



UNIVERSIDAD DEL AZUAY

FACTULAD DE CIENCIAS DE LA ADMINISTRACION

ESCUELA DE INGENIERIA DE SISTEMAS

**“CREACION DE UN DATA WAREHOUSE PARA LA EMPRESA FIBRO ACERO
S.A.”**

TESIS PREVIA A LA OBTENCION DEL TITULO DE INGENIEROS DE SISTEMAS

AUTORES:

VALERY MAGALY RODAS ORELLANA

CARLOS SEBASTIAN CALLE CEVALLOS

DIRECTOR:

ING. OSWALDO MERCHAN

CUENCA – ECUADOR

2008

AUTORIA

Los autores son los únicos responsables de los conceptos, conclusiones y observaciones emitidas en la presente Tesis.

Válery M. Rodas O.

Carlos S. Calle C.

AGRADECIMIENTO

Esta tesis es una muestra de la culminación de un gran paso en mi carrera, es por esto que quiero agradecer a muchas personas que por su amistad y apoyo me han ayudado de una u otra manera a terminarla.

A mis padres, Rodrigo y Elva, por todo el esfuerzo y confianza depositados en mí, que además de haberme dado la mejor educación profesional, me han dado el mejor crecimiento como ser humano. A mis hermanas Jéssica y Leydy, y a mis amigos por toda la amistad y ánimos brindados.

Al Ing. Oswaldo Merchán quien fue una guía y apoyo fundamental para la realización de este proyecto.

Al Ing. Geovany Vintimilla y al Ing. Demetrio Toledo, que a más de ser grandes amigos nos dieron su mano incondicionalmente en todo momento. A la empresa Fibro Acero S.A. por confiar en nuestro proyecto y facilitarnos todo lo necesario para culminar esta tesis.

A mi compañero de tesis Carlos Calle, quien puso todo su esfuerzo y conocimientos y fue una gran compañía tanto profesional como personal para realizar este importante proyecto.

Y por sobre todo a Dios, quien me ha dado el impulso necesario día a día para salir adelante en todo lo que me proponga.

Válery M. Rodas O.

AGRADECIMIENTO

Para la elaboración de esta tesis se necesitó de la ayuda de muchas personas a las cuales quisiera agradecer:

A Dios por darme salud y vida durante el desarrollo de este proyecto de tesis.

A mis padres: Iván y Marina, mis hermanos: Fernando Y Gina, a mi cuñada: Fabiola y a mi sobrino Nicolás, les agradezco por estar siempre a mi lado, brindarme su apoyo incondicional, alegrar mi vida y confiar siempre en mí.

A mi director de tesis: Ing. Oswaldo Merchán, gracias por aportar sus conocimientos y guiarme de manera correcta durante el desarrollo de esta tesis.

A mi compañera de tesis: Ing. Válerly Rodas, persona muy importante en mi vida. Gracias por ser quien me daba ánimos de seguir adelante en mis momentos de cansancio y debilidad, de aportar con grandes ideas y soluciones durante el desarrollo de esta tesis y por siempre confiar en mí. Gracias por todo, ya que sin ti esta tesis no hubiera podido ser realizada.

A los Ing.: Demetrio Toledo y Geovanny Vintimilla, participantes del desarrollo de esta tesis, les agradezco por compartir sus conocimientos no solo sobre este tema, sino en diferentes ámbitos los cuales nos permitieron realizar de una mejor manera este proyecto.

Y por último pero no el menos importante, un agradecimiento especial a la empresa Fibro Acero S.A., dirigida por el Eco. Marcelo Cordero. Gracias por abrirme las puertas de su empresa y facilitarme la documentación y recursos necesarios para la finalización de esta tesis.

Carlos S. Calle C.

Dedicatoria

Esta tesis la dedico a mis padres que gracias a su ejemplo y dedicación e logrado ir cumpliendo mis metas, siempre tomando en cuenta los valores que me han enseñado como la sencillez y humildad, a Dios por ser el pilar fundamental en mi vida, a mis hermanas y a mi compañero de tesis por todo el apoyo y confianza brindados.

Válery M. Rodas O.

Dedicatoria

Esta tesis la dedico a Dios por darme la vida durante del desarrollo de este proyecto, a mi familia por darme el soporte económico y moral durante el transcurso de mis estudios y siempre confiar en mi; y a mi compañera de tesis por ser parte fundamental no solamente en el desarrollo de esta tesis, si no a lo largo de mi formación profesional. Gracias a todos.

Carlos S. Calle C.

RESUMEN

Un Data Warehouse es un almacén de datos que permite obtener información confiable y valiosa para la toma de decisiones, es por esto que cada día más empresas tienen la necesidad de implementarlo debido a la globalización y competencia existente en el mercado.

Siendo Fibro Acero S.A. una empresa que se encuentra en el mercado por mas de quince años como fabricante y comercializador de cocinas, cocinetas y cilindros de gas, es de gran utilidad implementar esta herramienta para facilitar la toma de decisiones y así permitir a los gerentes mejorar sus estrategias. Para desarrollar el Data Warehouse utilizaremos la herramienta case Genexus en la creación y administración, Gxplorer el cual es un complemento OLAP de Microsoft Excel permitirá a los usuarios explotar la información, y SQL Server 2000 como base de datos.

ABSTRACT

A Data Warehouse is a data store that allows you to obtain reliable and valuable information in order to make decisions. Every day, more businesses find the need to implement this system due to globalization and competition in the marketplace.

Fibro Acero S.A. has been in the market for more than fifteen years as a manufacturer and retailer of kitchens, kitchenettes and gas cylinders. Implementing this system will be very useful in order to facilitate decision-making, and in this way permit managers to improve their strategies. In the development of the Data Warehouse, we will use Genexus in its creation and administration, SQL Server 2000 as the database and Explorer, a complement of Microsoft OLAP, which will permit users to exploit the information.



INDICE DE CONTENIDOS

Resumen.....	i
Abstract.....	ii
1. INTRODUCCION.....	3
2. ASPECTOS TEORICOS.....	6
2.1. Introducción	6
2.2. Sistemas de información	7
2.2.1. Tipos de sistemas	8
2.3. Data Warehouse	9
2.3.1. Características.....	10
2.3.2. Estructura de un data warehouse.....	12
2.3.3. Arquitectura de un data warehouse.....	14
2.3.4. Usos, ventajas y desventajas de un data warehouse.....	17
2.3.5. Ejemplos de un data warehouse.....	19
2.4. Modelo de datos.....	22
2.4.1. Modelo relacional.....	22
2.4.2. Modelo Multidimensional	23
2.5. Conclusiones	29
3. HERRAMIENTAS PARA LA CONSTRUCCIÓN, ADMINISTRACIÓN Y MANIPULACIÓN DE UN DATA WAREHOUSE	31
3.1. Introducción	31
3.2. Herramientas para la transformación del Data Warehouse	32
3.2.1. GeneXus y el Data Warehouse	32
3.3. Herramientas para la administración del data warehouse.....	33
3.3.1. Clasificación y características	34
3.3.2. Microsoft SQL Server 2000 y el Data Warehouse.....	36
3.4. Herramientas para manipular la información del data warehouse	38
3.4.1. Clasificación y características	38
3.5. Conclusiones:	41
4. METODOLOGIA DE DESARROLLO PARA LA CONSTRUCCION E IMPLEMENTACION DEL DATA WAREHOUSE.....	44
4.1. Introducción	44
4.2. Fases para la construcción del Data Warehouse	45
4.2.1. Recogida y Análisis de Requisitos.....	46

4.2.2.	Diseño Conceptual.....	51
4.2.3.	Diseño Lógico Específico.....	62
4.2.4.	Diseño Lógico General	78
4.2.5.	Implementación.....	79
4.3.	Conclusiones	96
5.	PRUEBAS SOBRE UN DATA WAREHOUSE	98
5.1.	Introducción	98
5.2.	Tipos de pruebas y características	99
5.2.1.	Prueba de Integración	99
5.2.2.	Prueba de Caja Negra	100
5.2.3.	Prueba de Caja Blanca.....	100
5.2.4.	Prueba de Aceptación	101
5.2.5.	Prueba de Rendimiento	101
5.2.6.	Prueba de Robustez	101
5.2.7.	Prueba de Resistencia	102
5.2.8.	Prueba de Transformación.....	102
5.3.	Ventajas y desventajas	102
5.4.	Conclusión	106
6.	CONCLUSIONES.....	109
7.	BIBLIOGRAFIA.....	112
	ANEXOS.....	115

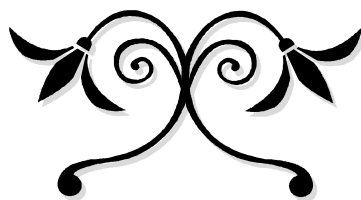
INDICE DE FIGURAS

Figura No. 2.1: "Procesos de un Sistema de Información"	8
Figura No. 2.2: "Niveles de la Organización"	8
Figura No. 2.3: "Data Warehouse Integrado"	11
Figura No. 2.4: "Data Warehouse Orientado a Temas"	11
Figura No. 2.5: "Data Warehouse no Volátil"	12
Figura No. 2.6: "Estructura de un Data Warehouse"	13
Figura No. 2.7: "Arquitectura real del Data Warehouse"	15
Figura No. 2.8: "Concurrencia en el Data Warehouse"	18
Figura No. 2.9: "Enfoque anticuado de procesamiento de la Información"	20
Figura No. 2.10: "Esquema Estrella"	24
Figura No. 2.11: "Esquema Copo de Nieve"	25
Figura No. 2.12: "Constelación de hechos"	25
Figura No. 2.13: "Ejemplo de un Cubo de tres dimensiones"	26
Figura No. 2.14: "Slice & Dice"	26
Figura No. 2.15: "Rotación"	27
Figura No. 2.16: "Drill – Up y Drill - Down"	27
Figura No. 2.17: "Roll - Up"	28
Figura No. 2.18: "Drill - Across"	28
Figura No. 2.19: "Drill - Through"	28
Figura No. 3.1: "Etapas del Data Warehouse - Transformación"	32
Figura No. 4.1: "Fases para la construcción de un DWH"	45
Figura No. 4.2: "Diagrama Entidad Relación del Sistema EcoSAC"	50
Figura No. 4.3: "Diagr. Entidad Relación de la base de datos Ventas"	51
Figura No. 4.4: "Notación Gráfica en CMDM del nivel Empleado"	52
Figura No. 4.5: "Notación Gráfica en CMDM de la Dimensión Empleado"	53
Figura No. 4.6: "Notación Gráfica en CMDM de la Relación Dimensional Faltas y Permisos"	54
Figura No. 4.7: "Representación Gráfica Dimensión Fecha"	54
Figura No. 4.8: "Representación Gráfica Dimensión Sucursal"	54
Figura No. 4.9: "Representación Gráfica Dimensión Cliente"	55
Figura No. 4.10: "Representación Gráfica Dimensión Producto"	55
Figura No. 4.11: "Representación Gráfica Dimensión Documentos"	55
Figura No. 4.12: "Representación Gráfica Dimensión Total Ventas"	56
Figura No. 4.13: "Representación Gráfica Relación Dimensional Ventas"	56
Figura No. 4.14: "Representación Gráfica Dimensión Tiempo"	56
Figura No. 4.15: "Representación Gráfica Dimensión Clasificación"	57

Figura No. 4.16: "Representación Gráfica Dimensión Proveedor"	57
Figura No. 4.17: "Representación Gráfica Dimensión Tipo Calificación"	57
Figura No. 4.18: "Representación Gráfica Dimensión Tipo Calificación"	57
Figura No. 4.19: "Representación Gráfica de la Relación Dimensional Calificación Proveedor"	58
Figura No. 4.20: "Representación Gráfica Dimensión Fecha"	58
Figura No. 4.21: "Representación Gráfica Dimensión Motivo"	58
Figura No. 4.22: "Representación Gráfica Dimensión Empleado"	59
Figura No. 4.23: "Representación Gráfica Dimensión Estado"	59
Figura No. 4.24: "Representación Gráfica Dimensión Permisos"	60
Figura No. 4.25: "Representación Gráfica Relación Dimensional Faltas y Permisos" ..	60
Figura No. 4.26: "Representación Gráfica Dimensión Fecha"	60
Figura No. 4.27: "Representación Gráfica Dimensión Empleado"	61
Figura No. 4.28: "Representación Gráfica Dimensión Remuneración"	61
Figura No. 4.29: "Representación Gráfica Relación Dimensional Remuneraciones" ..	62
Figura No. 4.30: "Materialización de la Relación Ventas, Cubo 1"	64
Figura No. 4.31: "Materialización de la Relación Calificación Proveedor, Cubo 1" ..	64
Figura No. 4.32: "Materialización de la Relación Faltas y Permisos, Cubo 1"	65
Figura No. 4.33: "Materialización de la Relación Remuneraciones, Cubo 1"	65
Figura No. 4.34: "Fragmentación Vertical de Dimensiones – Req. Dep. Ventas"	67
Figura No. 4.35: "Fragmentación Vertical de Dimensiones – Req. Dep. Logística" ...	68
Figura No. 4.36: "Fragmentación Vertical de Dimensiones – Req. Faltas y Permisos" ..	69
Figura No. 4.37: "Fragmentación Vertical de Dimensiones – Req. Remuneraciones" ..	70
Figura No. 4.38: "Mapeo de la Dimensión Empleado. Dep. Talento Humano"	72
Figura No. 4.39: "Mapeo de la Dimensión Producto. Dep. Ventas"	73
Figura No. 4.40: "Mapeo de Cubos Faltas y Permisos. Dep. Talento Humano"	73
Figura No. 4.41: "Mapeo de Cubos Remuneraciones. Dep. Talento Humano"	74
Figura No. 4.42: "Mapeo de Cubos Ventas. Dep. Ventas"	74
Figura No. 4.43: "Mapeo de Cubos Logística. Dep. Logística"	75
Figura No. 4.44: "Esquema Lógico del DWH de Talento Humano"	75
Figura No. 4.45: "Modelo Multidimensional en Estrella, Tabla de hecho Permisos"	76
Figura No. 4.46: "Modelo Multidimensional en Estrella, Tabla de hecho Remuneraciones"	77
Figura No. 4.47: "Esquema Lógico del DWH del departamento de Ventas"	77
Figura No. 4.48: "Esquema Lógico del DWH del departamento de Logística"	78
Figura No. 4.49: "Elección del objeto Dimensión o Medida"	80
Figura No. 4.50: "Estructura de la Dimensión Departamentos"	80
Figura No. 4.51: "Estructura de la Medida Permisos"	81

