un ambiente al otro
- c. El DW contiene un resumen de la información que no se encuentra en el ambiente operacional
- d. La mayor parte de los datos se alteran significativamente al ser seleccionados y movidos al Data Warehouse

En la figura No. 3 se representan las operaciones normales sobre el ambiente operacional y el de Data Warehouse.

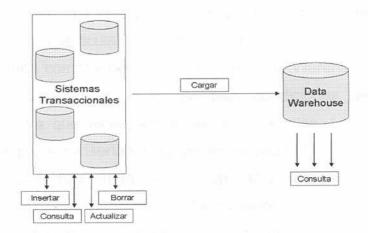


Figura 3. Data no volátil en Data Warehouse. Fuente propia.

#### II.1.2.1 Modelado De Datos Para Data Warehouse

Según Bohórquez (s.f.), "para sintetizar los datos que vamos a encontrar dentro de la bodega van a estar organizados y categorizados en:

**Dimensiones** 

Hechos

Datos de referencia

Datos agregados

Meta datos"

Las dimensiones, califican y manejan las consultas y restricciones del usuario. Generalmente, son las áreas temáticas de interés del negocio o se determinan a partir de las mismas. La interacción entre ellas es la que permite localizar los datos en las tablas de hechos y se relacionan, entre si, a través de las llaves primarias.

Las dimensiones se caracterizan a partir de atributos que pueden ser numéricos, alfanuméricos, imágenes, etc. y cada dimensión posee uno o más atributos y en muchos de ellos puede existir una jerarquía si este está normalizado. "La jerarquía es una relación lógica entre dos atributos dentro de una dimensión". La ventaja de manejar jerarquías está en poder analizar de lo general a lo particular (drill – down y drill – up).

Los hechos comprenden el volumen del Data Warehouse. Un hecho, es un dato instantáneo en el tiempo bajo condiciones definidas por las dimensiones ya que estos son accedidos por ellas. Los registros de los hechos siempre son insertados, jamás actualizados.

Los datos de referencia permiten apoyar la administración de los datos dimensionales, mientras que los datos agregados son todas aquellas sumarizaciones o acumulaciones preestablecidas y almacenadas dentro del Data Warehouse para agilizar consultas. Todo esto se complementa con la documentación de los procesos, tipos de datos y estructuras a través del metadata.

# II.2.1.2 Esquemas de Data WareHouse

Se definen 3 modelos o esquemas de Data Warehouse: esquema estrella, esquema copo de nieve y esquema de constelación. Estos diseños

son muy diferentes del diseño de un esquema de base de datos operacional y se adoptan para maximizar el performance de las consultas.

Según Nigro (2005), En un diseño de base de datos operacional, los datos están altamente normalizados para soportar frecuentes actualizaciones y para mantener la integridad referencial. En un diseño de Data Warehouse, los datos están desnormalizados para proporcionar acceso inmediato a los datos sin tener que realizar una gran cantidad de ensambles.

El esquema estrella es un paradigma en el cual un único objeto en el centro (conocido como tabla de hechos) está conectado radialmente con otros objetos circundantes (llamados tablas de dimensiones) formando una estrella. Este modelo está desnormalizado y puede estar compuesto de una o varias tablas de hechos haciendo centro a cientos de tablas de dimensiones.

Para Bohórquez (2005), las ventajas de este modelo es que es fácil de entender y responde a las consultas de los usuarios optimizando y reduciendo el número físico de uniones requeridas entre los hechos y las dimensiones. Sin embargo, es el menos robusto para la carga y más lento de construir por el nivel de desnormalización.

En la figura No. 4 se muestra un ejemplo de un esquema estrella.

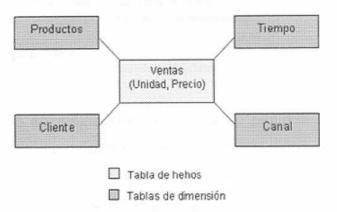


Figura 4. Esquema estrella.

Tomado de: http://www.exa.unicen.edu.ar/catedras/dwhouse/Clase3\_DW.zip

El esquema copo de nieve, es una extensión del esquema estrella donde cada punta de la estrella se explota en más puntas y su denominación se debe a que el diagrama del esquema se asemeja a un copo de nieve. Los esquemas copo de nieve normalizan dimensiones para eliminar redundancia y los datos de las dimensiones se agrupan en múltiples tablas en lugar de una tabla grande.

Con este tipo de esquema, se pueden desarrollar clases de jerarquías que permitan el análisis de lo general a lo particular y viceversa. Sin embargo, muchas uniones en las consultas, pueden sacrificar el desempeño del sistema.

La figura a continuación (No. 5) representa un ejemplo de un esquema copo de nieve.

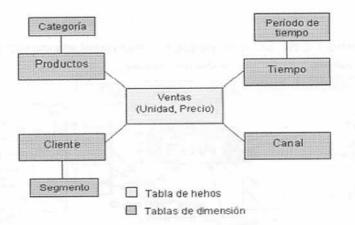


Figura 5. Esquema copo de nieve Tomado de: http://www.exa.unicen.edu.ar/catedras/dwhouse/Clase3\_DW.zip

El esquema de constelación está compuesto por una serie de modelos estrellas donde existe una tabla de hechos principal y una serie de tablas de hechos agregados o auxiliares que no necesariamente están relacionadas con las mismas dimensiones de la principal. La figura No. 6 representa un ejemplo de un esquema de constelación.

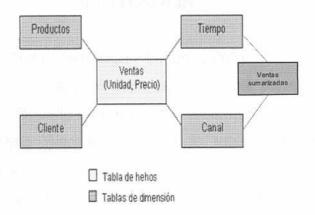


Figura 6. Esquema constelación
Tomado de: http://www.exa.unicen.edu.ar/catedras/dwhouse/Clase3\_DW.zip

La figura siguiente (No. 7) muestra el lugar que ocupa un Data Mart dentro de un ambiente empresarial.

#### The Corporate Information Factory and the Web Environment Departmental Data Marts DSS Applications Staging Area ETL -EDW BURINESS à LELLERAL Changed Data Capture Exploration Global ODS Oper Mart Cross Media Storage Manager Granularity Manager Corporate Applications Firewalls Session Analysis Local DBS Alternative Storage by Elli Iningn and Claudia limbolf Copyright 62201, all rights reserved Dialogue Internet Prefernatied Web Environment

Figura 7. Interrelación del Data Mart con el ambiente empresarial. Tomado de: http://www.inmoncif.com/cifresources/index.html

#### II.1.3 Data Mart

La definición de Data Mart según Inmon (2002): "Un Data Mart es un cuerpo de datos DSS para el departamento que tiene una fundación arquitectónica de un depósito de datos.". Es decir, que un Data Mart, es una extensión de un Data Warehouse y por lo tanto contiene información especializada sobre un departamento o área de la empresa.

Existen diversos nombres para los Data Marts. Entre ellos se tiene:

- 1. Bases de datos DSS departamentales.
- Bases de datos OLAP.
- 3. Bases de datos multidimensionales (md dbms).
- Bases de datos relacionales (r dbms).

Diferentes departamentos dentro de una organización quieren tener su propio Data Mart para el apoyo a la toma de dediciones DSS. Como se muestra en la figura No. 8, existen Data Marts para cada departamento que en conjunto forman el entorno de Data Warehousing de una compañía.

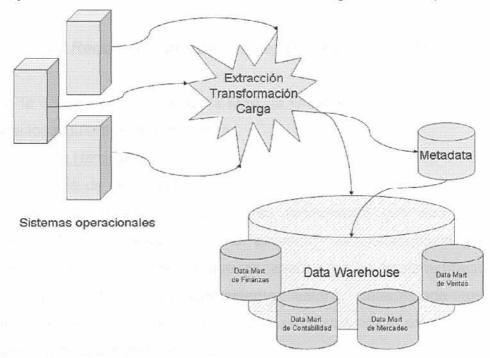


Figura 8. Data Marts por departamento. Fuente propia

# II.1.3.1 Objetivos de un Data Mart

Dentro de los objetivos de un Data Mart, según el autor presentado, están los siguientes:

- Permitir a los usuarios el acceso fácil y consistente a los datos producidos por las aplicaciones operativas de la organización.
- 2. Proveer un canal empresarial de datos para intercambiar y relacionar información de diferentes aplicaciones.
- Proporcionar un depósito estable para el archivo de datos de largo plazo.
- 4. Conservar la historia transaccional de la empresa para llevar a cabo un análisis de las variables en el tiempo y proceder a la toma de decisiones.
  - 5. Permitir el uso estratégico de información detallada.
- 6. Resumir, analizar y filtrar los datos operativos para obtener información valiosa que ayude a la toma de decisiones diaria y futura.

# II.1.3.2 Requerimientos para la construcción de un Data Mart

En la construcción de un Data Mart, la Universidad de Antioquia (2004) plantea los siguientes requerimientos:

# A Nivel de Hardware

Se requiere de un servidor para el almacenamiento y manejo de la base de datos corporativa; este servidor se recomienda que sea altamente escalable, pues algunas veces el proyecto de construcción presenta redimensionamiento a medida que se avanza en la implementación. La capacidad inicial de almacenamiento estará determinada por los requerimientos de información histórica presentados por la empresa y por la perspectiva de crecimiento que se tenga.

Dependiendo del diseño del sistema, puede ser necesario contar con un segundo servidor para las herramientas de consulta de datos. Este equipo debe tener el sistema operativo recomendado por el proveedor de la herramienta a utilizar, siendo el más usado alguna versión de Windows.

Las estaciones de trabajo de cada usuario deberán cumplir con las características recomendadas por el proveedor de la herramienta de consulta seleccionada.

## A Nivel de Software

Se requieren herramientas de software que se clasifican en cuatro categorías básicas: Herramientas de Almacenamiento (bases de datos, bases de datos multidimensionales), Herramientas de Extracción y Colección, Herramientas para Reportes de Usuario Final y Herramientas para Análisis Inteligentes.

<u>Sistemas operacionales base</u>: Las bases de datos operacionales se organizan como archivos indexados, bases de datos jerárquicas, o sistemas de bases de datos relacionales. Estos datos son la fuente principal de alimentación del Data Mart.

Herramientas de Almacenamiento: corresponde a la herramienta en la cual se irán a almacenar los datos. Existen muchas opciones, dependiendo del volumen de los datos, presupuesto, y capacidad del sistema. Cada uno de los sistemas de administración de bases de datos, como Oracle, DB2, Informix, TeraData, Sybase, etc, tienen una facilidad de Data Warehouse.

Herramientas de Extracción y Colección: Ayudan a definir, acumular, totalizar y filtrar los datos de sus sistemas transaccionales en el Data Mart. La mayoría de esas herramientas son desarrolladas por el personal interno de la compañía dado el gran conocimiento que tienen de los sistemas transaccionales. Se requieren herramientas de gestión para extraer datos desde las bases de datos o archivos operacionales, luego es necesario manipular o transformar los datos antes de cargar los resultados en el Data Mart.