Figura No. 4.52: "Objeto transacción para el Data View y tabla externa"	82
Figura No. 4.53: "Creación y estructura de un Data View"	82
Figura No. 4.54: "Creación y especificación de un Data Store"	83
Figura No. 4.55: "Algoritmo para la importación de información desde un archivo de Excel"	83
Figura No. 4.56: "Menú principal de la aplicación del DWH"	84
Figura No. 4.57: "Pantalla para el ingreso de las fechas de carga al DWH"	84
Figura No. 4.58: "Información del DWH"	85
Figura No. 4.59: "Creación del Metadata"	86
Figura No. 4.60: "Asignación de Nombre y Ubicación del Metadata"	87
Figura No. 4.61: "Menú de carga del Metadata"	88
Figura No. 4.62: "Elección del paradigma"	88
Figura No. 4.63: "Elección de la ubicación de la base de conocimiento"	89
Figura No. 4.64: "Elección de modelo de la base de conocimiento Genexus"	89
Figura No. 4.65: "Creación y asignación del ODBC para el Metadata"	90
Figura No. 4.66: "Visualización de la Metadata con sus dimensiones y medidas cargadas"	90
Figura No. 4.67: "Ingreso al modulo Gxplorer Settings"	91
Figura No. 4.68: "Creación de un nuevo usuario"	91
Figura No. 4.69: "Ubicación de la barra de Gxplorer Olap para MSEXcel"	92
Figura No. 4.70: "Nueva Consulta Gxplorer"	92
Figura No. 4.71: "Selección de las dimensiones y medidas a consultar"	93
Figura No. 4.72: "Resultado consulta con Gxplorer Olap para MSEXcel"	94
Figura No. 4.73: "Grafico explicativo de los resultados de la consulta DWH"	95
Figura No. 4.74: "Materialización de la Relación Ventas, Cubo 2"	115
Figura No. 4.75: "Materialización de la Relación Ventas, Cubo 3"	116
Figura No. 4.76: "Materialización de la Relación Calificación Proveedor, Cubo 2"	116
Figura No. 4.77: "Materialización de la Relación Faltas y Permisos, Cubo 2"	117
Figura No. 4.78: "Materialización de la Relación Faltas y Permisos, Cubo 3"	117
Figura No. 4.79: "Mapeo de la Dimensión Motivo. Dep. Talento Humano"	117
Figura No. 4.80: "Mapeo de la Dimensión Estado. Dep. Talento Humano"	118
Figura No. 4.81: "Mapeo de la Dimensión Fecha. Dep. Talento Humano"	118
Figura No. 4.82: "Mapeo de la Dimensión Fecha. Dep. Ventas"	119
Figura No. 4.83: "Mapeo de la Dimensión Cliente. Dep. Ventas"	119
Figura No. 4.84: "Mapeo de la Dimensión Sucursal. Dep. Ventas"	119
Figura No. 4.85: "Mapeo de la Dimensión Documentos. Dep. Ventas"	120
Figura No. 4.86: "Drill-up del Cubo Faltas1. Dep. Talento Humano"	120
Figura No. 4.87: "Drill-up del Cubo Faltas2. Dep. Talento Humano"	120

Figura No. 4.88: "Drill-up del Cubo Faltas2. Dep. Talento Humano"	121
Figura No. 4.89: "Drill-up del Cubo Ventas1. Dep. Ventas"	121
Figura No. 4.90: "Drill-up del Cubo Ventas2. Dep. Ventas"	121
Figura No. 4.91: "Drill-up del Cubo Ventas3. Dep. Ventas"	121
Figura No. 4.92: "Drill-up del Cubo Calificación Proveedor1"	122
Figura No. 4.93: "Drill-up del Cubo Calificación Proveedor2"	122



CAPITULO 1

INTRODUCCION

1. INTRODUCCION

En un mundo cada vez más competitivo en donde las empresas sienten la necesidad de crear estrategias que les permita mantenerse y sobresalir en el mercado, es de vital importancia contar con información precisa para tal fin. Hoy en día las empresas cuentan con sistemas transaccionales que almacenan las actividades diarias del negocio, pero no permiten obtener de manera rápida, resumida y objetiva la información que se necesita. Es por esto que surge un nuevo concepto para almacenar y analizar la información llamado Data Warehouse, que de ahora en adelante nos referiremos como DWH, el mismo que consolida los datos provenientes de diferentes bases de datos operacionales y los hace disponibles para la realización de análisis de tipo gerencial.

La información que almacenan los sistemas operacionales podría darnos los informes requeridos, pero para hacerlo se necesitaría que los usuarios sean expertos en lenguajes de consulta como SQL, además, al consultar grandes cantidades de datos el rendimiento del sistema puede verse perjudicado; es por esto que al momento de crear nuestro data warehouse se debe integrar toda la información relevante para apoyar en la toma de decisiones de una manera eficiente.

Para lograrlo es preciso realizar un análisis de lo que se desea construir, en este proyecto nos enfocamos en un análisis multidimensional, que en base a los requerimientos de los usuarios se identifican las dimensiones, medidas y hechos para representar la realidad, además de esto se requiere un modelo a seguir, pero al no existir todavía una metodología para la construcción de un DWH, en este trabajo se expondrá una metodología transaccional pero con variantes únicas asociada al data warehouse.

A lo largo de esta tesis podremos observar como se crea un data warehouse que cumpla con las necesidades de tres departamentos de Fibro Acero S.A.: ventas, logística y talento humano.

Para la construcción hemos utilizado la herramienta Genexus, con la cual, extraeremos la información por medio de objetos Data Views, construiremos el DWH ya que esta herramienta permite generar la estructura dimensional creando las medidas y dimensiones, y cargaremos la información al DWH con aplicaciones desarrolladas para tal fin.

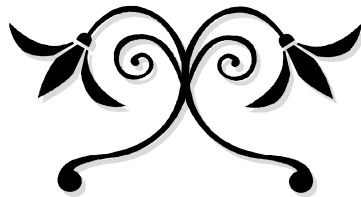
Como almacén de datos utilizaremos SQL Server 2000 y para explorar la información resultante de nuestro DWH utilizaremos Gxplorer Olap que lo encontraremos como complemento de Microsoft Excel lo cual es muy familiar para los usuarios.

En el capítulo 2 revisaremos de forma global lo que es un data warehouse, analizaremos su arquitectura y características, además nos adentraremos en los fundamentos del modelo multidimensional.

En el capítulo 3 analizaremos Genexus, Gxplorer, SQL Server 2000 además de otras herramientas existentes para desarrollar y explotar un DWH.

En el capítulo 4 desarrollaremos el DWH para la empresa Fibro Acero S. A., expandiendo la metodología que utilizamos y exponiendo como se utilizaron las herramientas en cada etapa del proceso.

En el capítulo 5 exponemos ciertas pruebas que pueden realizarse durante la implementación del DWH.



CAPITULO 2

ASPECTOS TEORICOS

2. ASPECTOS TEORICOS

2.1. Introducción

La competencia a nivel empresarial es cada vez mayor, y la necesidad de contar con información valiosa para ayudar a los directivos de la empresa a tomar decisiones de nivel estratégico crece aun más, es por esto que las empresas cuentan con sistemas transaccionales que les permiten procesar y almacenar los datos resultantes de la operación diaria del negocio, pero no les permite obtener información resumida para toma de decisiones, es por esto la importancia de crear un DWH.

En este capitulo nos introduciremos en los conceptos y estructura de un DWH, veremos desde información básica sobre los diferentes sistemas hasta adentrarnos mas y poder explicar mejor las características, arquitectura, herramientas front end, y los tipos de modelos de datos.

Como la mayor parte de información que conforma un data warehouse, proviene de los sistemas transaccionales, es necesario hacer referencia a la estructura de la organización con respecto a los sistemas de información.

2.2. Sistemas de información

Es un conjunto de elementos que interactúan entre sí, con la finalidad de apoyar a las actividades de una empresa.

La tarea principal de los sistemas de información es cumplir 3 objetivos básicos:

- Automatización de procesos operativos.
- Proporcionar información para ayudar en la toma de decisiones.
- Lograr ventajas competitivas gracias a su implementación y uso.

Estos sistemas realizan cuatro procedimientos, como se muestra en la figura 2.1, conocidos como: Entrada, Almacenamiento, Procesamiento y Salida de información.

- Entrada de información: es el proceso por el cual el sistema obtiene la información necesaria para su funcionamiento. Estas entradas pueden ser manuales, como teclados, scanners, etc., o automáticas, provenientes de sistemas externos o módulos.
- Almacenamiento de información: esta actividad es la más importante, ya que permite guardar los datos e información en discos duros, CD, etc., para luego ser reutilizada según las necesidades de la empresa.
- Procesamiento de información: es la capacidad de transformar los datos en información, basándose en cálculos o tareas preestablecidas, y así lograr conocer el estado de la empresa y en ciertos casos sirve de apoyo en la toma de decisiones.
- Salida de información: permite extraer la información procesada al exterior. La manera más común es a través de impresoras, discos extraíbles, etc. En algunos casos pueden ser entradas de información para otros sistemas.

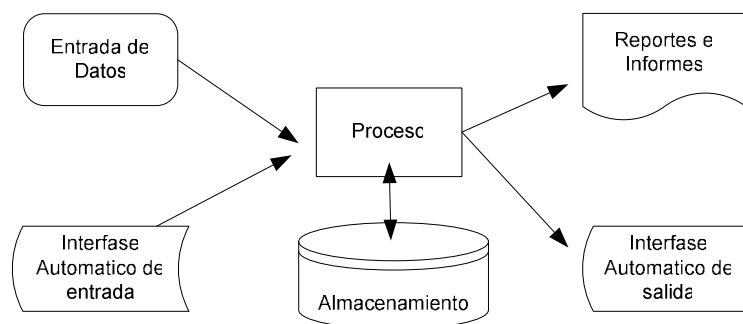


Figura No. 2.1: “Procesos de un Sistema de Información”

2.2.1. Tipos de sistemas

Existen cuatro tipos de sistemas, y se dividen de acuerdo al siguiente esquema:¹

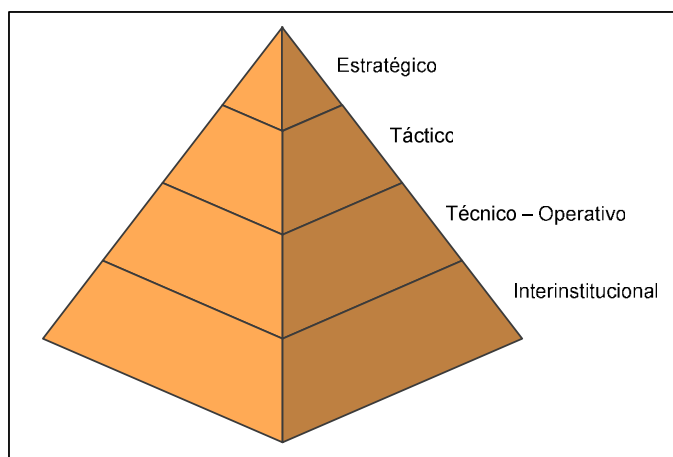


Figura No. 2.2: “Niveles de la Organización”

Sistemas estratégicos: Su objetivo es brindar el apoyo necesario para la toma de decisiones dentro de las empresas, mediante información filtrada y relevante. Su característica principal es la carga periódica de trabajo, es decir que su utilización no es predecible. Dentro de esta categoría tenemos los Sistemas de Información Gerencial (MIS), Sistemas de Información Ejecutivos (EIS), Sistemas de Información Georeferencial (GIS), Sistemas de Simulación de Negocios (BIS y que en la práctica son sistemas expertos o de Inteligencia Artificial - AI).

¹ Instituto Nacional de Estadística e Informática, 1997. Manual para la construcción de un Data Warehouse. <http://www.inei.gob.pe/web/metodologias/attach/lib619/INDEX.HTM>. [consulta 20 de noviembre de 2006]

Sistemas tácticos: Estos sistemas están diseñados para coordinar y manejar la documentación de una empresa. Su objetivo es facilitar consultas de la información almacenada en el sistema y la generación de informes. Entre los sistemas mas conocidos se encuentran los Sistemas de Transmisión de Mensajería (correo electrónico y servidor de fax), Control y coordinación de tareas (Work Flow), Sistemas Ofimáticos (OA), y tratamiento de documentos (Imagen, Trámite y Bases de Datos Documentarios).

Sistemas técnico-operativos: Son los sistemas tradicionales, utilizados para el ingreso masivo de información al sistema (contabilidad, facturación, control de personal, etc.). En la actualidad este tipo de sistemas tratan de mejorar su funcionalidad mediante sensores, bases de datos relacionales, data warehousing, etc.

Sistemas interinstitucionales: Es un nuevo tipo de sistema que está en constante crecimiento, y tiene como objetivo, crear una conexión entre organizaciones y el mercado mundial, por medio de redes informáticas de alcance nacional y global (Internet), sin importar donde se encuentren las organizaciones (intranet) ni en que mercado se encuentren (extranet).

2.3. Data Warehouse

El Data Warehouse surge por dos razones: primero, la necesidad de proporcionar una fuente única de datos limpia y consistente para propósitos de apoyo para la toma de decisiones, y segundo, la necesidad de hacerlo sin afectar a los sistemas operacionales.

Es por esto que se ha desarrollado una arquitectura sólida que permite procesar la información y poner su uso estratégico como generador de ventajas competitivas. Esta arquitectura es el Data Warehouse (en lo adelante nos referiremos como DWH) que nace a inicio de los 90. Para entender mejor su concepto, daremos la definición clásica de Bill Inmon, reconocido como el padre del DWH.

"Un Data Warehouse es una colección de datos orientados a un tema, integrados, no volátiles, e historizados, organizados para dar soporte al proceso de ayuda a la toma de decisiones" ²

Con esta definición simple y concreta se clarifica el concepto de un DWH, y si la profundizamos, podríamos decir que algunos de sus objetivos son:

- Comprender las necesidades de los usuarios por áreas dentro del negocio.
- Determinar qué decisiones se pueden tomar con la ayuda del DWH
- Seleccionar un subconjunto del sistema de fuentes de datos que sea el más efectivo y procesable para presentar el DWH.
- Asegurar que los datos sean precisos, correctos y confiables, y que mantengan la consistencia.
- Monitorear continuamente la precisión y exactitud de los datos y el contenido de los reportes generados.
- Publicar los datos.

2.3.1. Características

Siguiendo con la definición de Inmon, un DWH se caracteriza por ser:

2.3.1.1. Integrado

Un DWH recoge los datos de diferentes sistemas operacionales de la organización y/o fuentes externas, como se muestra en la Figura No. 2.3, por lo tanto el aspecto más importante del ambiente data warehousing es que la información encontrada en su interior este siempre integrada, para que al utilizarla, los usuarios se enfoquen en usar los datos y no preguntarse sobre la confiabilidad o consistencia de los mismos.

² Inmon, W.H. "Building the Data Warehouse". Tercera Edición, Jhon Wiley & Sons, 2002. [consulta 26 de octubre de 2006]

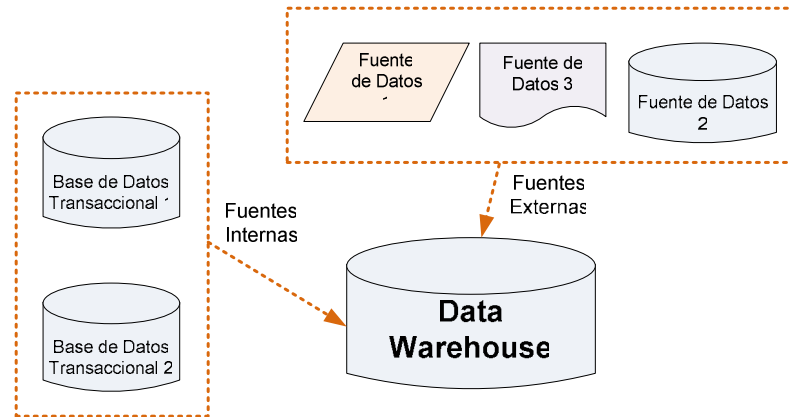


Figura No. 2.3: "Data Warehouse Integrado"

Cuando los datos se mueven de los sistemas operacionales al DWH, experimentan una transformación fundamental antes de entrar al depósito, y ésta integración se muestra de muchas maneras como: nomenclatura, medidas para las variables, formas de codificar, entre otros.

2.3.1.2. Orientado a temas

El DWH excluye la información que no será utilizada por el proceso de sistemas de soporte de decisiones, y la clasifica en base a los aspectos que son de interés para la empresa. Por ejemplo, consolidar toda la información de los vendedores en una sola tabla para que los usuarios puedan acceder más fácilmente.



Figura No. 2.4: "Data Warehouse Orientado a Temas"

2.3.1.3. De tiempo variante

La información del DWH es requerida en cualquier momento, a diferencia de los sistemas operacionales, que se la necesita al momento de acceder a ellos. Por esta característica, los datos encontrados en el depósito se llaman

de “tiempo variante”. Este tiempo variante se muestra de algunas maneras como:

- El horizonte largo de tiempo, de 5 a 10 años.
- Cada estructura clave del DWH contiene un elemento de tiempo como día, semana, mes, etc.
- Una vez almacenada la información en el DWH no puede ser actualizada.

2.3.1.4. No volátil

Para realizar la toma y análisis de decisiones, es necesaria una base de datos estable, y eso es lo que nos brinda un DWH, en este solo se realizan dos operaciones fundamentales, la carga inicial de los datos y el acceso a los mismos, a diferencia del ambiente operacional donde existe el ingreso, modificación y eliminación como se muestra en la figura No. 2.5.

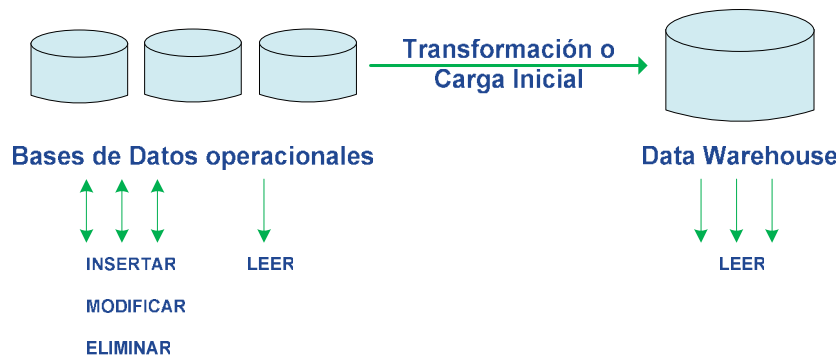


Figura No. 2.5: “Data Warehouse no Volátil”

2.3.2. Estructura de un data warehouse

La estructura que conforma un DWH es diferente a la de una base de datos normal, como observamos en la figura 2.6, en el DWH encontramos niveles y cada una de ellos tiene un diferente grado de detalle, esto nos permite conocer los límites que posee el almacén de datos. Los niveles anteriormente mencionados son los siguientes:³

³ Instituto Nacional de Estadística e Informática, 1997. Manual para la construcción de un Data Warehouse. <http://www.inei.gob.pe/web/metodologias/attach/lib619/INDEX.HTM>. [consulta 10 de diciembre de 2007]

- Detalle de datos actuales.
- Detalle de datos antiguos.
- Datos ligeramente resumidos.
- Datos completamente resumidos.
- Meta data

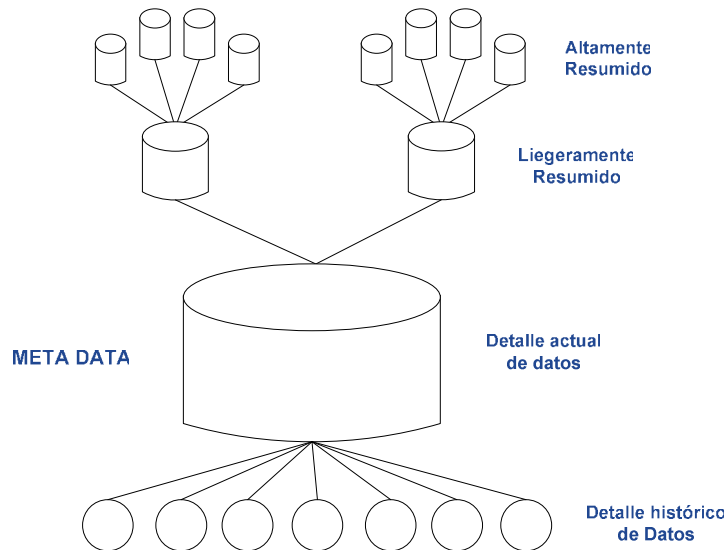


Figura No. 2.6: "Estructura de un Data Warehouse" ³

Detalle de datos actuales: En este nivel se encuentra almacenada la información más detallada de un DWH, que es obtenida desde las bases de datos operacionales y es considerada la más importante debido a que:

- Contiene la información más reciente y ésta es de gran interés.
- Mantiene su información con un gran detalle.
- Es de fácil acceso y su almacenamiento se lo hace en disco.

Detalle de datos antiguo: Como su nombre lo dice, en este nivel se almacena la información histórica del DWH y se lo hace de manera detallada, consistente con el detalle de los datos actuales. Su acceso no es frecuente y el método de almacenamiento depende de la cantidad de información que se tenga, esta puede ser almacenada en discos locales, discos alternos, etc.

Datos ligeramente resumidos: Comúnmente este nivel es almacenado en disco y su información proviene de los datos actuales. El grado de resumen que este presenta, depende del diseño establecido y se basa en dos factores: la unidad de tiempo con la cual está estructurado y el contenido o atributos que se tiene.

Datos completamente resumidos: Este nivel contiene información que es compacta y de fácil acceso, la cual es adquirida del nivel de datos ligeramente resumidos.

Estos dos últimos niveles son los más accedidos por los usuarios, ya que contienen la información relevante de la empresa y la más adecuada para el soporte en la toma de decisiones.

Meta data: Es el componente final de un DWH, el cual se encuentra en una dimensión diferente, por que su información es diferente a la almacenada en otros niveles y no son tomados de los sistemas operacionales. Aquí se encuentra:

- Un directorio para ayudar al analista a ubicar los contenidos del DWH.
- Una guía que indica la manera en la que los datos se transforman desde los sistemas operacionales al DWH.
- Indica las operaciones que se realizan para transformar, la información del detalle actual de datos, a los datos ligeramente resumidos y estos a su vez, a los datos completamente resumidos.

2.3.3. Arquitectura de un data warehouse

La tecnología Data Warehouse representa de mejor manera la amplia estructura de una empresa, por lo tanto, para saber cómo se relacionan todos los componentes involucrados en el DWH, es necesario analizar su arquitectura, la misma que establece el marco de trabajo, estándares y procedimientos a un nivel empresarial. Los objetivos de las actividades de la arquitectura son simples, integrar al DW las necesidades de información empresarial.

2.3.3.1. Elementos que constituyen la arquitectura de un data warehouse

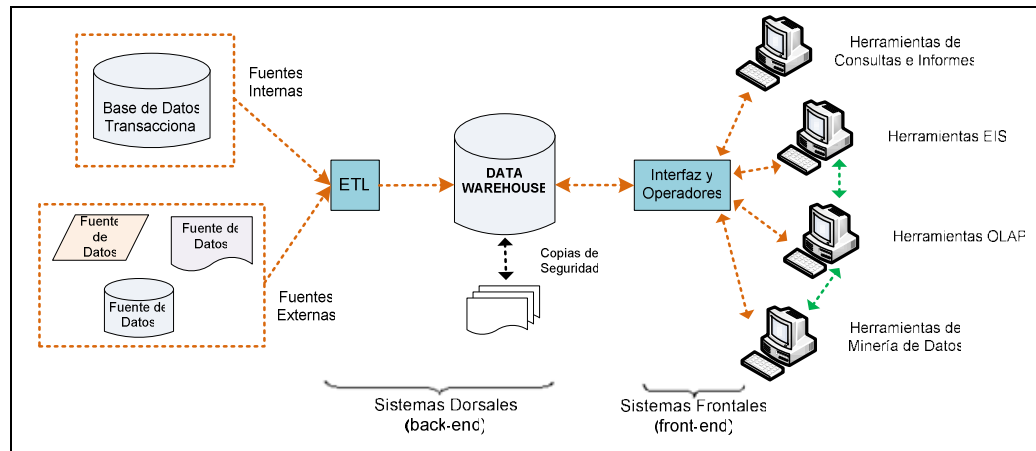


Figura No. 2.7: "Arquitectura real del Data Warehouse" ⁴

2.3.3.1.1 Fuentes de extracción de datos

Estas fuentes, tales como los sistemas OLTP, fuentes externas de datos, etc., son las que proveen toda la información que será almacenada en el DWH, sin antes haberla transformado e integrado.

OLTP (*On-Line Transaction Processing*) son aplicaciones que definen el comportamiento habitual de un entorno operacional de gestión, y ejecutan las operaciones del día a día.

2.3.3.1.2 Sistemas dorsales (back-end)

Son herramientas que se encargan de la extracción, limpieza, transformación, integración, carga y refresco de los datos de las fuentes externas. Dentro de la clasificación de sistemas dorsales, podemos mencionar los sistemas ETL y el mismo DWH, que se describen a continuación.

Sistemas ETL (*Extraction, Transformation, Load*): Realizan funciones como:

- Extracción de los datos desde las distintas fuentes de datos.
- Transformación o filtrado de los datos: limpieza, consolidación, etc.

⁴ Sepúlveda Castro, Jorge. Urrutia Suárez, Luis. "Diseño e Implementación de un Data Warehouse para la gestión de ventas de la empresa vitivinícola Miguel Torres Chile". Diciembre, 2004. [consulta 12 de marzo de 2007]

- Carga inicial del Data Warehouse: ordenación, agregaciones, etc.
- Refresco del Data Warehouse: es una operación periódica que propaga los cambios de las fuentes externas al Data Warehouse.

Data Warehouse: Posee información relevante de los datos o metadatos, estos metadatos, contienen información relativa a los datos y comúnmente, permiten mantener información sobre:

- La procedencia y frecuencia de refresco de los datos
- La fiabilidad y forma de cálculo de los datos
- Semántica de los datos y su localización en el DW, etc.

2.3.3.1.3 Sistemas frontales (front-end)

Los sistemas frontales o herramientas front-end, ofrecen al usuario final mecanismos de acceso a la información simples y potentes, permitiendo obtener una correcta y eficaz conexión con el DWH. Con estos sistemas se pueden realizar consultas, informes, análisis y extracción del conocimiento (data mining). Dentro de esta clasificación podemos mencionar las Interfaces y Gestores de Consultas, y los Sistemas de Integridad y Seguridad.

Las Interfaces y Gestores de Consulta permiten acceder a los datos y sobre ellos conectar herramientas más sofisticadas como OLAP (*Online Analytical Processing*, que lo revisaremos mas adelante), EIS, minería de datos (data mining).

2.3.3.1.4 Data Marts

A partir del depósito de datos integrado del DWH, se pueden crear los data marts. En síntesis, los data marts son repositorios de datos o pequeños DWH centrados en un tema o un área de negocio específico. En muchos casos, los Data Warehouse comienzan siendo Data Marts con el objetivo de minimizar los riesgos para luego ir ampliando su espectro gradualmente.

2.3.3.2. Tipos de Arquitectura

- Arquitectura real: (Figura No. 2.7) Es la más usada en los sistemas de soporte de decisión. La información es obtenida de extracciones

periódicas al sistema de producción y fuentes externas, se la integra, limpia y transforma para realizar la carga o almacenamiento en el DWH, que se lo hace en un SGBD (Sistema Gestor de Base de Datos) separado del sistema de producción de la empresa.

Mediante esta arquitectura, obtenemos datos que se ajustan a las necesidades del proceso de toma de decisiones.

- Arquitectura virtual: En esta arquitectura, la información del DWH es almacenada en el sistema tradicional, por lo que no existe un coste de almacenamiento suplementario, y el acceso a los datos se lo hace en tiempo real por medio de middlewares o pasarelas. Pero las desventajas que esta arquitectura presenta lo hacen poco utilizable, por ejemplo, los datos no son preparados y los accesos de decisión pueden entorpecer el funcionamiento y rendimiento del sistema de la empresa.
- Arquitectura remota: Es una combinación de las dos arquitecturas antes descritas. Trata de implementar físicamente los niveles agregados (los niveles de datos más utilizados) para conservar el nivel de detalle en los sistemas de producción, y facilitar el acceso que se lo hace a través de middleware o gateways.

2.3.4. Usos, ventajas y desventajas de un data warehouse

El acceso a un DWH o a una base de datos tradicional, depende del usuario y la información que necesite obtener. Un DWH esta orientado hacia las personas que necesitan hacer consultas complejas, de múltiples fuentes, que su respuesta no sea inmediata y en algunos casos que contenga información histórica.