Metadata: Describe los contenidos del Data Mart. Consiste en definiciones de los elementos de datos en el depósito, sistema (s) del (os) elemento (s) fuente y como la data se integra y transforma antes de ser almacenada.

Herramientas para Elaboración de Reportes a Usuarios Finales: Es la interfase vista por el usuario. Al usuario se le debe proveer un mecanismo para que vea los datos a un alto nivel, que entonces, obtenga con ello la solución a preguntas específicas. Existen muchas herramientas, incluyendo Cognos Powerplay, Business Objects, SAS, ShowCase Strategy, etc.

Herramientas de Análisis Inteligente: Entre ellas están las de empresas como IBM, SAS, Arbor, Cognos, Business Objects, entre otras. Estas herramientas han sido construidas utilizando inteligencia artificial, buscan alrededor del Data Warehouse modelos y relaciones en los datos. Estas herramientas utilizan una técnica conocida como Data Minning o Minería de datos.

Todas estas herramientas se juntan para crear una arquitectura de Data Mart compuesta de los siguientes niveles:

- 1. <u>Nivel de base de datos externo (base de datos operacional)</u>: Provee una estructura de procesamiento eficiente para un conjunto de transacciones bien definidas. El objetivo al aplicar Data Warehousing es liberar la información que está almacenada y combinarla con otras fuentes de datos externas.
- 2. <u>Nivel de acceso a la información</u>: Está representado por las herramientas que el usuario final utiliza en su día a día incluyendo software y hardware para mostrar la información en pantalla y emitir reportes de impresión.
- 3. <u>Nivel de acceso a los datos</u>: Se debe proveer al usuario final con un acceso a datos universales. Estos usuarios, sin tomar en cuenta la herramienta de acceso a la información o ubicación, deberían ser capaces de acceder a cualquier o a todos los datos necesarios para su trabajo. En

definitiva, este nivel "es responsable" de la comunicación entre las interfaces entre las herramientas de acceso a la información y las bases de datos operacionales.

- Nivel de directorio de datos (Metadata): Es la información alrededor de los datos dentro de la empresa.
- 5. <u>Nivel de gestión de procesos</u>: Tiene que ver con la programación de diversas tareas que deben realizarse para construir y mantener la información del Data Mart y la del directorio de datos.
- 6. <u>Nivel de mensaje de la aplicación</u>: Referido al transporte de la información alrededor de la red de la empresa.
- 7. <u>Nivel físico</u>: Se almacenan muchas copias de datos operacionales o externos en una forma fácil de acceder y de manera altamente flexible.
- 8. <u>Nivel de organización de datos</u>: Se incluyen todos los procesos necesarios para la extracción, manipulación y carga de los datos en el repositorio de información.

# II.1.4 Extract, Transform and Load (ETL)

Lane (1999) plantea que los procesos de Extracción, Transformación y Carga (ETL) son un conjunto de procesos en los cuales los datos son extraídos de numerosas bases de datos, aplicaciones y sistemas; transformados apropiadamente y cargados o almacenados en almacenes de datos, aplicaciones analíticas, etc.

Más de la mitad de todos los proyectos de desarrollo que tienen que ver con Data Warehousing, están prácticamente dedicados al diseño e implementación de procesos ETL. Los procesos ETL que no son diseñados adecuadamente son costosos en mantenimiento, cambio y actualización, por lo tanto lo más importante es hacer las elecciones adecuadas en términos de cuáles serán las tecnologías y herramientas a utilizar para el desarrollo y mantenimiento de procesos ETL.

Específicamente, la extracción es el primer paso para obtener la información que irá hacia el ambiente del Data Warehouse. Es una operación que consiste en copiar data desde los sistemas origen, en general las bases de datos operacionales, y llevarla a archivos que luego serán transformados y cargados en el Data Mart.

Diseñar y crear el proceso de la extracción es a menudo una de las tareas que más ocupa tiempo en el proceso de ETL y, de hecho, del proceso de almacenamiento de los datos enteros. Los sistemas fuente pueden ser muy complejos y la determinación de qué datos necesitan ser extraídos puede ser difícil. Por otra parte, el sistema fuente no puede ser modificado típicamente, ni pueden su funcionamiento o disponibilidad ser afectados para acoplarlo a las necesidades de los datos que se necesita almacenar en el proceso de la extracción. Éstas son las consideraciones más importantes para la extracción, y el ETL en general.

Las técnicas para la extracción se dividen en dos amplias categorías: técnicas que extraen datos de un sistema operacional y almacenan los datos en un archivo. Por otra parte, las técnicas de en las que se extraen datos de un sistema operacional y directamente se transportan datos a una base de datos en blanco, por ejemplo el uso de gateways y queries distribuidos.

Una consideración importante para la extracción es la extracción incremental, también llamada recogida de datos de cambio. Si un almacén de datos extrae datos de un sistema operacional cada noche, después de la primera carga, los únicos datos que ese almacén requiere son los datos que han cambiado desde la extracción pasada (es decir, los datos que se han modificado sobre las 24 horas pasadas). Si fuera posible identificar y extraer eficientemente solamente los datos con cambios más recientes, el proceso de la extracción (así como todas las operaciones en el proceso de ETL) podría ser mucho más eficiente puesto que necesitaría solamente extraer un volumen de datos mucho más pequeño.

Desafortunadamente, para muchos sistemas fuente, identificar los datos recientemente modificados puede ser difícil. La recolección de datos nuevos es el área técnica más ardua de la extracción de datos. Muchos almacenes de datos no utilizan cualquier técnica simple de captura y carga como parte del proceso de la extracción. En su lugar, las tablas enteras de los sistemas fuente se llevan al almacén de datos y estas tablas se comparan con un extracto anterior del sistema fuente para identificar los datos cambiantes. Mientras que este acercamiento puede no tener impacto significativo en las bases de datos operacionales, puede poner claramente una carga considerable a los procesos de almacenamiento, particularmente, si los volúmenes de los datos son grandes.

El segundo proceso del ETL es la transformación. Una vez que la información es extraída hacia el área de tráfico de datos, hay posibles pasos de transformación como; limpieza de la información, desechar lo que no sirve, seleccionar únicamente los campos necesarios para el Data Warehouse, combinar fuentes de datos, haciéndolas coincidir por los valores de los parámetros clave, creando nuevas claves para cada registro de una dimensión.

Las transformaciones de los datos son a menudo la parte más compleja y, en términos del tiempo de transformación, más costosa del proceso de ETL. Muchas, si no todas, las transformaciones de los datos pueden ocurrir dentro de una base de datos, aunque las transformaciones también se ponen en ejecución a menudo fuera de la base de datos (por ejemplo, en archivos "planos").

Al final del proceso de transformación, los datos están en forma para ser cargados. Para ello deben definirse los momentos e intervalos de carga y actualización de la data contenida en el Data Mart.

Resumiendo, se puede decir que en general un Data Mart permite poner a disposición de la alta dirección y los analistas del negocio, la información proveniente de las diferentes áreas de la empresa o institución, con el fin de compararla, analizarla y visualizarla en el nivel de detalle que cada uno requiera, gracias a la facilidad de generación de reportes y creación de relaciones de datos, muchas de ellas ocultas a simple vista, que nos brindan las herramientas informáticas. Sin embargo, para alcanzar este objetivo la empresa debe comprometerse a mantener actualizados los datos que están almacenados en estos repositorios de información especializada y asegurarse que todos los datos son válidos, exactos y oportunos.

## CAPÍTULO III

## III.1 METODOLOGÍA

# III.1.1 Three-Stage Process 1

Consiste en tres etapas: planificación, desarrollo e implantación. La planificación se resume en la entrega de un documento de la definición del proyecto y un plan detallado del mismo. En la etapa de desarrollo se entrega un prototipo y un cronograma de la etapa de implantación. Los resultados de la etapa de implantación son un Data Mart en producción y su documentación. Cada una de las etapas antes mencionadas son llevadas a cabo teniendo en mente las dos leyes de construcción de Data Marts, ley carga de complejidad y ley de diseño incierto, cada una de ellas está diseñada para ayudar a mantener el proyecto alejado de los problemas que puedan surgir.

# III.1.1.1 Etapa 1: Planificación

La etapa de planificación permite definir los objetivos de alcance y negocio para el Data Mart. Esta etapa establece, claramente y de manera realista, el cronograma del proyecto, presupuesto, recursos y restricciones. Un plan de reducción de riesgos anticipa riesgos potenciales y dicta medidas específicas para asegurar el éxito del proyecto.

Los dos puntos quizás más importantes, el documento del proyecto y el plan de proyecto, son creados en la etapa de planificación. Cada uno de

Original en inglés. Ver anexo B

éstos delimita un alcance claramente definido, un amplio cronograma de las actividades del proyecto, su organización y unos procesos de control determinados. El plan de proyecto incluye un cronograma detallado de actividades, asignación de estas actividades, presupuesto y etapas de planificación. Estos son los pasos:

- 1. Determinar el alcance del Data Mart, incluyendo el área de materia de negocios, número de esquemas estrellas y objetivos claves del negocio. Identificar los requerimientos de la organización y restricciones de tiempo y recursos. Identificar los riesgos y sus correspondientes medidas de disminución.
- 2. Identificar el nivel de esfuerzo necesario para crear el Data Mart, y desarrollar un presupuesto de acuerdo con ello.
- 3. Al final de la etapa de planificación, el equipo debe estar de acuerdo con el alcance del proyecto, el cronograma, y el presupuesto. Se debe asignar los recursos de la organización específicamente por nombre y responsabilidad, establecer líneas de comunicación y entregar el plan de proyecto.

### III.1.1.2 Etapa 2: Desarrollo

Utilizando un enfoque iterativo diseño-construcción, la etapa de desarrollo concluye con el prototipo funcional de un Data Mart cargando datos de las fuentes. Este es el momento para definir los objetivos claves del negocio, entrevistando a gerentes, analistas, y administradores de bases de datos para averiguar sus necesidades de información. El equipo de proyecto entonces construye un prototipo y lo carga con datos provenientes de los sistemas "legacy". Los usuarios consultan el prototipo para verificar que el Data Mart cumple con sus requerimientos. Utilizando este enfoque, el equipo descubre tempranamente cualquier problema asociado a la complejidad de la carga y mejora el diseño para conseguir un balance entre los requerimientos

del usuario, y la disponibilidad y calidad de los datos provenientes de las fuentes de información. El equipo desarrolla el prototipo lo suficiente para que los usuarios y desarrolladores consideren que está a un nivel muy cercano al Data Mart definitivo. La etapa de desarrollo generalmente toma entre 12 y 16 semanas.

La etapa de desarrollo genera dos productos: el prototipo del Data Mart y un cronograma para la ejecución de la etapa de implantación. Los resultados del prototipo del Data Mart incluyen una base de datos cargada con data operacional, scripts/sesiones de extracción y carga, instalación y configuración del ambiente de consultas y el diccionario de datos. El plan de proyecto para la etapa de implantación incluye un cronograma detallado de actividades, asignación de las actividades, presupuestos, y el alcance de dicha etapa.

Para cumplir con esta etapa se debe:

- 1. Determinar el número de iteraciones para el Data Mart y la dependencia entre ellas. Determinar qué data será analizada y cargada para cada iteración y cómo serán presentados los resultados a los usuarios.
- 2. Entrevistar a gerentes y analistas de sistemas en el área de interés para determinar sus requerimientos en cuanto a la data que servirá para apoyar la toma de decisiones. Determinar las preguntas que el Data Mart debe responder. Definir los requerimientos de desempeño y de arquitectura para el Data Mart.
- 3. Entrevistar a los administradores de bases de datos y demás individuos que tengan el conocimiento de los datos y las estructuras de los datos de los sistemas fuentes. Analizar la data para identificar los problemas y áreas de carga de complejidad. Localizar las áreas que necesiten de información para cumplir los requerimientos de los usuarios.
- 4. Diseñar el esquema para el Data Mart. Analizar la transformación que los datos tendrán desde que salen de los sistemas fuentes y llegan a almacenarse en el esquema del Data Mart. Diseñar el ambiente de consultas

para el usuario final, incluyendo los reportes requeridos. Verificar el diseño de manera iterativa con gerentes, analistas y administradores de bases de datos.

- 5. Extraer la data de los sistemas fuentes. Verificar que los datos se cargan en el Data Mart de la manera adecuada. Desarrollar las consultas y los reportes requeridos por los usuarios finales. Dejar diccionarios de datos para los datos de los sistemas fuentes, los datos del Data Mart, los datos de la base de datos temporal donde se encuentran los datos recién extraídos de los sistemas fuente, y los pasos aplicados para las limpieza cada uno de los datos.
- 6. Presentar el prototipo a los analistas de negocio, gerentes y administradores de bases de datos. Validar que el Data Mart funcione adecuadamente, cuidando que la data capturada y cargada son las correctas y que la data que se carga en el Data Mart siga los requerimientos de los usuarios. Identificar las áreas problema, donde no se cumplen los requerimientos. Reconfirmar el alcance del proyecto para verificar que se sigue por el camino indicado.

## III.1.1.3 Etapa 3: Implementación

Durante la etapa de implementación, el equipo va construyendo sobre el prototipo. El equipo pone la carga de datos en ejecución, la transformación de la captura de datos, cambios suaves de dimensión (los expedientes que cambian infrecuentemente en un cierto plazo, tal como estado civil de un cliente) y los procedimientos de calidad-aseguramiento.

Se debe entrenar a usuarios para navegar con éxito en el Data Mart, usando una herramienta consulta. Para el momento en que el Data Mart se mueva en el ambiente de la producción y la responsabilidad se ha cerrado al personal de ayuda de la producción, el equipo de proyecto habrá probado a fondo el prototipo y lo habrá templado para el funcionamiento óptimo. El

equipo establece los procedimientos para supervisar, usar y recoger datos del perfil de consulta.

La etapa de implementación crea dos productos principales: un Data Mart y una documentación de los datos de la producción.

Para el Data Mart de producción se incluye una base de datos completamente poblada optimizada para el funcionamiento, el extractor de la producción y los scripts de carga, un ambiente de consulta para el usuario final de la producción y los informes de la producción. Los productos de la documentación incluyen la documentación del sistema, la documentación del usuario, documentos finales del diseño (esquema incluyendo metadata de la estrella) y el entrenamiento inicial de usuarios y del personal de ayuda. Para finalizar esta etapa se deben realizar:

- 1. La construcción del prototipo, determina las actividades necesarias para convertir al prototipo en un Data Mart operacional. Se Desarrolla una estrategia para identificar y manejar dimensiones que cambian lentamente. Se especifican los pasos detallados y las líneas del tiempo. Se asignan los recursos, los papeles del personal, y las responsabilidades.
- 2. Concluya el diseño del prototipo y el extractor, transformador y cargador de la data. Ponga la estrategia en ejecución para identificar y manejar dimensiones que cambian lentamente. Establezca procedimientos de transformación de la carga y sesiones de la carga. Concluya el ambiente y los informes de la herramienta de consulta. Conduzca la optimización del funcionamiento de los componentes del Data Mart. Construya los índices para mejorar funcionamiento de las consultas.
- 3. Con las preguntas del negocio identificadas durante el análisis de requerimientos, desarrolle y ejecute los casos de prueba contra los datos cargados para asegurarse de que contesta a preguntas del negocio correctamente. Asegúrese de que el Data Mart esté agregando datos correctamente, y establezca los procedimientos para asegurar la calidad de cada carga de los datos.

- 4. Con el depósito de metadata del desarrollo, prepare la documentación del sistema. Escriba la documentación del usuario, incluyendo la información sobre el tema y los objetos del Data Mart disponibles para las consultas. Registre el metadata del usuario final en el depósito del Data Mart.
- 5. Determine los procedimientos para la administración y la gerencia del sistema. Establezca los procedimientos de la ayuda del usuario final del Data Mart. Cree los planes de la reserva y recuperación. Desarrolle un plan de entrenamiento, y conduzca el entrenamiento de usuario.
- 6. Realice la prueba de aceptación, evalúe los resultados, y tome la acción correctiva. Realice los chequeos de punto de aseguramiento de calidad en los datos cargados y los scripts de extracción y carga. Pruebe formalmente todos los informes predefinidos. Realice la carga de prueba en la base de datos.
- 7. Cargue el Data Mart y establezca el horario en ejecución para las cargas regulares. Instale un servidor para la herramienta de consulta, al depósito, y a todos los clientes. Conduzca un chequeo final para asegurarse de que todo el hardware y software está instalado correctamente y de que el sistema es completamente funcional. Entregue las responsabilidades del sistema al personal de ayuda del Data Mart. Inicie las operaciones vivas del Data Mart.

## CAPÍTULO IV

#### IV.1 DESARROLLO

En las siguientes páginas se habla del desarrollo del trabajo especial de grado basado en la metodología Three-Stage Process (explicada en el capítulo III). El presente capítulo trata sobre las actividades realizadas en cada una de las etapas planteadas por la metodología.

#### IV.1.1 Etapa de Planificación

La primera etapa que debió ser cubierta fue la de **Planificación**. El paso inicial consistió en la definición de los objetivos. Para ello se llevaron a cabo reuniones con el tutor (Roberto Castañeda) quien es el promotor de la idea de implantación del Data Mart. En estas reuniones se discutió lo referente al tema que sería cubierto por el Data Mart, se estableció el área de facturación y ventas enfocado a las facturas y el contexto que lo rodea como alcance general.

El esquema seguido en las antes mencionadas reuniones fue, inicialmente, de carácter informativo. En las primeras conversaciones se obtuvo una idea general sobre la orientación o dirección hacia dónde se enfocaría el Data Mart. Luego, ya teniendo en cuenta que el proyecto estaba enfocado al tema de facturas, se realizaron las siguientes reuniones con la finalidad de especificar qué rodearía el tema central del Data Mart, es decir, conocer los elementos involucrados en el contexto de las facturas. De estas segundas reuniones se obtuvieron los objetivos de alcance y del negocio.

El siguiente paso fue puntualizar el cronograma de trabajo basado en la metodología seleccionada, el presupuesto, recursos y restricciones, número de esquemas estrella y un plan general de reducción de riesgos. Para poder establecer los elementos del plan de proyecto se realizaron consultas bibliográficas referentes a los temas de Data Warehouse y Data Mart acompañado de referencias básicas a cerca de la inteligencia de negocios. A través de esta documentación se logró conocer las piezas necesarias para definir la base tecnológica y los datos fuentes que alimentarán al Data Mart.

La selección de los recursos a utilizar se basó en los conocimientos adquiridos. A través de ellos se establecieron los requerimientos de construcción del Data Mart a niveles de hardware y software (ver páginas 29 - 31 en el marco referencial). Para cubrir estas necesidades, la compañía puso a disposición equipos específicos y a nivel de software, se cuenta con las bases de datos operacionales de las tiendas y de oficina central, Oracle 8i con instalación para Data Warehousing para el desarrollo de los procesos de extracción, transformación y carga de data.

Para el desarrollo de la aplicación Web de usuario final (de aquí en adelante EIA), el equipo de trabajo hizo selección del servidor Web a utilizar y las herramientas de desarrollo para ello (Ver apéndice A).

El esquema general del Data Mart se organizó en tres esquemas estrella basado en las tablas de hechos **líneas de factura**, pagos de facturas e **impuestos de facturas** relacionadas con las dimensiones de cliente, tiempo, artículos y tiendas. La figura No. 9 representa un esquema general de las diferentes dimensiones y tablas de hechos enunciadas.

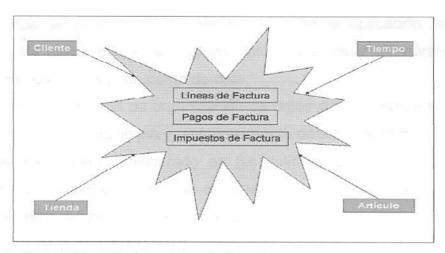


Figura 9. Relación de tablas de hechos y tablas de dimensiones. Fuente propia

El establecimiento del plan de reducción de riesgos se enfocó en cubrir las vulnerabilidades del sistema en los diferentes niveles de la arquitectura de un Data Mart propuestos en el marco referencial.

Para el nivel de base de datos externo se tomó en consideración el manejo de riesgos referentes al control de acceso y manipulación de los datos almacenados por los sistemas operacionales y así no perjudicar su funcionamiento. Para esto se planteó el uso de perfiles de usuario que cuenten con privilegios limitados sobre las operaciones que se puedan realizar sobre la data. Adicionalmente, se puede tener en cuenta el uso de servidores de desarrollo, con data real, para las pruebas de los procesos ETL.

En segundo lugar, el nivel de acceso a la información estará limitado a la definición de cuentas de usuario final de solo consulta de datos a través de la herramienta Web (EIA).

En cuanto al nivel de directorio de datos (Metadata) se hizo consideraciones referentes a la información que debe manejar el personal de soporte a la herramienta determinando que esta debe ser real, específica y referente a la totalidad del sistema.

En lo que se refiere al nivel de mensaje de la aplicación se hizo sugerencias a cerca del acceso a los datos externos y de horarios de ejecución de los procesos de carga de data.

Finalmente, en el nivel de organización de datos debe atacarse el riesgo de disponibilidad de data limpia y real al momento de ser solicitada.

El nivel físico está considerado en las políticas de seguridad de los equipos de la compañía. En gestión de procesos no se identificaron riesgos específicos, esto debido a que este nivel engloba el funcionamiento del resto de los niveles de la arquitectura.

Todos los datos anteriores componen un documento de planificación de proyecto que consta de una descripción general y esquemas específicos agrupados en el plan de proyecto (Ver apéndice A). Este plan de proyecto es el resultado final de esta primera etapa.