En una empresa, los usuarios que utilizan las bases de datos operacionales necesitan que la información consultada tenga una respuesta inmediata y normalmente son consultas predefinidas, las cuales en muchos de los casos acceden a una sola base de datos para satisfacer la petición. Por el contrario, un DWH permite realizar cualquier tipo de consulta, según la

información que se necesite (detallada, ligeramente resumida o completamente resumida).

Un DWH tiene una característica notable, que es la concurrencia. Los recursos que se necesitan para poder realizar las consultas son altos, y mientras más usuarios accedan al DWH más recursos serán necesarios. Como muestra la figura 2.8, se debe distribuir de manera adecuada el acceso a los diferentes niveles del DWH, puesto que mientras más alto sea el nivel de acceso menos recursos se deberán utilizar, y la concurrencia podrá ser mayor.

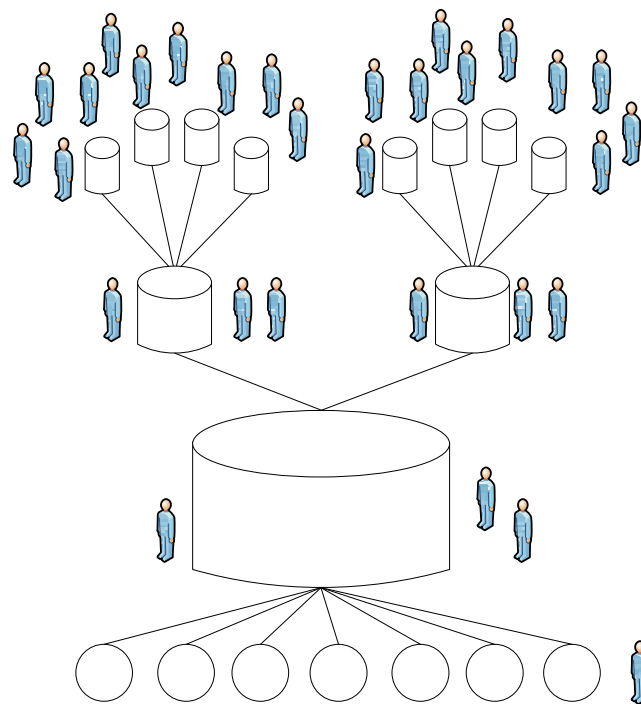


Figura No. 2.8: "Concurrencia en el Data Warehouse"

A continuación se describen algunas ventajas y desventajas que tiene un DWH.⁵

Ventajas:

- Gran poder de procesamiento de la información.
- Permite mayor flexibilidad y rapidez en el acceso de la información.

⁵ Asesoría, Diseño y Desarrollo de Aplicaciones
<http://www.navactiva.com/web/es/atic/aseso/desarrollo/asesor1/2005/32086.jsp>, [consulta 23 de noviembre de 2006]

- Facilita la toma de decisiones en los negocios.
- Las empresas obtienen un aumento en la productividad.
- Proporciona una comunicación fiable entre los departamentos de la empresa.
- Permite conocer lo que sucede en la empresa.
- Transforma los datos en información y la información en conocimiento.
- Reduce los tiempos de respuesta y los costes de operación.

Desventajas:

- Requiere una revisión del modelo de datos, objetos, transacciones y además del almacenamiento.
- Tiene un diseño complejo y multidisciplinar.
- Requiere una reestructuración de los sistemas operacionales.
- Tiene un alto coste.

2.3.5. Ejemplos de un data warehouse

Ejemplo 1: Preparación de un reporte complejo.

El informe incluye las finanzas actuales, el inventario y la condición de personal, acompañado de comparaciones del mes actual con el anterior y el mismo mes del año anterior, con una comparación adicional de los 3 años precedentes. Se debe explicar cada desviación de la tendencia que cae fuera de un rango predefinido.

Sin un data warehouse, el informe es preparado de la siguiente manera:

La información financiera actual se obtiene desde una base de datos mediante un programa de extracción de datos, el inventario actual de otro programa de extracción de otra base de datos, la condición actual de personal de un tercer programa de extracción y la información histórica desde un backup de cinta magnética o CD-ROM.

Lo más interesante es que se ha pedido otro informe que continúe al primer informe (debido a que las preguntas se originaron a partir del anterior). El hecho es, que ninguno de los trabajos realizados hasta aquí (por ejemplo,

diversos programas de extracción) se pueden usar para los próximos o para cualquier reporte subsiguiente. Imagine el tiempo y el esfuerzo que se ha desperdiciado por un enfoque anticuado. (Ver Figura 2.9).

Las inconsistencias deben identificarse en cada conjunto de datos extraídos y resolverse, por lo general, manualmente. Cuando se completa todo este procesamiento, el reporte puede ser formateado, impreso, revisado y transmitido.

Nuevamente, el punto importante aquí es que todo el trabajo desempeñado para hacer este informe no afecta a otros reportes que pueden solicitarse es decir, todos ellos son independientes y caros, desde el punto de vista de recursos y productividad.

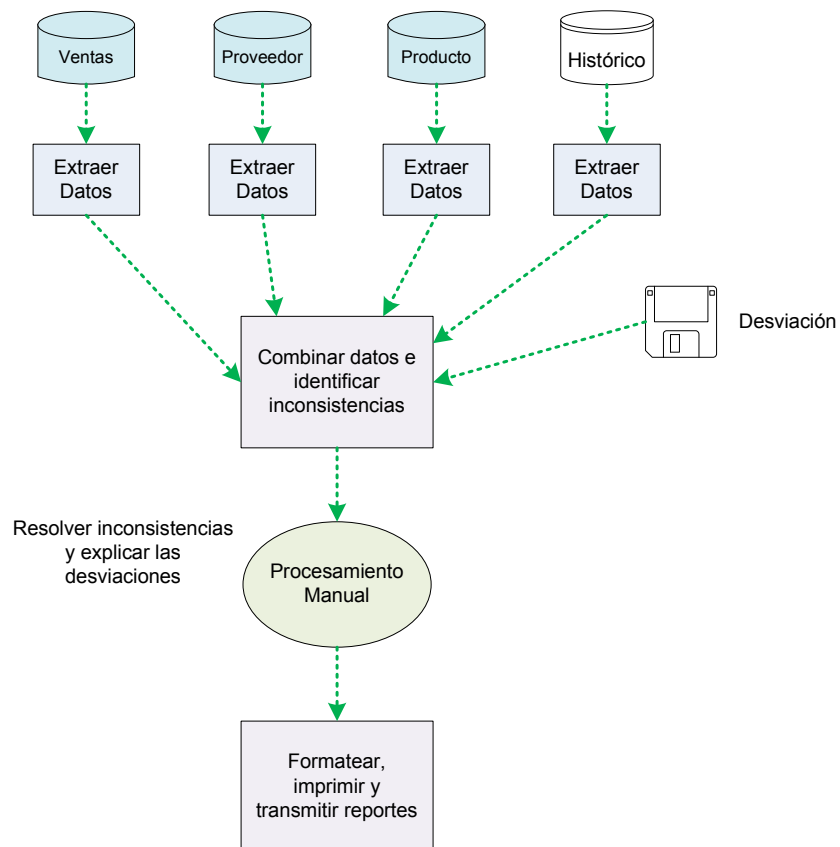


Figura No. 2.9: "Enfoque anticuado de procesamiento de la Información"

Al crear una data warehouse y combinar todos los datos requeridos, se obtienen los siguientes beneficios:

- Las inconsistencias de los datos se resuelven automáticamente cuando los elementos de datos se cargan en el data warehouse, no manualmente, cada vez que se prepara un reporte.
- Los errores que ocurrieron durante el proceso complejo de la preparación del informe, se minimizan porque el proceso es ahora mucho más simple.
- Los elementos de datos son fácilmente accesibles para otros usos, no sólo para un reporte particular.
- Se crea una sola fuente.

Ejemplo 2: Transporte de Cargas y Pasajeros

Se utilizan DWH para almacenar y acceder a meses o años de datos de clientes y sistemas de reservas para realizar actividades de marketing, planeamiento de capacidad, monitoreo de ganancias, proyecciones y análisis de ventas y costos, programas de calidad y servicio a clientes.

Las empresas de transporte de cargas llevan datos históricos de años, de millones de cargamentos, capacidades, tiempos de entrega, costos, ventas, márgenes, equipamiento, etc.

Las aerolíneas utilizan sus DWH para sus programas de viajeros frecuentes, para compartir información con los fabricantes de naves, para la administración del transporte de cargas, para compras y administración de inventarios, etc. Hacen un seguimiento de partes de repuesto, cumplimiento con las regulaciones aeronáuticas, desempeño de los proveedores, seguimiento de equipaje, historia de reservas, ventas y devoluciones de tickets, reservas telefónicas, desempeño de las agencias de viajes, estadísticas de vuelo, contratos de mantenimiento, etc.

Algunas empresas que cuentan con DWH de magnitud: Corrail, Union Pacific, Norfolk Southern, American President Lines, Delta, Lufthansa, QANTAS, British Airways, American Airlines, Canadian Airlines, SNFC.

2.4. Modelo de datos

2.4.1. Modelo relacional

Es un modelo de datos basado en la lógica de predicado y en la teoría de conjuntos. Tras ser postuladas sus bases en 1970 por Edgar Frank Codd, de los laboratorios IBM en San José (California), no tardó en consolidarse como un nuevo paradigma en los modelos de base de datos⁶. La estructura fundamental del modelo es precisamente la relación, una tabla bidimensional la cual está constituida por tuplas (registro) y columnas (atributos). Usa el álgebra relacional, y para manipular la información utiliza ocho operaciones: restrict, project, times, union, minus, join, intersect y divide.

- Restrict: restituye una relación que contiene un subconjunto de las tuplas de la relación a la que se aplica. Los atributos se quedan como estaban.
- Project: restituye una relación con un subconjunto de los atributos de la relación a la que viene aplicado. Las tuplas de la relación resultado se componen de las tuplas de la relación original, de manera que siguen siendo un conjunto en sentido matemático.
- Time: se aplica a dos relaciones y efectúa el producto cartesiano de las tuplas. Cada tupla de la primera relación está concatenada con cada tupla de la segunda.
- Union: aplicando este operador a dos relaciones compatibles, se obtiene una que contiene las tuplas de ambas relaciones. Dos relaciones son compatibles si tienen el mismo número de atributos y los atributos correspondientes en las dos relaciones tienen el mismo dominio.

⁶ Wikipedia. Modelo Relacional. http://es.wikipedia.org/wiki/Modelo_relacional. [consulta 3 de enero de 2007]

- Minus: aplicado a dos relaciones compatibles restituye una tercera que contiene las tuplas que se encuentran sólo en la primera relación.
- Join: se concatenan las tuplas de dos relaciones de acuerdo con el valor de un conjunto de sus atributos.
- Intersect: aplicado a dos relaciones compatibles restituye una relación que contiene las tuplas que existen en ambas.
- Divide: aplicado a dos relaciones que tengan atributos comunes, restituye una tercera que contiene todas las tuplas de la primera relación que se puede hacer que correspondan con todos los valores de la segunda relación.

Este es el modelo más difundido y es considerado el pilar fundamental de las bases de datos que existen hoy en día en grandes y medianas empresas, el único problema es:

- Legibilidad limitada: Los usuarios finales no pueden interpretar el modelo entidad relación, por lo que no pueden navegar por este modelo en busca de información.
- Dificultad de las herramientas para encontrar información: Cuando se maneja una cantidad muy grande de información, las herramientas no siempre pueden realizar una selección correcta de los datos y en algunos casos pueden devolver información errónea.

2.4.2. Modelo Multidimensional

El modelo multidimensional es una técnica que les permite visualizar a los usuarios finales las relaciones que existen en el modelo de una manera mas simple y entendible. Este modelo permite el empleo de cualquier base de datos como MOLAP (matrices multidimensionales), base de datos relacionales, ROLAP (transforma datos multidimensionales en operaciones relacionales en SQL), etc. Los objetivos del modelo son: ⁷

⁷ Peralta, Verónica. "Diseño Lógico de Data Warehouse a partir de esquemas conceptuales multidimensionales". Noviembre, 2001. <http://www.fing.edu.uy/vperalta.pdf>. [consulta 10 de enero de 2007]

- Representación de los datos en forma cercana a la intuición del usuario.
- Resolver problemas planteados en sistemas relacionales.

Un esquema multidimensional esta formado por dos elementos: tablas y esquemas.

Las tablas dentro de un modelo multidimensional se dividen en dos tipos:

- Tabla de hechos: Contienen información obtenida generalmente por una aplicación de una manera estadística, es decir, un conjunto de valores agrupados en un solo valor. Estos son conocidos como atributos de hechos o síntesis.
- Tabla de dimensiones: Poseen atributos conocidos como atributos de dimensión o clasificación y contienen el detalle de los valores obtenidos en los atributos de hecho o síntesis. Las dimensiones poseen jerarquías, que son varios atributos unidos mediante algún tipo de jerarquía, de aquí nace el concepto de que el modelo multidimensional se puede representar mediante un esquema tipo estrella o copo de nieve.⁸
- o Esquema estrella: Su nombre se da por la distribución que tiene este tipo de esquema. Tiene una tabla de hecho en el centro y en las puntas están las tablas de dimensiones.

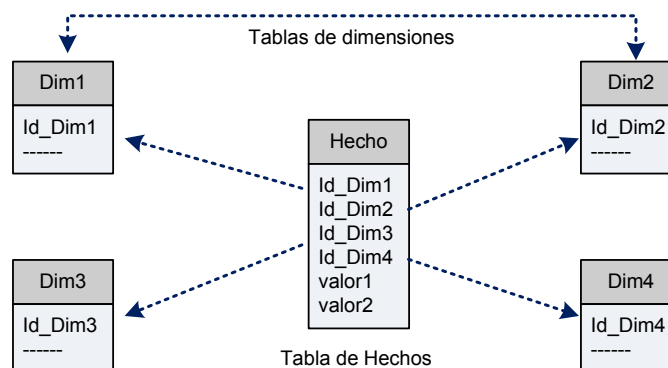


Figura No. 2.10: "Esquema Estrella"

⁸ Sepúlveda Castro, Jorge. Urrutia Suárez, Luis. "Diseño e Implementación de un Data Warehouse para la gestión de ventas de la empresa vitivinícola Miguel Torres Chile". Diciembre, 2004. [consulta 24 de abril de 2007]

- Esquema copo de nieve: La diferencia con el esquema estrella radica en que las tablas de dimensiones están normalizadas y cada una de estas tiene un solo nivel.