#### IV.1.2 Etapa De Desarrollo

La segunda etapa, que plantea la metodología propuesta, es la de desarrollo. Con el objetivo de construir el prototipo funcional del Data Mart, se realizaron diferentes entrevistas, como sugiere la metodología, para conocer las necesidades de información que tuvieran los futuros usuarios finales de la herramienta. La primera recomendación recibida fue revisar una antigua herramienta (que no ha sido utilizada por problemas de instalación) con la que se esperaba obtener reportes del histórico de facturación referente a las ventas por tienda, por productos, proveedores, clientes, etc. Pudiendo reproducir una factura casi tal cual como fue entregada al cliente, además de algunos reportes estadísticos de ranking de artículos vendidos, de clientes, de proveedores, etc.

Por otra parte, se realizaron reuniones con el tutor de tesis para conocer más de las necesidades de información. En una primera instancia se recogieron como preguntas generales a responder la búsqueda de tendencias de ventas de productos y de compras de artículos por parte de los clientes. También se hizo énfasis en poder consultar las facturas de venta de una base de datos histórica de las tiendas, cosa que ya estaba contemplada en el EIA. En base a esta herramienta y a las respuestas obtenidas de los especialistas, se construyeron los objetivos de negocio del Data Mart y se especificaron las preguntas que debería contestar el mismo (Ver apéndice G).

El siguiente paso fue determinar qué data apoya a la toma de decisiones sobre clientes, artículos, proveedores, tiendas y las ventas que cada una produce (facturación de artículos).

En general, las decisiones deben ser tomadas para mejorar las ventas, como es el plan de cualquier distribuidor. Para las ventas en si, se mide si existe aumento en temporada makro mail. Se espera que las ventas suban en estas temporadas y no sólo de los artículos incluidos en cada uno de estos periodos, sino también en el resto de los artículos.

En cuanto a los *clientes*, se hace observaciones sobre el ranking de clientes con mayores compras y se estudia la posibilidad de hacer ofertas especiales dirigidas a ellos. Por otra parte, el ranking de *artículos* más vendidos permite planificar ofertas sobre ellos y con esto lograr aumentar las ventas de otro artículo. En conjunto, se puede segmentar el mercado, es decir, para un tipo de cliente específico ver qué productos adquieren y hacer planificaciones de ofertas para ese grupo.

También se define el tráfico como el volumen de ventas por líneas de factura. El objetivo es determinar si aumentan las ventas con la afiliación de clientes nuevos o si cada ves que van a alguna tienda compran más artículos. De esto se hacen todas las combinaciones hasta poder determinar si hay más ventas pero no nuevas afiliaciones de clientes.

Finalmente, también se toman decisiones sobre los diferentes proveedores. Se observa si los artículos que el proveedor ofrece pertenecen al ranking de productos más vendidos para ofrecer mejores planes de

compra (como solicitar mayor cantidad o mejorar fechas de pago...) al igual que si las ventas de sus productos disminuye se tomará otra decisión.

Todas estas decisiones son tomadas sobre estudios en un rango de tiempo determinado y aplicable a una tienda en particular, a varias o a todas las sucursales de Makro en Venezuela.

Conociendo las necesidades de información y la data para la toma de decisiones se pasó al establecimiento de los requerimientos de desempeño. Esto se hizo en base a los siguientes aspectos:

- 1. <u>Disponibilidad</u>: Los datos almacenados en el Data Mart podrán ser consultados en cualquier momento por los usuarios designados. La información disponible será la obtenida en la última carga, es decir, contiene información hasta el día anterior a la consulta. Esto se debe a que el día en curso se incluirá como histórico cuando se haga el cierre diario de ventas.
- 2. <u>Confiabilidad y consistencia</u>: Se asegura que la data almacenada proporciona información real y es un fiel reflejo de la data que ha sido extraída de los sistemas operacionales. Para ello el producto de la extracción de data será comparado con la data original. En la transformación solo habrá organización y universalidad de los datos extraídos. Finalmente, la data a cargar tendrá una estructura que satisfaga las características del repositorio de datos.
- 3. Robustez: En esencia recuperación de errores cubierto desde la extracción transformación y carga de data, hasta las consultas que harán los usuarios finales.
- 4. <u>Basado en los requerimientos del negocio</u>: El diseño del Data Mart está orientado a contestar las preguntas establecidas anteriormente y en base a la data que apoya a la toma de decisiones.
- 5. <u>Amigable</u>: La interfaz ofrecida para hacer las consultas de usuario final es una herramienta Web diseñada en base a prototipos entregados a los usuarios finales. Corre en la Intranet de la compañía.

# Definición de arquitectura del Data Mart

Esta sección de la etapa de desarrollo fue definida con base en la arquitectura por niveles planteada en el marco teórico.

- 1. Nivel de base de datos externos: En este nivel se ubican las bases de datos operacionales situadas en cada una de las tiendas llamada STORE (de aquí en delante ST) y la base de datos central ubicada en la sede de La Urbina llamada Head Office (de aquí en adelante HO). De ellas se hará la extracción de toda la data que se necesita para mantener el Data Mart.
- 2. <u>Nivel de acceso a la información</u>: se contará con la herramienta de consulta del Data Mart, el EIA (Electronic Invoice Archive).
- 3. <u>Nivel de acceso a los datos</u>: Para este nivel se cuenta con el manejador de base de datos ORACLE 8i (8.1.6). Se realizará una instalación tipo Data Warehousing destinado a la consulta y almacenaje de información.
- 4. Nivel de directorio de datos (Metadata): Se establecerá como contenido de la Metadata el diccionario de datos de la base de datos del Data Mart, los esquemas estrella que especifican dimensiones y tablas de hechos, explicación detallada del proceso de extracción, transformación y carga de la data que proviene de ST y HO y que alimenta al Data Mart.
- Nivel de gestión de procesos: tareas y especialistas que verifiquen periódicamente el correcto funcionamiento de los procesos de mantenimiento del Data Mart.
- 6. <u>Nivel de mensaje de la aplicación</u>: la información viajará al servidor central desde cada tienda a través de enlaces privados (Frame Relay) de CANTV. Se establecerán comunicaciones por medio del protocolo uucp.
- 7. <u>Nivel físico</u>: Para efectos de desarrollo la data se almacenará en los servidores establecidos en el plan de proyecto y serán accedidos desde los computadores personales de los usuarios finales.
- 8. <u>Nivel de organización de datos</u>: en este nivel se ubican todas las tareas programadas para mantener los datos almacenados en el Data Mart.

Específicamente, los procesos incluidos en el ETL que correrán en cada servidor de tienda, para la extracción de los datos ubicados en ST, y los procesos que correrán en HO. Los procesos de extracción serán desarrollados en Pro\*C para Oracle 8i generando un archivo plano con la data solicitada de las diferentes bases de datos operacionales. Este archivo será transformado a través de un nuevo programa en C dando como resultado el archivo plano transformado y listo para ser cargado en el Data Mart a través de la utilidad SQL\*Loader para Oracle 8i.

En la figura No. 10 se muestran los diferentes niveles de la arquitectura del Data Mart y la relación entre ellos.

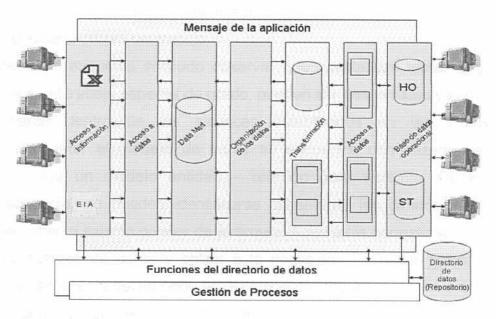


Figura 10.Arquitectura del Data Mart. Fuente propia.

Determinar el número de iteraciones no fue posible debido al poco conocimiento que se tiene sobre las herramientas de extracción y carga. Sin embargo, se pudo determinar a través de estudio de la data necesaria para la toma de decisiones que debe existir una carga inicial con la que se logre reunir el historial de ventas, así como también, lograr reunir en el repositorio la data referente a los clientes, artículos, proveedores, etc. que pertenecen a la tienda. Las siguientes cargas, denominadas cargas incrementales,

contendrán data producida en el día a día de los sistemas operacionales ya nombrados y las nuevas inclusiones o cambios sobre los datos restantes.

Con el fin de lograr estas dos cargas, en una primera instancia se utilizarán datos contenidos en dos servidores de desarrollo que simulan los dos ambientes (HO y ST) y que cuentan con data real cargada de las tiendas. Esto para hacer pruebas de extracción, transformación y carga sin comprometer la data real.

Para este momento se cuenta con información suficiente para hacer el diseño del esquema del Data Mart. Con este fin, se hizo consulta al esquema general de Data Mart que ha sido planteado en las oficinas centrales de Makro a nivel mundial para establecer ideas iniciales referentes a lo que sería el diseño propio del trabajo.

De dicho esquema se pudo observar que cuenta con tablas que, para objetivos del trabajo especial de grado, pueden ser descartadas ya que no se consideran dentro de los alcances establecidos en las entrevistas. Adicionalmente, el modelo de datos se encuentra normalizado de igual manera que un modelo entidad — relación para sistemas OLTP y en consecuencia, no pueden identificarse claramente las dimensiones ni se puede saber si existen niveles de jerarquía. Además la normalización elimina la redundancia de datos y obliga a la necesidad de hacer "joins" y esto precisamente resta desempeño en el tiempo de respuesta.

Se concluye que el esquema planteado no es la mejor solución, ya que parece no seguir las reglas para el diseño de esquemas para Data Warehouse. Además no puede ser clasificado dentro de los tipos de esquema que existen en la teoría de Data Warehousing.

Finalmente, se optó por realizar otro diseño. Uno que posea niveles de jerarquía, sea más desnormalizado y que pueda clasificarse dentro de los tipos de esquemas de la teoría de Data Warehouse. Sin embargo, hay que aclarar que el esquema planteado no puede ser obviado del todo ya que se trata de una solución que será implantada a nivel mundial.

Para el diseño propio del esquema se parte por las tablas de dimensión. Se busca en los esquemas HO y ST las tablas que definirán a estas dimensiones. Luego, se verifica con el tutor y en ocasiones con personal del departamento de informática, si para cada dimensión las tablas halladas satisfacen los requerimientos de información del negocio, de esta manera se puede conocer si se requiere agregar algunas columnas adicionales a la tabla, las cuales a su vez pueden estar en otras tablas, o descartarlas si no traen aporte alguno al Data Mart.

Realizando consultas al EIA surge una nueva dimensión, de nombre makro mail, dentro del esquema del Data Mart. Posteriormente en otra reunión con el tutor surge otra dimensión, de nombre comprador, la cual también es adicionada al modelo de datos.

Para las tablas de hechos, ya mencionadas en la etapa de planificación, se establece una entrevista con el tutor para determinar el grado de granularidad para cada una. Aquí se obtiene como resultado qué significa cada registro para cada una de estas tablas.