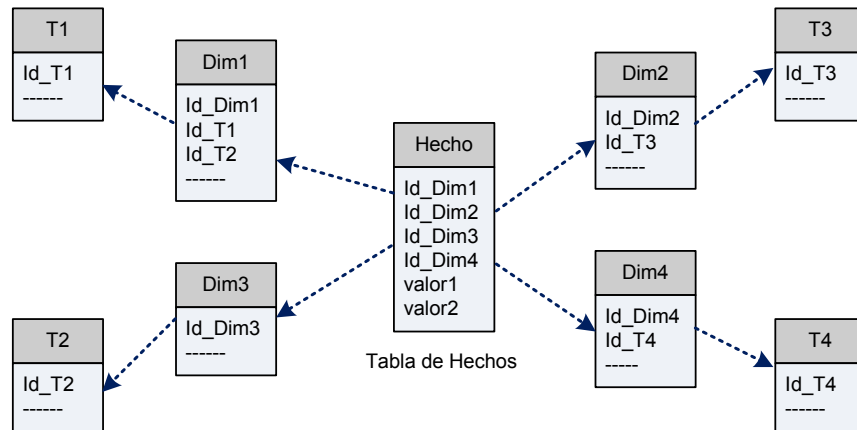


Figura No. 2.11: "Esquema Copo de Nieve"

- Constelación de hechos: Son tablas de hecho que comparten varias tablas de dimensiones.

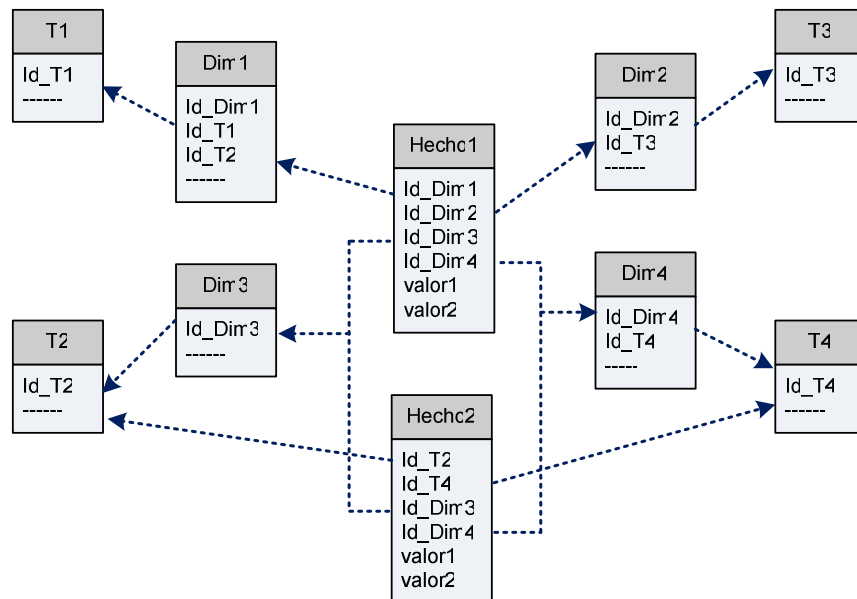


Figura No. 2.12: "Constelación de hechos"

Bases de datos multidimensionales: Dentro de las bases de datos multidimensionales la información se representa como una matriz multidimensional conocida como cubo o hipercubo, en cuyos ejes se encuentran las dimensiones o criterios de análisis, y en los cruces las medidas, que representan los indicadores o valores a analizar, tal como lo

muestra la siguiente figura, en donde las dimensiones año, mes y motivo se cruzan entre si para indicarnos el numero de faltas, que seria nuestra medida a analizar.

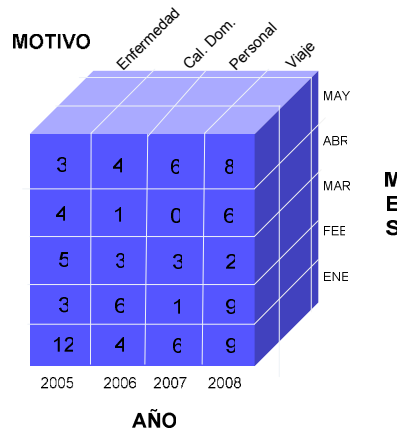


Figura No. 2.13: "Ejemplo de un Cubo de tres dimensiones"

Operaciones en Modelos Multidimensionales:

- Slice & Dice: Permite seleccionar los valores de las dimensiones que se requieren. Como se muestra en la figura 2.14, en donde seleccionamos ciertos motivos en años y meses específicos.

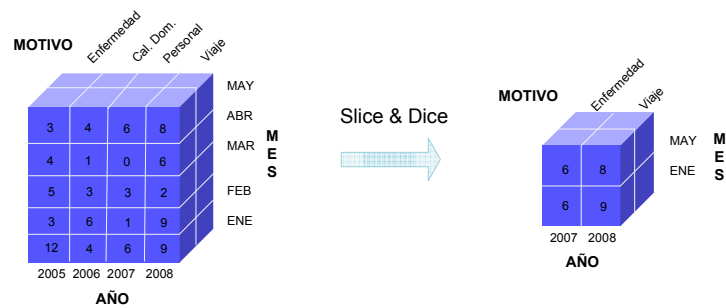


Figura No. 2.14: "Slice & Dice"

- Rotación: Este operador selecciona el orden de visualización de las dimensiones, como lo muestra la siguiente figura.

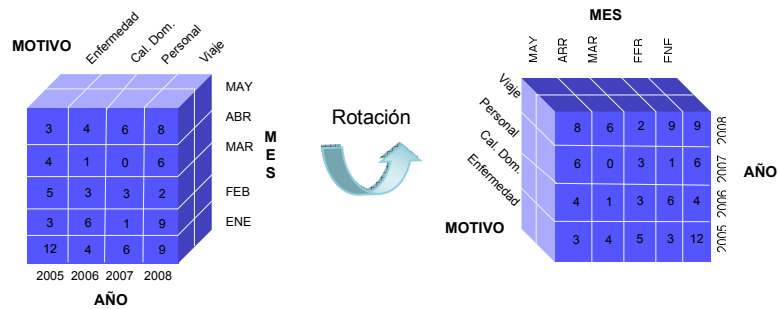


Figura No. 2.15: “Rotación”

- Drill Down – Drill Up: permite bajar o subir a los niveles más atómicos de nuestro esquema multidimensional según la jerarquía establecida en las dimensiones. En la figura 2.16 se pueden apreciar las dos operaciones según la jerarquía de las dimensiones empresa, sección y subsección.

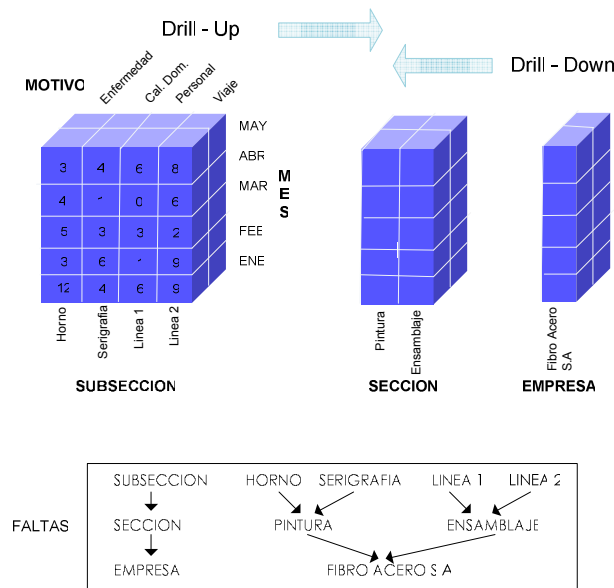


Figura No. 2.16: “Drill – Up y Drill - Down”

- Roll Up: Permite hacer cálculos a las medidas en función de agrupamientos. Estas operaciones pueden ser: count, sum, min, max, avg, etc, en el siguiente ejemplo se realiza un roll up suma sobre las faltas de los empleados.

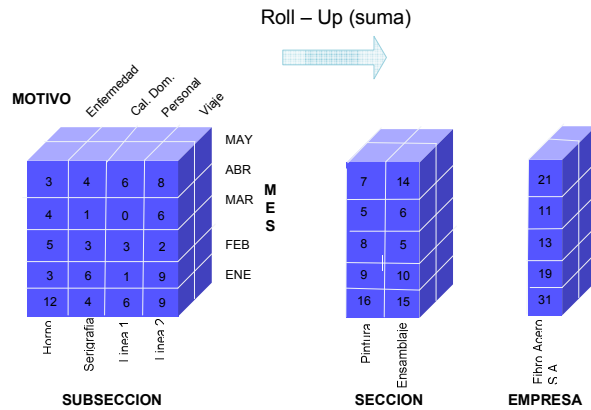


Figura No. 2.17: “Roll - Up”

- Drill-Across: Permite combinar datos de varios cubos, como lo muestra la figura 2.18.

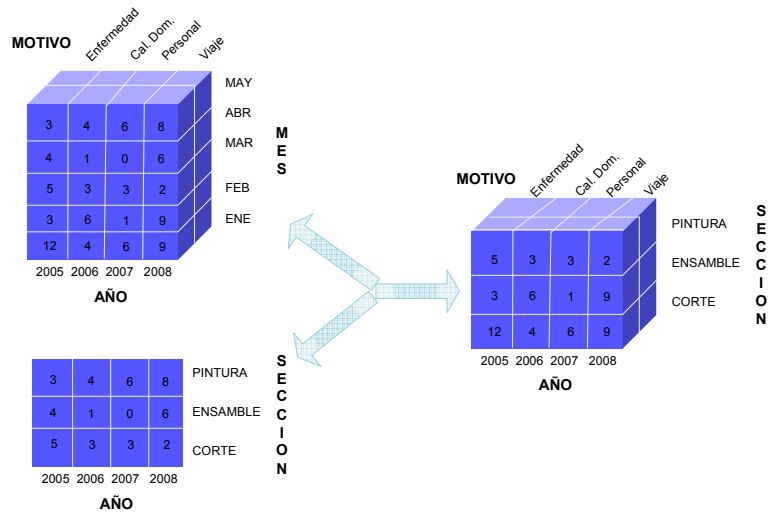


Figura No. 2.18: “Drill - Across”

- Drill-Through: Operador que permite acceder a datos descriptivos, un ejemplo de esto en la figura 2.19

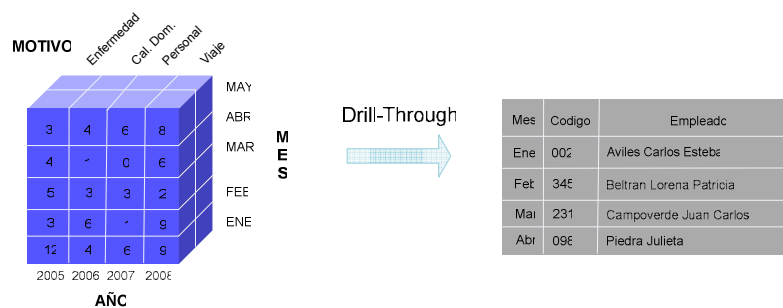


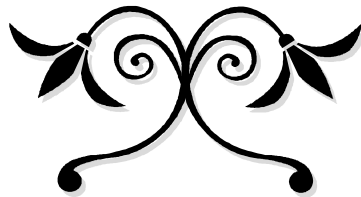
Figura No. 2.19: “Drill - Through”

2.5. Conclusiones

La necesidad en las empresas de información más detallada y de fácil acceso para tomar decisiones acertadas y bien fundamentadas, hace necesaria la creación de aplicaciones para soporte de decisiones basadas en un data warehouse.

El DWH es un repositorio, donde se almacenan datos estratégicos, tácticos y operativos. Toda la información esta previamente depurada para garantizar la calidad, y su orientación hacia el negocio, su información se explota mediante herramientas flexibles que independizan en lo posible al usuario del desarrollo informático, llamadas front-end.

La información que contendrá este Data Warehouse será elaborada a partir de las diferentes fuentes de datos que existen en la empresa, y el modelo a utilizar para esquematizarlo será el multidimensional, ya que su representación de los datos es cercana a la intuición del usuario.



CAPITULO 3

HERRAMIENTAS PARA LA CONSTRUCCIÓN, ADMINISTRACIÓN Y MANIPULACIÓN DE UN DATA WAREHOUSE

3. HERRAMIENTAS PARA LA CONSTRUCCIÓN, ADMINISTRACIÓN Y MANIPULACIÓN DE UN DATA WAREHOUSE

3.1. Introducción

Para la utilización de un DWH es de gran importancia conocer las tres fases que nos garantizarán una correcta administración del mismo, las cuales son: Transformación, Administración de la base de datos y por último la Manipulación de los datos.

Cada una de estas fases tienen diferentes herramientas para poder ser creadas, y deben ser elegidas cuidadosamente según su funcionalidad, robustez, escenarios en los que van a trabajar, etc.

En este capítulo se indicarán las herramientas que pueden ser útiles en cada una de las fases, indicando su definición y características principales.

3.2. Herramientas para la transformación del Data Warehouse

En esta fase se realiza la transformación de inconsistencias y formatos de los datos de las diferentes bases de datos operacionales según las reglas de transformación establecidas (Figura No. 3.1).

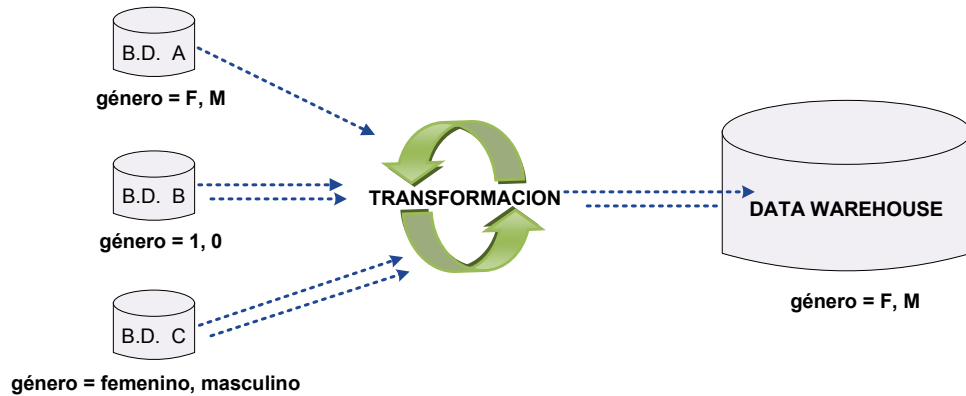


Figura No. 20 3.1: “Etapas del Data Warehouse - Transformación”

La transformación de los datos es muy importante dentro de la construcción del DWH ya que de ello depende el correcto funcionamiento y la consistencia de la información. Las herramientas que se pueden utilizar para realizar el objetivo de este nivel son innumerables, basta con utilizar una herramienta de programación que permita interactuar con las bases de datos operacionales; entre las más conocidas tenemos: GeneXus, Oracle, Microsoft Visual Basic, Microsoft Visual FoxPro, etc.

3.2.1. GeneXus y el Data Warehouse

Para la fase de transformación de la información del DWH en la empresa Fibro Acero S.A. se utilizará la herramienta de programación GeneXus. La principal razón para utilizar esta herramienta es que la utilizan en la empresa y es el lenguaje mas conocido por los desarrolladores de este proyecto. A continuación algunas características de dicha herramienta.

GeneXus: Es una herramienta inteligente generadora de código y base de conocimiento para el diseño y desarrollo de complejas aplicaciones multiplataforma, que permite el desarrollo de aplicaciones partiendo del diseño de una base de conocimiento, lo cual ofrece un ahorro de hasta un 30% en tiempo y recursos, y de un 70% aproximadamente en la fase de

mantenimiento⁹, realizando los cambios desde la base de conocimiento y transmitiéndolos hacia las aplicaciones generadas de ésta fuente.

Miles de medianas y grandes empresas y casas de software en todo el mundo utilizan GeneXus para desarrollar complejos sistemas de misión crítica con grandes bases de datos que comprenden desde sistemas centralizados, distribuidos hasta aplicaciones Web. Estos sistemas integran módulos de data warehouse, Web services, portales corporativos y mucho más.

- GeneXus es una poderosa herramienta para el diseño y desarrollo de software multiplataforma. Permite el desarrollo incremental de aplicaciones críticas de negocio de forma independiente de la plataforma.
- GeneXus genera el 100% de la aplicación. Basado en los requerimientos de los usuarios realiza el mantenimiento automático de la base de datos y del código de la aplicación, sin necesidad de programar.
- GeneXus soporta las plataformas y lenguajes líderes y los DBMS más populares.

GeneXus y GXplorer apoyan el ciclo vital entero de un almacén de los datos, desde el desarrollo hasta el análisis de datos.

3.3. Herramientas para la administración del data warehouse

La segunda fase que tiene un DWH es el almacenamiento y administración de la información, esto se lo puede lograr por medio de una buena base de datos, la cual nos brinde un tiempo de respuesta menor, datos consistentes y aminorar el costo de mantenimiento y administración de información.

⁹ Genexus. <http://www.technologies.cl/hxwec20.exe?1,1,64>. [consulta 29 de noviembre de 2006]

3.3.1. Clasificación y características

A continuación se detallan algunas herramientas para la administración y almacenamiento de información en un DWH.

3.3.1.1. Oracle

La construcción de su almacén de datos es la mejor manera de consolidar y organizar todos sus datos para que puedan ser manejados, alcanzados y analizados fácilmente. Con una única interfaz de administración y capacidades de self-tuning y autodiagnóstico, Oracle simplifica el mantenimiento de su almacén de los datos en continuo crecimiento, con el funcionamiento más rápido y la mejor relación precio/rendimiento del mercado.

El Oracle Warehouse Builder 10g es la herramienta para el diseño de la integración del BI corporativo que gestiona el ciclo vital completo de datos y metadatos para la base de datos Oracle 10g. Algunas de sus características son las siguientes:

- Posee igual interacción en todas las plataformas (Windows, Unix, Macintosh, Mainframe).
- Soporta bases de datos de todos los tamaños, desde severas cantidades de bytes y gigabytes en tamaño.
- Provee salvar con seguridad de error lo visto en el monitor y la información de acceso y uso.
- Soporta un verdadero ambiente cliente servidor. Este establece un proceso entre bases de datos del servidor y cliente para la aplicación de programas.

3.3.1.2. PostgreSQL

Es un manejador de bases de datos relacional y de objetos de alta escalabilidad y compatible con SQL. Cuenta con más de 15 años de experiencia en desarrollo y corre en la mayoría de los sistemas operacionales: Linux, UNIX (AIX, BSD, HP-UX, SGI IRIX, Mac OS X, Solaris,

SunOS, Tru64), BeOS, y Windows¹⁰. Esta base de datos cuenta con todas las características de los motores de base de datos privados como: full ACID compliant, tiene soporte a foreign keys, joins, views, triggers y stored procedures (en diferentes lenguajes), incluye soporte a SQL92 y SQL99 data types, incluyendo INTEGER, NUMERIC, BOOLEAN, CHAR, VARCHAR, DATE, INTERVAL, y TIMESTAMP. También soporta almacenamiento de BLOBs. Cuenta con interfaces de programación para C/C++, Java, Perl, Python, Ruby, Tcl, ODBC y una excelente documentación. Entre algunas de sus características tenemos:

- Agregación en memoria usando tablas de hashing para hacer que las consultas sean más rápidas.
- Mejoras en el manejo de subconsultas por el optimizador, que resultan en hasta un 400% de incremento en la velocidad de respuesta de consultas complejas.
- Nuevo protocolo de conexión que incrementa la velocidad de las transferencias de datos.
- Implementación mejorada de índices funcionales que permite una mejor indexación en tipos de datos personalizados y compuestos.

3.3.1.3. Db2

Es una plataforma de depósito de datos totalmente integrada de IBM, ofrece una integración más completa de los metadatos y la infraestructura de ejecución. DWE (Data Warehouse Edition) le ayuda a conseguir información en demanda al integrar los componentes principales de la administración del depósito, la minería de datos, el proceso analítico online (OLAP) y la creación de informes/análisis en línea, junto con el rendimiento y la utilización mejorada de funciones de business intelligence (BI), cuyo fin es obtener un conocimiento en tiempo real y facilitar la toma de decisiones. Amplía el valor del depósito de datos empresarial con:

¹⁰ Asociación Colombiana de Ingenieros de Sistemas. <http://www.acis.org.co/index.php?id=623>.
Noviembre 2005. [consulta 30 de noviembre de 2005]

- Proceso analítico online (OLAP) y análisis más rápidos, mediante DWE Cube Views y Tablas de consultas materializadas.
- Gestión de tablas derivada y resumen mediante DWE SQL Warehousing Tools.
- Minería que se conecta fácil y directamente con el depósito de datos empresariales (EDW) mediante DWE Intelligent Miner Modeling.
- Técnicas de minería de datos que aprovechan la potencia del EDW distribuido en varios sistemas mediante DWE Intelligent Miner Scoring.
- Comprensión y análisis de datos muy eficaz mediante DWE Intelligent Miner Visualization.
- Análisis integrado de DB2 Alphablox, que aprovecha la comprensión EDW en sus procesos empresariales.
- Entorno de herramientas común para el modelado de datos, el modelado de cubos y el flujo/control de datos SQL mediante DWE Design Studio.

3.3.2. Microsoft SQL Server 2000 y el Data Warehouse

La herramienta de administración del DWH que va a ser utilizada en el desarrollo de este proyecto es Microsoft SQL Server 2000, debido que la empresa cuenta con los permisos y licencias necesarias para su utilización y es el administrador de base de datos más conocido por los desarrolladores de este proyecto. A continuación se define y se detallan algunas características de Microsoft SQL Server 2000.

Microsoft SQL Server 2000 es uno de los motores de base de datos más conocido en grandes y pequeñas empresas, debido a la fiabilidad, calidad y facilidad de uso, y es considerado una excelente plataforma para procesos transaccionales en línea (OLTP), almacenamiento de datos y aplicaciones de comercio electrónico. Esta herramienta tiene cierta evolución en comparación a su versión anterior, entre las principales tenemos:

- Analysis Services de SQL Server. Conocido anteriormente como servicios OLAP, el cual incluye componentes para la minería de datos.
- Meta Data Services de Microsoft SQL Server que es su nuevo repositorio de datos.

Entre sus características principales encontramos:

- Integración con Internet. SQL Server 2000 incluye compatibilidad integrada con XML y con diferentes componentes de almacenamiento de datos para sitios Web de mayor tamaño.
- Escalabilidad y disponibilidad. Puede soportar diferentes plataformas, como equipos con Microsoft Windows 98, por medio de servidores que tengan Microsoft Windows 2000. También admite características de servidores federados, vistas indizadas y soporte para memorias grandes, aspectos importantes para diferentes sitios Web.
- Características de base de datos corporativas. Protege la integridad de los datos a la vez que minimiza la carga de trabajo de la administración de miles de usuarios modificando la base de datos simultáneamente. De igual manera, permite hacer referencia a datos de varios orígenes como si fueran parte de la base de datos de SQL Server 2000.
- Facilidad de instalación, distribución y utilización. Incluye un conjunto de herramientas administrativas y de desarrollo para mejorar el proceso de instalación, distribución, administración y uso del SQL Server 2000.
- Almacenamiento de datos. Incluye herramientas para extraer y analizar datos de resumen para el procesamiento analítico en línea.

3.4. Herramientas para manipular la información del data warehouse

Esta es la tercera y última fase en la elaboración de un DWH, la cual nos permite visualizar, utilizar y manipular la información almacenada en nuestro almacén de datos. Existen diversas herramientas las cuales nos permiten manipular un DWH, unas más conocidas y potentes que otras.

3.4.1. Clasificación y características

A continuación se describirán las herramientas más conocidas para la manipulación de información en un DWH.

SBI Discoverer Application: Esta es una herramienta de tipo OLAP (tratamiento analítico en línea) que tiene como objetivo integrar una plataforma para el análisis multidimensional de información destinada a optimizar el proceso de toma de decisiones. SBI Discoverer está diseñada para ser utilizada por empresas de diversos sectores y usuarios de diferentes niveles en las organizaciones.

Para la utilización de esta herramienta se utilizan 2 componentes conocidos como SBI Discoverer y SBI Generator. El primero es un potente navegador multidimensional, generador de reportes y buscador de información oculta, el cual tiene la posibilidad de realizar graficas y dar formato de presentación a los informes. Además es de fácil utilización, ya que posee la flexibilidad para que los usuarios puedan ir descubriendo las respuestas a sus preguntas, detectando indicadores y desvíos ocultos en la información. Por otra parte, SBI Generator permite generar cubos multidimensionales para ser analizados por SBI Discoverer.

Entre algunas de las características de SBI Discoverer Application tenemos:

- Análisis multidimensional por planilla.
- Construcción de gráficos totalmente configurables.
- Drill Down / Drill Up.
- Creación de medidas calculadas.
- Generación de reportes con actualización automática.
- Aplicable a cualquier modelo de datos.

- Conexión a múltiples bases de datos vía ODBC.
- Creación de dimensiones, niveles y medidas de una manera sencilla.
- Permite particionar los cubos de información para una mayor velocidad.

GXplorer: Es una herramienta nos permite realizar la construcción y administración de un DWH. Esta compuesta de 2 elementos: GXplorer OLAP para las consultas de un DWH y GeneXus Query para las consultas dinámicas sobre la base de datos. Este es un producto que permite obtener la información estratégica necesaria para realizar una correcta toma de dediciones en una organización, y a la vez le da a conocer el estado verdadero en el que se encuentra su empresa de manera sencilla y con la interfaz que prefiera: Windows o Web.

GXplorer se encarga de minimizar la complejidad en la construcción y utilización del DWH, permitiendo al ejecutivo una utilización muy simple e inmediata dentro de un ambiente que le es muy natural: EXCEL. Permite administrar todo el ciclo de vida de un DWH, desde la construcción y carga de los datos operacionales hasta la utilización de la información.

A continuación se detallan algunas de las características de GXplorer.

- Permite implementar business intelligence de una manera mas ágil y con un costo total menor (desarrollo, implementación y mantenimiento).
- Se puede acceder a la información de manera local y remota a través de la Web.
- Seguridad en el acceso de la información según el perfil del usuario.
- Con GXplorer se pueden realizar consultas de una manera sencilla y utilizarlas en EXCEL.

- Se puede administrar todo el ciclo de vida de un DWH, automatizando los procesos de creación, carga de datos y mantenimiento.
- Nos permite utilizar un método incremental en el desarrollo del DWH.

3.5. Conclusiones:

Las herramientas dentro del desarrollo de un proyecto son lo más importante a tomar en cuenta, ya que de estas depende la rapidez, robustez y en si la calidad y funcionalidad del proyecto final.

De acuerdo a cada etapa se han escogido las siguientes herramientas para la implementación del DWH en Fibro Acero S.A.

Genexus: Se lo utilizará para la construcción y transformación de la información por las siguientes razones.

- Es el lenguaje utilizado en la empresa por lo que se cuenta con licencia, además es el mas conocido por los desarrolladores de este proyecto.
- Al ser una herramienta generadora de código, permite ahorrar hasta un 30% en tiempo y recursos, y un 70% en la fase de mantenimiento
- Es una herramienta para el diseño y desarrollo de software multiplataforma.
- Soporta las plataformas y lenguajes líderes y los DBMS más populares.

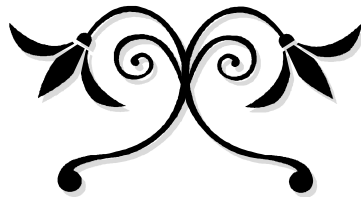
Microsoft SQL Server 2000: es la herramienta de administración del DWH que va a ser utilizada en el desarrollo de este proyecto.

- Permite la integración con Internet.
- Escalabilidad y disponibilidad.
- Protege la integridad de los datos.
- Facilidad de instalación, distribución y utilización.
- Incluye herramientas para extraer y analizar datos de resumen para el procesamiento analítico en línea.

Explorer y Gxquery: Serán las herramientas front-end a utilizar en nuestro proyecto.

- Los usuarios utilizan y manipulan la información por medio de EXCEL.

- Permite el manejo de perfiles, lo cual provee seguridad en el acceso a la información.
- Se puede administrar todo el ciclo de vida del DWH
- Permite implementar BI de una manera más ágil y a un menor costo.
- Se puede acceder a la información a través de la Web.
- Nos permite utilizar un método incremental en el desarrollo del DWH.



CAPITULO 4

METODOLOGIA DE DESARROLLO PARA LA CONSTRUCCION E IMPLEMENTACION DEL DATA WAREHOUSE

4. METODOLOGIA DE DESARROLLO PARA LA CONSTRUCCION E IMPLEMENTACION DEL DATA WAREHOUSE

4.1. Introducción

Durante los últimos años, se han propuesto diferentes metodologías de análisis y diseño de sistemas de información. Sin embargo, aún no hay una metodología específica para el análisis y diseño de sistemas Data Warehousing que sea ampliamente aceptada.