- Para la tabla fa\_invoice\_line, cada registro representa una línea de una factura.
- 2. Para la tabla **fa\_invoice\_vat**, cada registro representa el tipo de impuesto que se le aplica a una factura.
- 3. Para la tabla **fa\_invoice\_payment**, cada registro representa la forma de pago aplicada a una factura.
- 4. Para la tabla **fa\_invoice\_multisave**, cada registro representa una promoción aplicada a una línea de una factura.

Una vez que se tiene claro los atributos o columnas a nivel del esquema, se establece el número de tablas para cada dimensión. Esto determinará el tipo de esquema del Data Mart. Las dimensiones CUST (cliente), BUYER (comprador), TIME (tiempo) y STORE (tienda) son desnormalizadas ya que toda la información se encuentra en una sola tabla. Adicionalmente estas

dimensiones poseen una jerarquía desnormalizada. Los niveles que conforman su jerarquía provienen también de una misma tabla.

Sin embargo, las dimensiones ARTICLE (artículo), SUPLLIER (proveedor) y MAKRO MAIL (makro mail) son normalizadas. Por lo tanto estas dimensiones se componen de varias tablas al igual que sus jerarquías respectivas, donde cada nivel proviene de una tabla diferente. Esto hace que el modelo de datos del Data Mart se clasifique como copo de nieve (snowflake), en lugar de considerarse como esquema estrella (star).

A pesar de que el esquema estrella es mencionado a lo largo de la metodología seleccionada hubo que cambiar a snowflake, ya que los datos, para estas dimensiones, no se pueden obtener de un solo esquema, es decir, que provienen de HO y ST.

Por otra parte, los índices que se emplean para desarrollo de Data Warehose son los llamados Bitmap Index. Este tipo de índice se aplica a las columnas de las tablas que se encuentran involucradas en las preguntas que responderá el EIA.

Con el conocimiento obtenido sobre la data para la toma de decisiones y en conjunto con el plan de proyecto generado en la etapa anterior, se establece como dimensiones TIME (tiempo), ARTICLE (artículo), CUST (cliente), SUPPLIER (proveedor), STORE (tienda), MAKRO MAIL (makro mail) y BUYER (comprador). Como tablas de hechos están INVOICE LINE (líneas de factura), INVOICE VAT (impuesto de facturas, INVOICE PAYMENT (pago de facturas) e INVOICE\_MULTISAVE (promociones por factura) (ver apéndice E).

La figura No. 11 representa el esquema copo de nieve utilizado para la tabla de hechos de líneas de factura (FA\_INVOICE\_LINE). En este esquema se refleja la relación de esta tabla con las diferentes dimensiones que componen el Data Mart.

La siguiente tabla de hechos determinada es la de pagos de facturas (FA\_INVOICE\_PAYMENT), la figura No. 13 representa el esquema de copo de nieve para esta tabla.

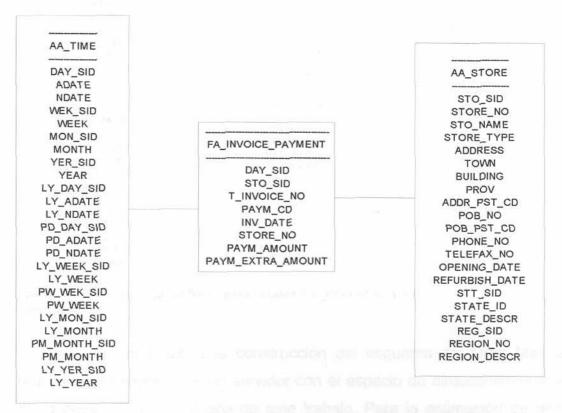


Figura 13. Esquema Copo de Nieve para la tabla FA\_INVOICE\_PAYMENT. Fuente propia.

Finalmente, la última tabla de hechos definida es la de impuestos de facturas. La figura No. 14 representa el esquema de copo de nieve para esta tabla (FA INVOICE VAT).

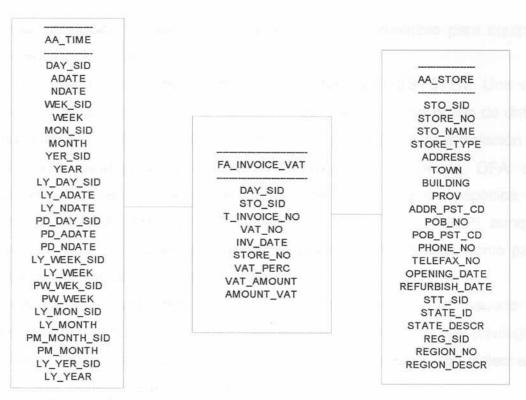


Figura 14. Esquema Copo de Nieve para la tabla FA\_INVOICE\_VAT. Fuente propia.

Para dar comienzo a la construcción del esquema del Data Mart se requiere, inicialmente, de un servidor con el espacio de almacenamiento en disco necesario para efectos de este trabajo. Para la estimación de dicho espacio se realizarán consultas a las bases de datos operacionales de las tiendas, estos queries dan información sobre el espacio, expresado en MB (mega bytes), que ocupan los segmentos de datos y los segmentos de índice de las tablas del historial de facturas las cuales guardan información de un período correspondiente a siete meses aproximadamente.

En vista de que se dispone de tres discos duros cuyas capacidades suman 36GB (dos de 9GB y uno de 18GB), se obtuvo como resultado que para efectos del desarrollo podría almacenarse en disco data de hasta aproximadamente un año de antigüedad. Esto por medio de la proyección a un año de los valores previamente obtenidos (ver apéndice F). Se prosigue

con la instalación del sistema operativo Solaris 8 compatible para equipos con procesador Pentium Intel.

El siguiente paso consiste en crear y montar los FILESYSTEM. Una vez definidos, se continuará con la instalación del manejador de bases de datos relacional Oracle 8i versión 8.1.6.0 en el equipo servidor. Esta instalación se realizará basándose en el modelo OFA propuesta por Oracle. OFA, del acrónimo Optimal Flexible Architecture, consiste en que los espacios de tablas (SYSTEM, USERS, DATA, INDEX, TOOLS, RBS, TEMP) aunque guardan ciertas relaciones a nivel lógico, deben separarse físicamente para evitar problemas de performance.

Luego de la instalación de Oracle se crea el usuario EIA, usuario de desarrollo del esquema, con los privilegios del rol EIA\_DEV. Los privilegios que se obtendrán a través de este rol son: connect, resource, dba, select any table, alter system y create dimension.

Con el usuario EIA se crean los espacios de tablas (tablespaces) de datos e índices para que éste sea dueño de ellos. Se crean los objetos de la base de datos del esquema EIA (tablas, índices, dimensiones, etc.) en los espacios de tablas antes mencionados.

El diseño de la herramienta de consulta para usuarios finales está enfocado a un desarrollo Web que estará disponible en la intranet de la compañía. Este desarrollo se hará, como se estableció anteriormente, utilizando como lenguaje de programación JSP bajo el modelo MVC (Model View Controller). Las aplicaciones de MVC pueden ser implementadas con J2EE utilizando JSP para las vistas, servlets como controladores y JDBC para el modelo.

Fue seleccionado este patrón de programación, porque permite obtener una separación total entre lógica de negocio y presentación siendo su principal objetivo el aislar los datos tanto de la aplicación como del mecanismo para representarlos. Para ello, las aplicaciones se dividen en tres grandes capas: Vista (la presentación de los datos), Controlador (el que

atenderá las peticiones y componentes para toma de decisiones de la aplicación) y Modelo (la lógica del negocio o servicio y los datos asociados con la aplicación) (ver apéndice G).

La separación en capas como presentación, lógica de negocio, acceso a datos es fundamental para el desarrollo de arquitecturas consistentes, reutilizables y más fácilmente mantenibles, lo que al final resulta en un ahorro de tiempo. Es una arquitectura preparada para los cambios, que permite la actualización y desarrollo independiente de cada uno de los citados componentes.

Las consultas que están disponibles en esta aplicación son consultas básicas que muestran respuestas a las preguntas que debe contestar el Data Mart definidas en la etapa anterior. La data será mostrada en forma de tablas dinámicas que resumen la información solicitada. Además, contará con la opción de exportar resultados a Microsoft Excel e imprimir.

Finalizado el diseño de la herramienta EIA y del esquema del Data Mart, se inició el análisis de la data de los sistemas operacionales. Como ya es conocido, la data será extraída de dos esquemas de datos (ST y HO). El primer paso fue ubicar la data que llenaría cada uno de los atributos de los registros del Data Mart. Para ello, se entrevistó a los administradores de bases de datos y demás individuos que tengan el conocimiento de los datos y las estructuras de los datos de los sistemas fuentes, acompañado de una revisión de los modelos y diccionarios de datos de los sistemas operacionales.

En este proceso se determinaron algunos problemas de extracción de data, como por ejemplo, información duplicada en los dos esquemas de datos, para ello se determinó cuál podría ser el mejor origen de los datos estableciendo HO como el repositorio preferido. Esta elección se basó en que en él se centraliza toda la información de las tiendas a nivel nacional. Además, eliminaría complejidad a la hora de hacer carga de datos, ya que la

data sería extraída de un solo esquema y no se tendría que repetir el mismo proceso para cada una de las tiendas.

Otro aspecto que agrega complejidad a la extracción de datos en ST es que no se cuenta con toda la información necesaria para alimentar el Data Mart, por ejemplo, no se almacenan todos los cardholders de todas las tiendas.

Después de este proceso se obtuvo un filtrado de atributos que deberán ser transformados para poder ser almacenados en el repositorio de información. Como por ejemplo, el número de factura debe ser concatenado con el número de caja donde se emitió esta o el caso de algunos datos como la hora de inicio y fin de makro\_action que también deben ser concatenados.

Finalmente, sólo restaba identificar qué datos serían generados o agregados en el proceso de transformación de data. Estos datos son, básicamente, los SERIAL ID de cada uno de los registros de cada una de las tablas que servirán para identificar y hacer único a cada registro. Como otro ejemplo de este tipo de atributos, están los datos que serán almacenados en la dimensión TIEMPO. Además de los datos que vienen vacíos o no existen.

De este proceso de identificación de extracción de data se elaboró una tabla (ver apéndice C) que resume, para cada atributo, su origen, si es extraído de forma directa de los sistemas operacionales, si es transformado en el proceso de ETL o si es agregado en el proceso de carga.

Habiendo precisado qué datos sería necesario extraer de las bases de datos operacionales y la ubicación específica de estos datos, se procedió a hacer el diseño del proceso de extracción de la data (ver apéndice D). Los primeros pasos llevados a cabo ayudaron a definir las sentencias de consulta (SELECT) para cada una de las tablas que serían consultadas. El objetivo principal de esto era hacer consultas directas sobre la data para luego poder comparar los resultados con los datos extraídos por el proceso de extracción.