Por lo tanto, para desarrollar nuestro proyecto nos basaremos en las fases de desarrollo de sistemas transaccionales, pero con variantes únicas asociadas al DWH.

Durante el desarrollo de este capítulo explicaremos cada una de las fases mientras se va desarrollando el DWH para la empresa Fibro Acero S.A. utilizando la herramienta case Genexus y base de datos SQL Server 2000 para tal fin.

4.2. Fases para la construcción del Data Warehouse

En total son cinco fases las que implementaremos para construir nuestro DWH, se muestran a continuación en la Figura 4.1.

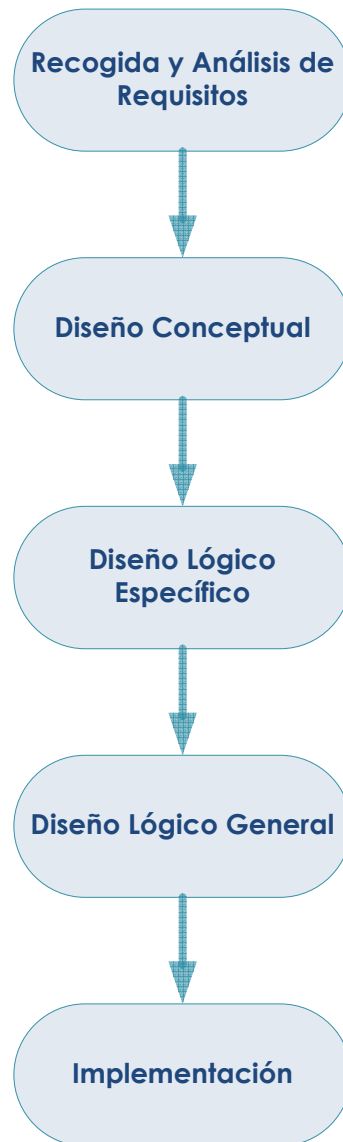


Figura No. 4.1: "Fases para la construcción de un DWH"

- Recogida y Análisis de Requisitos: aquí se especifican las funciones, servicios y demás resultados que se esperan obtener del DWH a través de la determinación de fuentes de extracción de información, tanto para el sistema OLTP de la empresa, como para las fuentes externas, y también los datos obtenidos de entrevistas y consultas a los usuarios.

- **Diseño Conceptual:** en esta etapa se especifica el problema a resolver, a través de la construcción del diseño o esquema conceptual del Data Warehouse en base a los requerimientos y la información obtenida en la etapa anterior.
- **Diseño Lógico Específico:** el objetivo de esta etapa es construir el esquema lógico partiendo del modelo conceptual generado en la etapa anterior y del análisis sobre el performance y almacenamiento de la información.
- **Diseño Lógico General:** definición del esquema ROLAP o MOLAP y diseño de las ETL (Extraction, Transformation, Load).
- **Implementación:** construcción del Data Warehouse, y preparación de la interfaz con el usuario a través de las herramientas front-end

4.2.1. Recogida y Análisis de Requisitos

Esta fase es una de las más importantes para el desarrollo del DWH, pues aquí delimitaremos los objetivos y resultados que se esperan obtener en base a las necesidades de los usuarios y a las fuentes de información disponibles en la empresa.

Además, con el progreso del análisis que realicemos, se detectaran ciertos inconvenientes, como la transformación de la información si existiera incompatibilidad en el formato de los datos, o la falta de la misma.

Para el desarrollo de esta fase, utilizaremos ciertos métodos que nos permitan obtener y depurar los requerimientos de los usuarios, como las reuniones de definición de requerimientos con los implicados en el desarrollo del proyecto, reuniones con el personal de sistemas para determinar como se llevaran acabo dichos requerimientos y analizar la información existente en la empresa Fibro Acero S.A.

4.2.1.1. Reuniones de definición de requerimientos

Como primer paso realizamos entrevistas a los jefes de los diferentes departamentos en donde se implementara el DWH. A continuación se muestran los requerimientos obtenidos.

- Departamento de Ventas:
Se necesita mostrar la cantidad y el valor de las ventas, tanto locales, nacionales y de exportación, las cuales se puedan dimensionar por el año y mes de venta, tipo de venta, sucursal, ciudad, tipo de cliente, grupos, líneas, productos, estado del cliente y tipo de documentos. El objetivo de este requerimiento es el poder visualizar o explorar las ventas de una manera detallada y acumulada, por medio de los parámetros antes mencionados.

- Departamento de Logística:
Se necesita generar un informe comparativo entre los proveedores que tiene la empresa, para conocer la eficiencia de cada uno de ellos tomando en cuenta la calidad del producto, tiempo de entrega y cantidad entregada. Cada proveedor esta clasificado según la ingerencia del material en la línea de producción, pudiendo llegar a pertenecer al Tipo A, B o C. También es de importancia para el departamento el poder dimensionar las calificaciones según el proveedor, ciudad y país del proveedor, año y periodo de entrega, calificación, clasificación, calidad del producto, entrega a tiempo y cantidad entregada.

- Departamento de Talento Humano:
En este departamento encontramos 2 tipos de necesidades:
Faltas y Permisos: Conocer la cantidad en horas y el numero de faltas y permisos que solicitan los empelados de la empresa en un rango de tiempo especifico y así poder llevar un mejor control sobre las justificaciones y motivos por los cuales se han realizado dichos permisos. Es de vital importancia para el departamento, que estos datos puedan ser analizados por: la empresa en la que trabajan,

sección, subsección, departamento, turno, motivo, año y mes, si esta justificado, si es descontable o si el permiso está anulado.

Remuneraciones: Conocer la variación del valor de las remuneraciones según la Empresa, Sección, Subsección, Departamento, Empleado, Año y Mes.

4.2.1.2. Reuniones con el personal de Sistemas

Una vez establecidas las necesidades de los usuarios, nuestro siguiente paso es entrevistarnos con los encargados del área de sistemas, el Ing. Geovanny Vintimilla, jefe del área y el Ing. Demetrio Toledo, desarrollador de la empresa, para analizar y determinar las fuentes de información que utilizaremos y verificar si es posible satisfacer todas las necesidades por medio de la información que la empresa tiene almacenada en sus bases de datos.

4.2.1.2.1 Análisis de las fuentes de información

La empresa Fibro Acero S.A. cuenta con dos sistemas desarrollados por el departamento de sistemas, EcoSAC y Producción, creados con la herramienta case Genexus y utilizan una base de datos en SQL Server 2000 que se encuentra almacenada en un servidor con Windows Server 2003.

Junto al personal de sistemas se procedió a revisar los sistemas antes mencionados para analizar los requerimientos establecidos y descartar las necesidades que actualmente se encuentran solucionadas por los sistemas oltp, también se verificó que exista la información necesaria en las bases de datos para poder implementar el DWH.

Como se van a crear 3 DWH (uno por cada departamento) veremos por separado a cada uno de ellos.

Departamento de Talento Humano: El departamento tiene dos requerimientos por cumplir, el primero es de Permisos y Faltas y el segundo de Remuneraciones.

- Permisos y Faltas: Por medio del análisis antes mencionado, la información se obtendrá del programa EcoSAC, el cual guarda la información en SQL Server 2000 en la base de datos "fa002". Aquí se

necesitará acceder a 10 tablas para poder obtener la información necesaria. En la Figura 4.2 se muestra el modelo entidad relación para esta base de datos.

- Remuneraciones: La información que requiere el DWH se encuentra en archivos externos de tipo Excel, por lo que es necesario realizar un proceso de carga directa. Estos archivos son generados por el propio departamento y han sido modificados para que su nueva estructura sea aprovechada en el DWH.

Departamento de Ventas: La información de las ventas que realiza la empresa y sus sucursales la maneja el sistema Adviser, que esta desarrollado en cobol y la información la almacena en archivos en un Servidor con Red Hat version 8.0, para este caso el departamento de sistemas tiene ya creado un proceso, el cual, permite importar la información del sistema Adviser en Linux a Windows y almacenarlo en una base de datos en SQL Server 2000 llamada "Ventas", de donde nuestro DWH obtendrá la información. El modelo entidad relación de esta base de datos será indicada en la Figura 4.3.

Departamento de Logística: Para desarrollar este DWH se encontró un problema, el cual no permitirá cumplir con todas las necesidades del departamento, puesto que actualmente el programa que este departamento utiliza "Producción" se encuentra en etapa de pruebas. En nuestro caso, se procederá a realizar una simulación de cual sería el almacén de datos resultante, mediante la importación de archivos de Excel los cuales contienen ya la información requerida por el departamento.

Diagramas Entidad Relación: A continuación se muestran los diagramas entidad relación de las tablas que utilizamos en el desarrollo del DWH. En el caso de los archivos de Excel se muestra un extracto en el Anexo 4.1.

Diagrama Entidad Relación de la base de datos fa002 del sistema EcoSAC para el DWH de Talento Humano.

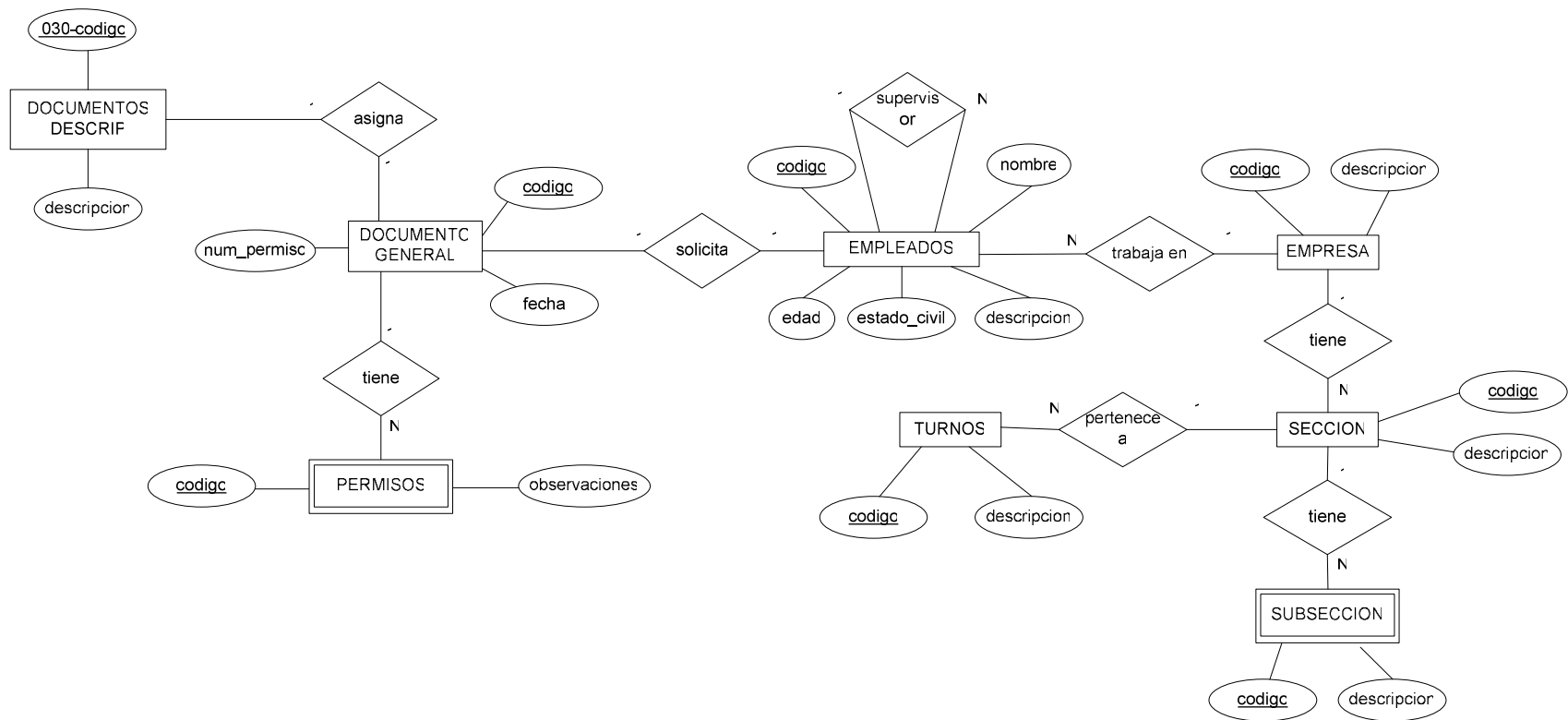


Figura No. 4.2: “Diagrama Entidad Relación del Sistema EcoSAC” ¹¹

¹¹ Vintimilla, Geovanny. Modelo Entidad Relación Sistema EcoSAC. [consulta 23 de noviembre de 2007]

Diagrama Entidad Relación de la base de datos Ventas.

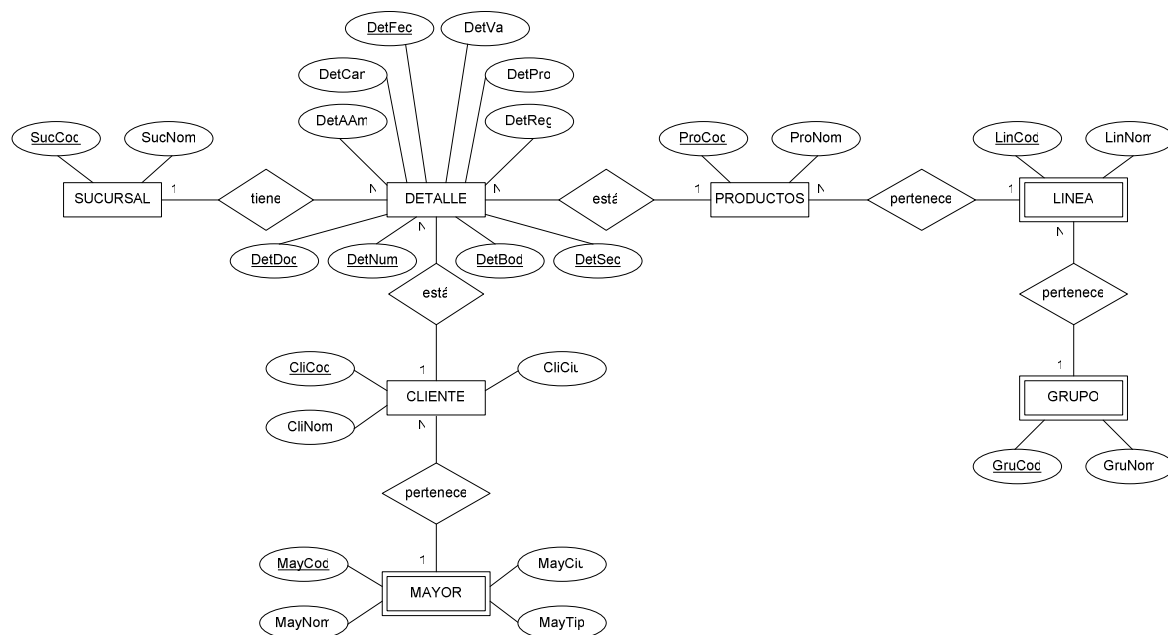


Figura No. 4.3: "Diagrama Entidad Relación de la base de datos Ventas" ¹²

4.2.2. Diseño Conceptual

Para desarrollar el esquema conceptual, utilizaremos los modelos multidimensionales, q están formados básicamente por dimensiones, medidas y hechos, para representar la realidad.