Junto a este conocimiento y las decisiones anteriormente tomadas a cerca del número de cargas necesarias, se procedió a hacer un diseño global

de extracción de data en PRO\*C. Básicamente se esquematizó en un procedimiento de extracción para cada tabla que posee un cursor que recorre la tabla indicada con la sentencia de selección correspondiente. Al recorrer el cursor, cada uno de los atributos extraídos será almacenado en una variable global que luego es impresa en un archivo plano (con la definición de otro procedimiento para ello), el mismo para todos los datos y uno diferente para cada esquema (en total se obtiene un tipo de archivo para ST – uno por cada tienda – y uno para HO).

Para resolver el problema de identificación de cada una de las consultas diferentes, se antepone en cada línea un valor alfanumérico que identifica cada lista de atributos extraídos (ver apéndice D).

Esta idea general fue individualizada para cada una de las tablas de consulta. Más aún para cada uno de los tipos de cargas diferentes como es el caso de extracción para la primera carga (en el cual viajará el historial de facturación de cada una de las tiendas, el total de artículos con todas sus características asociadas, el total de clientes con todas sus características asociadas al igual que el resto de datos para completar todas las dimensiones) y el caso de las extracciones de carga incremental donde llegarán al repositorio los datos nuevos o que hayan cambiado en el último día.

El problema más grande a solucionar en esta etapa fue la determinación de qué datos eran los que deberían cargarse en las cargas incrementales, ya que sólo debe agregarse la data que haya cambiado desde la última carga. Para ello se hizo uso de las tablas SHADOW, ubicadas tanto en HO como en ST, donde se almacena el nombre de la tabla y la clave del atributo que ha cambiado. Con estos datos se busca en la tabla indicada los nuevos registros o aquellos que hayan sido alterados de alguna forma para ser cargados en el repositorio de información del Data Mart.

El segundo problema a enfrentar es referente a los datos que, en el modelo OLTP del sistema operacional, podían ser de valor NULL y en el

repositorio de datos final son NOT NULL. Para solucionarlo, se colocaron valores que cubran el espacio vacío de acuerdo al tipo de dato a almacenar.

Definido el proceso de carga, se hizo pruebas sobre la data de desarrollo con el fin de conocer el archivo plano que este proceso genera como respuesta. Este proceso de pruebas fue elaborado de forma incremental, agregando una consulta a la vez y haciendo ajustes sobre la data de respuestas (eliminar valores repetidos corrigiendo alguna consulta errada...) para obtener los resultados esperados y además estimar tiempos de respuesta para generar cada uno de los archivos planos. El resultado final de este paso fue la obtención de un archivo plano para cada esquema en el que se concentra la información extraída de los sistemas operacionales.

Sobre los archivos planos obtenidos se aplica el proceso de transformación. Inicialmente, este proceso fue pensado para ser desarrollado en lenguaje C. Al estudiar los datos faltantes, se determinaron tres casos esenciales de transformación:

- Los datos que no serán transformados ya que vienen "listos" para ser cargados en el Data Mart.
- 2. El generar los SERIAL ID de identificación de los registros pertenecientes a cada tabla de dimensión.
- Asignar valores de NULL para las columnas que permiten este valor (y lo tienen) tanto en los sistemas operacionales como en el Data Mart.

El objetivo de este desarrollo fue crear un proceso de transformación que consistiera en leer cada una de las líneas del archivo plano y ajustar su estructura a una línea válida de inserción en el Data Mart, generando un nuevo archivo plano (ya transformado) que será utilizado por utilidad SQL\*LOADER para actualizar el repositorio de información.

El primer caso no presenta mayor inconveniente, ya que simplemente se toman los mismos datos y se colocan en la base de datos.

El segundo caso (el manejo de los SERIAL ID) hizo cambiar la opción inicial de lenguaje de programación de C a PRO\*C, esto porque para obtener

los nuevos SERIAL ID debe accederse al Data Mart. Por lo tanto se decidió utilizar una herramienta propia del manejador de base de datos y que ofrece facilidades de acceso a los datos. El valor de SERIAL ID es un número de ocho posiciones (NUMBER 8) lo que permite almacenar valores desde 0 hasta 9999999 (una secuencia de cien millones de valores distintos). El valor del SERIAL ID se llevará de manera secuencial para cada conjunto de valores que conformaban claves primarias en las tablas normalizadas de los modelos OLTP de los sistemas fuentes.

Para obtener cada uno de los nuevos valores de identificación se pregunta por el valor más reciente que ha sido asignado como SERIAL ID para el conjunto de datos antes mencionado. El valor siguiente en secuencia, es decir, incrementado en una unidad, será asignado como el nuevo SERIAL ID cada vez que se introduzca un nuevo registro en el Data Mart (ver apéndice D). Este nuevo valor será "escrito" en el archivo plano de carga.

En el caso de la primera carga, no existen valores previos almacenados de secuencia de SERIAL ID por lo que fue necesario buscar un almacén temporal para los valores que se van generando. No sólo para poder llevar la secuencia incremental sino también para poder hacer relación entre las diferentes tablas del Data Mart, por nombrar un ejemplo, las tablas de hechos llevan como claves primarias las claves foráneas de las diferentes dimensiones (la tabla fa\_invoice\_line está relacionada con los SERIAL ID de las dimensiones de tiempo, artículo, cliente, etc. Y así todas las demás). El proceso normal buscaría los valores de SERIAL ID en las diferentes tablas de dimensiones y al no hallar ningún valor quedarían vacíos los campos de identificación de la tabla.

Esto se solucionó almacenando estos valores, a medida que son generados, en arreglos unidimensionales o bidimensionales. Esto para el caso de las tablas que poseen como clave única el SERIAL ID y que además los valores válidos no superen la cantidad de posiciones aceptadas para definir un arreglo en PRO\*C.

En el caso de no cumplirse ninguna de las dos condiciones anteriores, se hizo uso de una tabla auxiliar denominada SERIAL\_ID en la que se almacena el valor generado, el identificador del atributo y el nombre de la tabla de origen de este valor. Tal es el caso de los clientes cuya cantidad de registros supera el millón seiscientos.

Para almacenar valores NULL en los campos que lo permiten, ultimo caso a abordar, se escribirá en el archivo valores como N/D para indicar que ese dato no está disponible.

Luego de algunas pruebas iniciales, este proceso fue descartado, ya que el tiempo de respuesta obtenido era muy grande (aproximadamente 5 días solo para carga inicial). La razón principal por la que este tiempo era tan prolongado era el acceso a lectura de un archivo plano.

Finalmente, y contando ya con el archivo de extracción de data, se decidió ejecutar carga masiva, sobre una base de datos temporal con el uso de **staging tables**, de los datos extraídos. Sobre esta base de datos temporal se realiza el proceso de transformación con ayuda de arreglos y una tabla para almacenar los SERIAL ID generados que permiten resolver los problemas anteriormente mencionados.

El resultado de este proceso es un archivo plano con extensión .dat que será utilizado por la utilidad SQL\*LOADER para finalmente ser cargado en el Data Mart.

El proceso de carga se abordó utilizando procesos de especificación de la posición del archivo de data. Para cargar data del archivo de datos, "el SQL\*Loader debe conocer la posición de cada uno de los campos y su longitud. Para especificar la posición en el registro lógico, se usa la palabra POSITION en la especificación de la columna. La posición puede o no ser expresada de forma explícita o relativa al campo precedente. Los argumentos de posición deben estar encerrados entre paréntesis". Este proceso se ejecuta en el archivo stngtp94b.ctl.

### IV.1.3 Etapa de Implementación

La última etapa de la metodología planteada es la etapa de implementación. El primer punto abordado fue el emplear una estrategia para manejar los cambios suaves sobre las dimensiones. Dicha estrategia consiste en el uso de las tablas SHADOW tanto del esquema HO como de ST en los servidores de desarrollo (al igual que en los servidores de producción).

Las tablas SHADOW almacenan valores de los registros que han cambiado o hayan sido creados, por tabla, en los sistemas operacionales durante el día en curso. Los datos que se guardan son el nombre de la tabla alterada y, separados por comas (,), los valores claves de cada registro cambiado / creado. Con estos datos se busca, en cada tabla de los sistemas operacionales, el registro a incluir en el Data Mart por cualquiera de las condiciones antes mencionadas. Con ellos se obtiene un archivo con los datos extraídos de los sistemas operacionales.

A esto se anexan los datos producidos diariamente en las tablas de líneas de factura (invoice\_line), impuestos por factura (invoice\_vat) y pago de facturas (invoice\_payment).

Una vez extraídos los datos que corresponden al día en curso, cambios suaves en las dimensiones y la nueva data de facturas, se aplican los procesos de transformación correspondientes, obteniéndose como resultado un archivo de datos (.dat) donde la data posee la estructura adecuada para ser cargada.

Para finalizar, se cargan en el Data Mart los datos transformados. El proceso de carga con la utilidad SQL\*Loader de Oracle toma el archivo .dat y vacía su contenido en un primer esquema de pruebas de carga. Este esquema no es el definitivo para el Data Mart, cuenta con las siguientes características:

Tablas con la misma estructura.

- No posee índices. Este esquema no será consultado y por lo tanto no se requiere agilizar las consultas.
- Uso de constraints para garantizar la consistencia e integridad de los datos.

El objetivo de estas pruebas es verificar el comportamiento de los procesos ETL al encontrarse con estos cambios. De acuerdo a los resultados obtenidos en los bad files, discard files y los log files se hace depuración del código de transformación o carga según sea el caso.

Al momento de hacer la primera prueba, se presentaron problemas tales como:

- Formatos de fechas que no se reconocían y que debieron ser ajustados en el proceso de transformación.
- Para el archivo ctl, necesario para la carga, debió ajustarse algunas posiciones de datos ya que en algunos casos se obtuvo data truncada o que no se ajustaba al formato por falta de algún valor
- Revisar algunos CHECKS de las tablas destino ya que fueron creados en base a los OLTP de los sistemas transaccionales, pero estos con el paso del tiempo han sido modificados algunos de los cambios no han sido documentados.
- 4. La data de desarrollo proporciona algunos casos de información incorrectos con el objeto de poder hacer pruebas de comportamiento del sistema en caso de fallas.

La figura No. 15 muestra ejemplos del archivo log generado por el SQL\*Loader en algunas de las pruebas. Se observa el número del registro que ha sido rechazado, la tabla donde este iba a ser almacenado y la razón de la falla de inserción.

```
Record 8028: Rejected - Error on table STG_CUST.

ORA-01400: cannot insert NULL into ("EIA"."STG_CUST"."ADDRESS")

Record 8029: Rejected - Error on table STG_CUST.

ORA-01400: cannot insert NULL into ("EIA"."STG_CUST"."ADDRESS")

Record 8030: Rejected - Error on table STG_CUST.

ORA-01400: cannot insert NULL into ("EIA"."STG_CUST"."ADDRESS")

Record 8068: Rejected - Error on table STG_CUST.

ORA-01400: cannot insert NULL into ("EIA"."STG_CUST"."ADDRESS")

Record 8069: Rejected - Error on table STG_CUST.

ORA-01400: cannot insert NULL into ("EIA"."STG_CUST"."ADDRESS")
```

Figura 15. Ejemplo de archivo log. Fuente propia.

El archivo de registros descartados se copia las líneas del archivo a cargar que han sido rechazadas. En la figura No. 16 se muestra un ejemplo de este archivo.

	700		1 (a (i.e.)	THE RESERVE THE PROPERTY OF TH				
A	10	1	101	254282000121020010115	6900.00	0.000	0.0000	0.00020050616
A	15	1	101	272742000121020010115	12950.00	0.000	0.0000	0.00020050616
A	19	1	101	555232000121020010115	750.00	0.000	0.0000	0.00020050616
A	20	1	101	555882000121020010115	6069.86	0.000	0.0000	0.00020050616
A	21	1	101	578892000121020010115	3851.53	0.000	0.0000	0.00020050616
À	22	1	101	586042000121020010115	6541.49	0,000	00000.0	0.00020050616
l.	25	1	101	614512000121020010115	3851.53	0.000	0.000	0.00020050616
A	27	1	101	615162000121020010115	5842.80	0.000	0.0000	0.00020050616
A	28	1	101	615422000121020010115	3851.53	0.000	0.0000	0.00020050616
A	29	1	101	616332000121020010115	3851.53	0.000	0.0000	0.00020050616
A	30	1	101	617762000121020010115	5842.80	0.000	0.0000	0.00020050616

Figura 16. Ejemplo de archivo .dsc. Fuente propia.

El proceso de pruebas de carga se ejecutó en 4 oportunidades hasta que se logró obtener carga limpia de la totalidad de los datos extraídos y transformados. En la figura No. 17, se muestra el ambiente de prueba de los procesos de extracción, transporte, transformación y carga. Con esto, se procedió a ejecutar la carga de la data en el modelo definitivo de Data Mart. A diferencia del modelo anterior, en este modelo fue necesario establecer los índices para agilizar las consultas sobre los datos extraídos del servidor de desarrollo. En cambio, ya asegurado que la data se carga correctamente y

para agilizar este proceso, se desactivan los constraints de claves primarias y claves foráneas.

Entomo de desarrollo

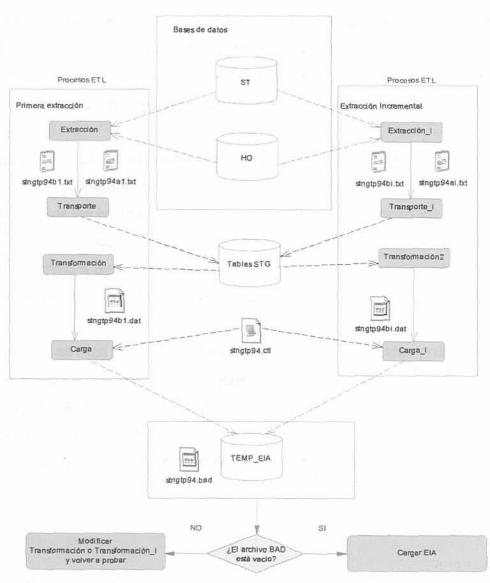


Figura 17. Ambiente general de desarrollo. Fuente propia.

Esta carga alimentó el Data Mart con todos los registros extraídos del servidor de desarrollo. Ya que toda esta data es data real extraída de los dos esquemas (HO y ST) se puede asegurar que al pasar los procesos a

producción se obtendrá de igual manera la extracción, transporte, transformación y carga exitosa de los registros.

Superado el primer escalón, la data almacenada y lista para ser consultada, se procedió a llevar a consultas en SQL las preguntas del negocio identificadas en las etapas anteriores. En una primera instancia, se fueron desarrollando las consultas directamente sobre la data a través de la herramienta SQL Plus. Con esto se hizo la verificación de que la data cargada responde a las preguntas del negocio establecidas al principio del proyecto

El siguiente paso fue la construcción final de la herramienta de consulta EIA. Como fue planteado en la etapa anterior, se hizo uso del patrón de diseño MVC. Por las bondades del patrón escogido se desarrolló en paralelo la interfaz de búsqueda de información es decir, lo referente a las vistas y el modelo de acceso a datos que ejecuta las sentencias de consulta definidas y permite mostrarlas en la interfaz Web a través del controlador (ver apéndice G).

Finalmente, se hizo pruebas de funcionamiento del EIA. Estas pruebas se realizaron en conjunto con el tutor quien ejecutó consultas sobre la data cargada y de acuerdo a los criterios de búsqueda plasmados en la herramienta. Los resultados obtenidos se muestran en el apéndice G.

Con este último paso se concluyó el trabajo especial de grado.

### CAPÍTULO V

#### V.1 RESULTADOS

Los resultados obtenidos en el desarrollo de este trabajo especial de grado son los siguientes:

- 1. Con la documentación teórica básica acerca de los temas referentes al desarrollo de un Data Mart, empezando por la parte más amplia como lo es la Inteligencia de Negocios, hasta llegar a detalles tales como Data Warehousing y Data Mart, también incluyendo documentación referente a los procesos necesarios para alimentar la estructura; se logró obtener conocimiento para enfrentar la puesta en marcha del proyecto.
- 2. De la primera etapa de la metodología implementada (etapa de planificación) se obtuvo un plan de proyecto en el que se sentaron las bases para su desarrollo
- 3. Se definió, sobre el servidor de trabajo, las estructuras y espacios de almacenamiento para tablas, índices, usuarios, dimensiones, jerarquías, etc.
- 4. Con ayuda de las entrevistas al personal de la empresa se logró determinar las necesidades de información que sirvieron como base para establecer las preguntas que contesta el Data Mart y con ellas la definición de las dimensiones y tablas de hechos para almacenar la data necesaria
- 5. En base a los puntos anteriores unido a la consulta al modelo de Data Mart propuesto en Makro a nivel mundial se logró definir un esquema de almacenamiento que cumpliera con las dimensiones y tablas de hechos propuestas
- 6. Con el estudio de los modelos OLTP de los sistemas operacionales, específicamente de los esquemas de Head Office (HO) y Store (ST), se logró

determinar qué datos serían extraídos de los mismos para alimentar a la estructura de almacenamiento. Así mismo, determinar el mejor origen de los datos (dependiendo de cada uno de los casos) ya que hay presencia de redundancia de información en los dos esquemas. Además se pudo establecer qué data no estaba disponible.

- 7. Para cumplir con la solicitud de almacenamiento de data histórica y data actual, a través de la consulta al modelo OLTP se pudo definir los procesos de carga inicial y de carga incremental. Las dos cargas constan de procesos de extracción, transporte, transformación y carga. Los procesos de extracción asociados a la carga inicial hacen consultas sobre las tablas históricas de facturación de cada tienda así como también en el resto de las tablas necesarias (indicadas en el apéndice B). En cambio, la extracción para carga incremental consulta directamente los cambios realizados sobre las tablas de HO y ST asentados en la tabla SHADOW (en la que se almacenan los nombres de las tablas que han sufrido cambios y las claves primarias de los registros alterados) y las tablas de líneas de factura, pagos de facturas e impuestos de facturas.
- 8. Se desarrollaron cuatro archivos de extracción de data en el lenguaje PRO\*C stngtp94a1.pc y stngtp94b1.pc para los esquemas de HO y ST respectivamente y stngtp94ai.pc y stngtp94bi.pc para los esquemas de HO y ST respectivamente. Los dos primeros para cumplir con la primera extracción y los dos últimos para realizar la extracción incremental. Con ellos se hace consulta de la data, contenida en estos dos esquemas, necesaria para el Data Mart. Esta información es vaciada en dos archivo planos diferentes (uno para HO y otro para ST según sea el caso).
- 9. Para acortar los tiempos de ejecución del proceso de transformación se hizo uso de un proceso de transporte (stngtp94a.ctl) de la data extraída a unas tablas temporales (staging tables). Estas estructuras fueron creadas para cargar de forma masiva, con el uso de la herramienta SQL Loader, la

información contenida en los archivos planos obtenidos del proceso de extracción.

- 10. El proceso de transformación se plasmó en un programa, en lenguaje PRO\*C, llamado stngtp94t.pc encargado de hacer lectura de la data transportada a las tablas temporales después de la extracción completando la información necesaria para alimentar el Data Mart y generando un nuevo archivo plano con la data limpia y lista para ser cargada.
- 11. Se desarrolló un segundo proceso de carga masiva (stngtp94b.ctl) para cargar la data transformada directamente al Data Mart.
- 12. Acompañado del desarrollo del Data Mart, se generó documentación necesaria para conocer los elementos que componen a la herramienta, tales como: esquema de dimensiones, esquema de ambiente general de funcionamiento, diccionario de datos para conocer las tablas y los atributos pertenecientes a cada una de ellas especificando claves primarias, índices, tamaños y valores predeterminados entre otras cosas (ver apéndice E).
- 13. Se cargó en la estructura de base de datos la data contenida en el servidor de desarrollo de la compañía (6 meses de facturación). Esta data es real y extraída de los diferentes esquemas, fue utilizada para probar los diferentes procesos y asegurar que están listos para ponerse en producción.
- 14. Se implementó una herramienta Web de consulta para la data almacenada en el Data Mart. Esta herramienta fue denominada Electronic Invoice Archive (EIA) y permite hacer consultas, por diferentes criterios, de las facturas almacenadas en la base de datos así como también consultar algunos ranking de artículos, clientes y ventas.
- 15. Con el desarrollo de este trabajo especial de grado, se logró dar el primer paso para la implementación de Data Warehousing en la compañía Makro Comercializadora S.A. sede la Urbina, ya que el desarrollo de este Data Mart es el inicio de un proyecto de Inteligencia de Negocios que se está llevando a cabo en la compañía a nivel mundial.

16. Este desarrollo también contribuye a eliminar la carga de data histórica de facturación de los sistemas operacionales. Actualmente, esta información se almacena en los esquemas de operación, con la utilización correcta de este sistema esta información ahora puede ser almacenada en el Data Mart, eliminando así carga de data no necesaria para el proceso diario.

## CAPÍTULO VI

#### VI.1 CONCLUSIONES

- 1. Los sistemas operacionales producen a diario gran cantidad de data que, por si sola, no brinda información que permita hacer análisis del proceso diario del negocio. Es por ello que debe ser organizada y clasificada para que pueda traducirse en información útil que permita ver la integración de los datos de los procesos.
- 2. La aplicación de herramientas de Inteligencia de Negocios permite capturar y organizar la data que se produce a diario en los sistemas operacionales de la empresa con ello se puede rastrear información vital para la toma de decisiones.
- 3. Data Warehouse es, en general, un depósito en el que se almacenan los datos que se generan de las operaciones básicas de funcionamiento de una empresa. Esta información se utiliza para saber cómo está "marchando" el negocio. A través de la implementación de un Data Mart se puede sectorizar la información haciéndola corresponder a un área o departamento en particular.
- 4. Un Data Mart puede ser utilizado para incorporar y reunir la información de diferentes esquemas y dependencias, permitiendo generar una visión integral del funcionamiento del área de estudio. Los datos se encuentran en múltiples lugares hay que transformarlos para traducirlos a un idioma común
- 5. La creación del proceso de transformación de la data extraída consume la mayor parte del tiempo del desarrollo de un Data Mart.

- 6. Oracle Data Warehousing es una instalación de Oracle que permite crear jerarquías y dimensiones para la implementación de un Data Mart. Además, ofrece opciones especiales de consulta para este tipo de estructuras.
- 7. Para el desarrollo de un Data Mart, es fundamental tener en cuenta las necesidades de información que plantean los usuarios finales. De nada sirve una implementación que no de respuesta a las preguntas definidas por los futuros usuarios.
- 8. La aplicación Web desarrollada (EIA) solo responde a consultas fijas sobre la data cargada es por ello que el Data Mart debe ser acompañado de una herramienta de consulta de nivel superior para toma de decisiones ya que la información almacenada es solo el nivel base de almacenamiento.
- Los datos almacenados en el Data Mart se convierten en información que aporta conocimiento a la compañía Makro Comercializadora S.A. referente a sus transacciones de facturación diaria.

#### VI.2 RECOMENDACIONES

Después de desarrollado el trabajo especial de grado, se hacen las siguientes recomendaciones para llevar el sistema a producción:

- 1. En base al tamaño de los registros calculados en el apéndice F y tomando en consideración la preferencia de la compañía en la adquisición de equipos desarrollados por SUN, se recomienda la adquisición de un servidor con las siguientes características:
  - a. Data Warehouse Server: Servidor V490 (2\*1.5GHz UltraSparc IV+, 8GB RAM).
    - b. Servidor de Desarrollo: Servidor V240 (2\*1.5GHz, 2GB RAM).
  - c. Servidor de Aplicación: Servidor Sun Fire X4100 (2\*AMD Opteron 252, 2GB RAM).
  - d. Ambiente SAN: Storage: Sun StorEdge 3510 FC (12\*146GB) –
     Switch Brocade 3250.
    - e. Unidad de respaldo C4 (2 Drives LTO3 conexión a FC).
- 2. Como Herramienta de usuario final, en base al apéndice H, se recomienda la adquisición de MicroStrategy 8. Ya que, de las herramientas consultadas, ofrece diferentes módulos de análisis de data y generación de reportes que además, son compatibles con diferentes estilos de Business Intelligence.
- 3. Incluir los procesos de extracción de data en el proceso de night run antes de que este proceso elimine el contenido de las tablas SHADOW.
- 4. Los procesos de carga incremental deben ser ejecutados a diario, ya que la información que se desea almacenar (principalmente las líneas de factura) se genera diariamente.
- Revisar a diario los archivos de log y bad files generados por el SQL Loader en la carga masiva de la información para verificar que los datos se hayan agregado correctamente.

- 6. Complementar el EIA con nuevas vistas, aprovechando la implementación en MVC, para los diferentes tipos de usuarios.
- Entrenar a los usuarios finales en el uso de la herramienta final que se seleccione para trabajar sobre la plataforma de Data Mart desarrollada.
- 8. Entrenar a los administradores del repositorio de información para mantener la data con revisiones oportunas de los archivos log que se generan de la carga masiva de información.

## REFERENCIAS BIBLIOGRÁFICAS

#### Material Impreso:

Hanna, P. (2002). JSP. Manual de referencia. España: Madrid

Inmon, W. (2002). Building the Data Warehouse (3rd Edition). United States of America.

Portfolio, T. (1992). Programmer's Guide to the ORACLE Precompailers.

Versión 1.5. United States of America: Oracle.

Portfolio, T. (1992). Pro\*C Supplement to the ORACLE Precompilers Guide.

Versión 1.5. United States of America: Oracle.

#### Material electrónico:

Bohórquez, J. (s.f.). Aproximación metodológica de un Spatial Data Warehouse. Disponible en: gis.esri.com/library/userconf/ latinproc00/colombia/spatial\_data.pdf [Documento en base electrónica]

Lane, Paul. (1999). Oracle 8i Data Warehousing Guide, Release 2 (8.1.6). [Documento en base electrónica].

Nigro, Oscar. (2005). Estructura del Data Warehouse. Disponible en: http://www.exa.unicen.edu.ar/catedras/dwhouse/Clase3 DW.zip

Van De Put, Jacco. (2005). MBS2000HOTables. [Documento en base electrónica]. Disponible en Makro, sede la Urbina.

Van De Put, Jacco. (2005). MBS2000STTables. [Documento en base electrónica]. Disponible en Makro, sede la Urbina.

#### Páginas Web:

Bitam. (2005). Soluciones. [Documento en Línea]. Disponible en: http://www.bitam.com/spanish/Soluciones.htm. [Consulta: 2005, septiembre 30]

Bitam. (2005). Papiro. [Documento en Línea]. Disponible en: http://www.bitam.com/spanish/Papiro.htm. [Consulta: 2005, septiembre 30]

Bitam. (2005). Artus. [Documento en Línea]. Disponible en: http://www.bitam.com/spanish/Artus.htm. [Consulta: 2005, septiembre 30]

Elizondo, Mónica. (2002). Sistemas de toma de decisiones, [P{agina Web en Línea].

Disponible en: http://www.netmedia.info/netmedia/articulos.php?id\_sec=32&id\_art=3678&nu m\_page=10777. [Consulta: 2005, septiembre 26]

Elliot, J. (1997) Manual para la construcción de un Data Warehouse, [Documento en Línea]. Lima: Instituto nacional de estadísticas e informática INEI. Disponible en:

http://200.14.84.223/apuntesudp/docs/civil\_ind/(ICI2423)Sistemas\_de\_Inform acion/(2003-06-11)\_818\_datawarehouse\_o\_bodegaje\_de\_datos.pdf [Consulta: 2005, 25 de junio].

Gaxiola M., Jesús A. Sistema de Soporte a la Toma de Decisiones (DSS).

[Página Web en Línea]. Disponible en http://www.geocities.com/SiliconValley/Pines/7894/introduccion/dss.html

[Consulta 2005,septiembre 27].

Gómez (s.f.). Análisis de gastos usando inteligencia de negocios, [Documento en Línea]. Chile: Universidad técnica Federico Santa María. Departamento de informática. Disponible en:

http://portal.inf.utfsm.cl/PTE2003/Memorias/MEMORIA\_4.pdf [Consulta: 2005, junio 25].

Gutiérrez E., Damián. (s.f.) Data Warehouse, [Página Web en Línea]. México: Universidad Iberoamericana. Disponible en:

- http://www.monografias.com/trabajos17/data-warehouse/data-warehouse.shtml#losproce. [Consulta: 2005, junio 20].
- Inmon. (2003). Bill Inmon's Resource Center [Página Web en Línea].

  Disponible en: http://www.inmoncif.com/cifresources/index.html [Consulta: 2005, febrero 10]
- Inmon. (2003). Bill Inmon Library [Página Web en Línea]. Disponible en: http://www.inmoncif.com/library/methods/dwmeth\_frame.asp [Consulta2005, 10 de febrero]
- Jones, G. (1998). Building effectives Data Marts [Documento en Línea]. United States of America: Oracle Corporation. Disponible en: http://members.fortunecity.com/rhubarb404/orahtml/oramag/bld\_dm.html [Consulta: 2005, febrero 14]
- Makro. (s.f.) Quienes Somos. [Página Web en línea]. Disponible: http://www.makro.com.ve/makro\_perfil.htm [Consulta: 2005, junio 26]
- Makro. (s.f.) Makro Company Profil. [Página Web en línea] Disponible en: http://www.makro-sa.com/southamerica/makro.htm [Consulta: 2005, junio 26]
- Makro. (s.f.) Makro Mission Statement [Página Web en línea] Disponible en: http://www.makro-sa.com/southamerica/makro.htm. [Consulta: 2005, marzo 30]
- Makro. (s.f.) Makro Formula. [Página Web en línea] Disponible en: http://www.makro-sa.com/southamerica/makroformula.htm. [Consulta: 2005, marzo 30]
- Microstrategy. (2005). Microstrategy 8. [Documento en línea]. Disponible en: http://www.microstrategy.com.ar/Software/. [Consulta 2005, octubre 08].
- Nigro, Oscar. (2005) [Página Web en Línea]. Introducción al Data Warehouse, Olap y Data Mining. España: Universidad Nacional del Centro de la Pcia de Bs. As. Facultad de Ciencias Exactas. Disponible en: http://www.exa.unicen.edu.ar/catedras/dwhouse/ [Consulta: 2005, junio 12]

Oliver, A. y Stampoultzis, G. (2005). The Apache Jakarta Project [Página Web en línea]. Disponible en: http://jakarta.apache.org/poi/trans/es/index.html [Consultada: 2005, agosto 20].

Oracle. Oracle 8i Utilitis. (s.f.). [Página Web en Línea]. Disponible en: http://www.csee.umbc.edu/help/oracle8/server.815/a67792/ch05.htm#2511 [Consulta 2005, agosto 10].

Oracle Oracle 9i Database Utilitis. (s.f.). [Página Web en línea]. Disponible en:

http://www.lc.leidenuniv.nl/awcourse/oracle/server.920/a96652/ch03.htm [Consulta 2005, agosto 10]

Oracle. (2004). ORACLE INFORMATION ARCHITECTURE SERVICES.

Oracle Discover for Business Intelligence. Executive Discovery Workshop.

[Página Web en Línea]. Disponible en:

http://www.oracle.com/consulting/technology/business-intelligence.html.

[Consulta: 2005, octubre 04].

Ramírez, Luz M. (2004). [Página Web en Línea]. Disponible en: http://www.enterate.unam.mx/Articulos/2004/octubre/inteligen.htm 10-11-05. [Consulta: 2005, julio 12]

Sánchez M, R. (s.f.). Bussines Inteligence... To BI or not to BI. México. [Página Web en Línea] Disponible en: http://www.monografias.com/trabajos14/bi/bi.shtml [Consulta: 2005, agosto 18].

Universidad de Antioquia. (2004). Proyecto Data Mart [Página Web en Línea]. Colombia: Medellín. Disponible en: http://planeacion.udea.edu.co/datamart/definicion.htm [Consulta: 2005, febrero 14]

Universidad de Antioquia. (2004). Proyecto Data Mart [Página Web en Línea]. Colombia: Medellín. Disponible en: http://planeacion.udea.edu.co/datamart/requerimientos.htm [Consulta: 2005, febrero 14]

Vela L., Mónica S. (s.f.) Sistemas de soporte a la decisión: una ventaja competitiva al alcance de todo tipo de empresas [Página Web en Línea]. España. Disponible en: http://www.gestiopolis.com/canales2/gerencia/1/ssdmonica.htm [Consulta: 2005, septiembre 28].

#### Revistas electrónicas:

Gómez. Decisiones inteligentes de negocios. PC World Venezuela [En línea].

Disponible en: http://www.pcworld.com.ve/n69/articulos/informe1.html
[2005,12 de julio].