Para seleccionar los objetos relevantes y determinar cuales serán dimensiones o medidas, existen tres enfoques, el primero se basa en el análisis de requerimientos, el segundo en un análisis de las bases de datos fuente y el tercero un intermedio entre ambos. ¹³

¹² Vintimilla, Geovanny. Modelo Entidad Relación de la migración de la base de datos Ventas [consulta 28 de noviembre de 2007]

¹³ Peralta, Verónica. "Diseño Lógico de Data Warehouse a partir de esquemas conceptuales multidimensionales". Noviembre, 2001. <http://www.fing.edu.uy/vperalta.pdf>. [consulta 15 de enero de 2007]

En el enfoque basado en requerimientos se analizan los requerimientos de los usuarios obtenidos en la fase anterior y se identifican en ellos los hechos, dimensiones y medidas relevantes.

En el enfoque basado en las bases fuentes se construyen cubos multidimensionales transformando un esquema conceptual de las bases fuentes, en general se utiliza el modelo E/R, por lo general se comienzan por identificar en el esquema fuente los posibles hechos relevantes para la toma de decisiones, y a partir de los hechos identificados se navega por las entidades y relaciones construyendo las jerarquías de las dimensiones.

Para nuestro trabajo escogimos el modelo CMDM (Conceptual Multidimensional Data Model) desarrollado por Carpani, cuyo enfoque esta dirigido a analizar los requerimientos e identificar en ellos los hechos, dimensiones y medidas relevantes. Aparte incluye nuevas estructuras llamadas relaciones dimensionales, que representan las relaciones entre los objetos.¹⁴

4.2.2.1. Estructuras básicas en el modelo CMDM

Nivel: conjunto de objetos que son de un mismo tipo. Gráficamente los atributos de los niveles o ítems aparecen debajo del nombre del nivel. Los ítems que identifican un nivel llevan un numeral (#) a la derecha (clave absoluta), y los que lo identifican sólo con respecto a su padre (clave débil o relativa) aparecen subrayados. En la figura 4.4 se puede observar el nivel Empleado con sus respectivos atributos.

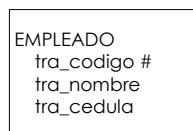


Figura No. 4.4: "Notación Gráfica en CMDM del nivel Empleado"

¹⁴ Carpani, Fernando. "CMDM: Un modelo Conceptual para la Especificación de Bases Multidimensionales". Agosto, 2000. [consulta 20 de abril del 2007]

Dimensión: Es un conjunto de niveles relacionados jerárquicamente, del nivel con más detalle (hijo) al nivel con menos detalle (padre). En cada jerarquía se tiene una relación *uno a muchos* entre objetos del nivel superior e inferior.

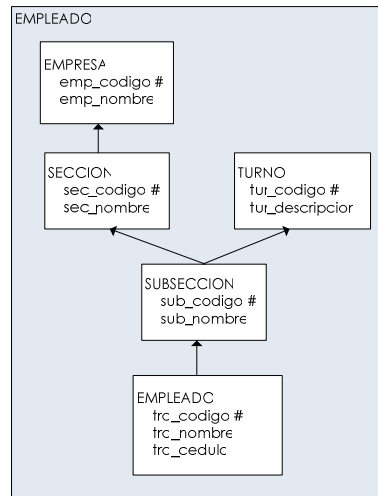


Figura No. 4.5: "Notación Gráfica en CMDM de la Dimensión Empleado"

Relaciones Dimensionales: representa el conjunto de todos los cubos que se pueden construir a partir del cruzamiento de los niveles de un conjunto dado de dimensiones¹⁵, en cada uno de los cubos debe aparecer al menos un nivel de cada una de las dimensiones que participan en la relación y uno de ellos puede representar la medida.

En CMDM, un cubo representa un cruzamiento concreto entre niveles determinados de las Jerarquías de la relación dimensional, es decir, un conjunto de datos relacionados con un tema en particular como Ventas, Logística, Talento Humano, etc.

Por lo tanto, el esquema de una relación dimensional está dado por un nodo central que tiene el nombre de la relación dimensional y los nodos de cada una de las dimensiones que participan de la relación.

¹⁵ Peralta, Verónica. "Diseño Lógico de Data Warehouse a partir de esquemas conceptuales multidimensionales". Noviembre, 2001. <http://www.fing.edu.uy/vperalta.pdf>. [consulta 15 de enero de 2007]

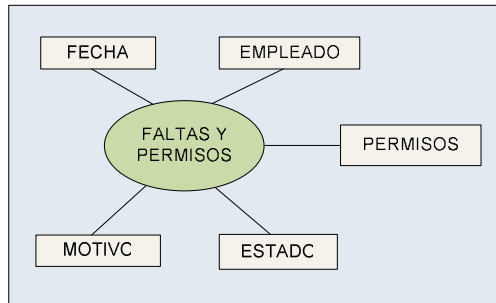


Figura No. 4.6: “Notación Gráfica en CMDM de la Relación Dimensional Faltas y Permisos”

4.2.2.2. Implementación del Diseño Conceptual

Con los requerimientos obtenidos en el capítulo anterior y el análisis realizado, procederemos a implementar el diseño conceptual CMDM definiendo las estructuras necesarias para el proyecto de acuerdo a cada departamento.

- Departamento de Ventas: En este departamento se obtuvieron seis dimensiones que permitirán implementar el requerimiento de ventas. A continuación se muestran las dimensiones con sus respectivos atributos y la relación dimensional resultante del cruzamiento entre las dimensiones.

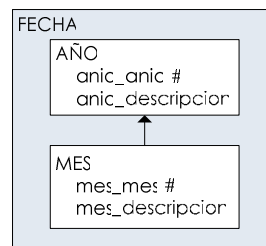


Figura No. 4.7: “Representación Gráfica Dimensión Fecha”

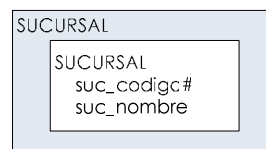


Figura No. 4.8: “Representación Gráfica Dimensión Sucursal”

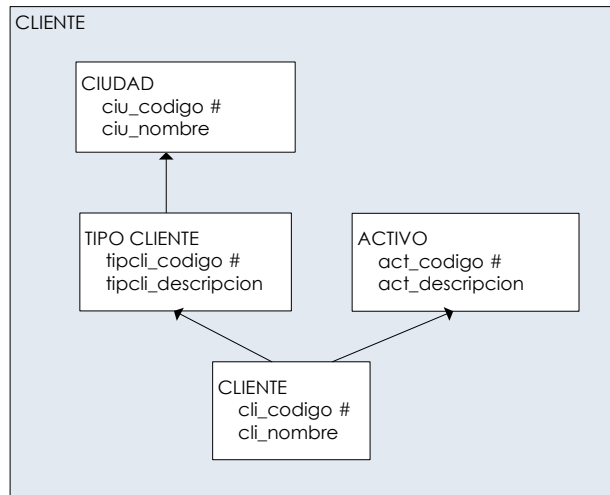


Figura No. 4.9: "Representación Gráfica Dimensión Cliente"

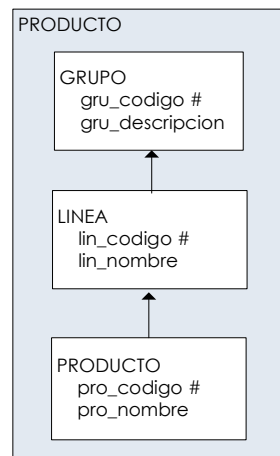


Figura No. 4.10: "Representación Gráfica Dimensión Producto"

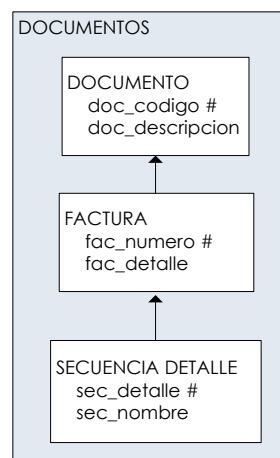


Figura No. 4.11: "Representación Gráfica Dimensión Documentos"

Dimensión Total Ventas: Al manejarse en CMDM la dimensionalidad genérica, las medidas también se las representa como dimensiones. La dimensión Total Ventas posee un solo nivel compuesto por los atributos ven_total que representa el total de ventas en dinero y ven_cantidad que es la cantidad en unidades vendidas.

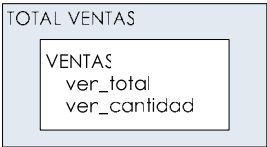


Figura No. 4.12: “Representación Gráfica Dimensión Total Ventas”

Relación Dimensional Ventas: Como resultado del cruzamiento entre las diferentes dimensiones obtenidas para este departamento se obtuvo la relación dimensional que se muestra en la figura 4.13.

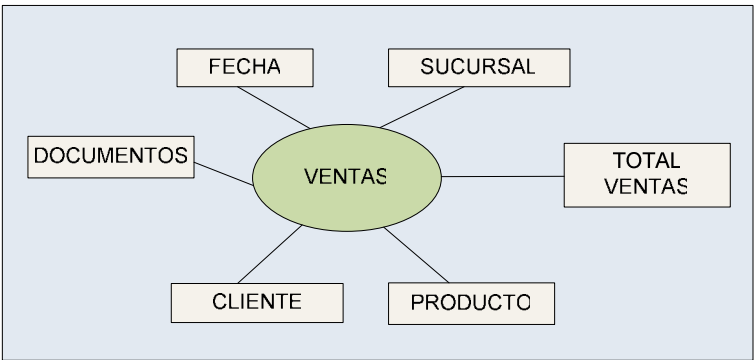


Figura No. 4.13: “Representación Gráfica Relación Dimensional Ventas”

- Departamento de Logística: como resultado del análisis del requerimiento solicitado por este departamento se han creado cinco dimensiones con sus respectivos atributos, y se muestran a continuación representados gráficamente.

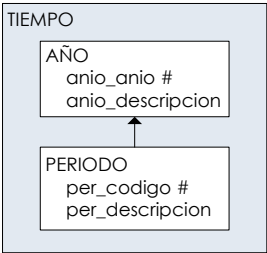


Figura No. 4.14: “Representación Gráfica Dimensión Tiempo”

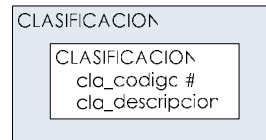


Figura No. 4.15: "Representación Gráfica Dimensión Clasificación"

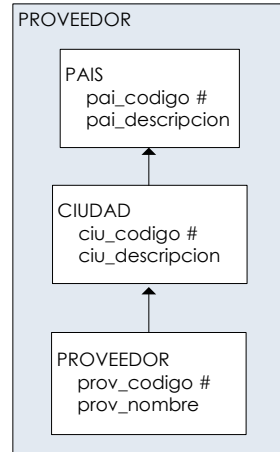


Figura No. 4.16: "Representación Gráfica Dimensión Proveedor"

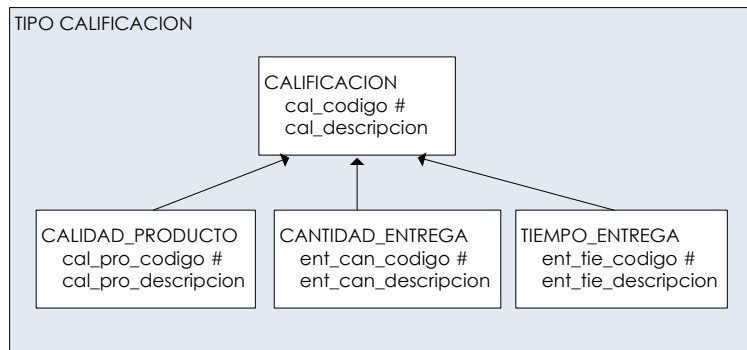


Figura No. 4.17: "Representación Gráfica Dimensión Tipo Calificación"

Dimensión Calificación: La dimensión calificación representa la medida y posee un solo nivel compuesto por los atributos cal_calidad, cal_cantidad_entrega, cal_tiempo_entrega, cal_proveedor que muestran la calificación de los proveedores.

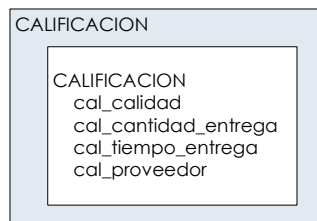


Figura No. 4.18: "Representación Gráfica Dimensión Tipo Calificación"

Relación Dimensional Calificación Proveedor: en la figura 4.19 se muestra el cruce de las dimensiones obtenidas en una única relación dimensional.

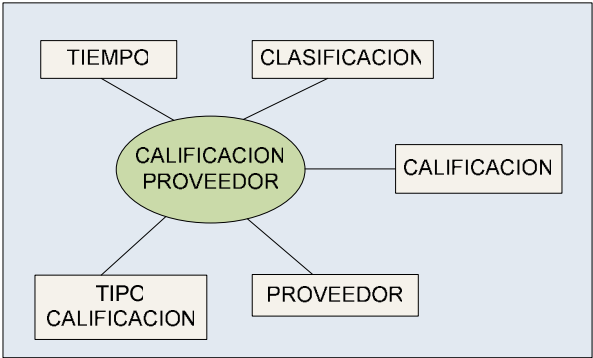


Figura No. 4.19: “Representación Gráfica de la Relación Dimensional Calificación Proveedor”

- Departamento de Talento Humano: en este departamento existen dos requerimientos, obtener información de acuerdo a las faltas y permisos, y de acuerdo a las remuneraciones. En las siguientes figuras se muestran las dimensiones obtenidas para cumplir con estos requerimientos y las relaciones dimensionales resultantes.

Faltas y Permisos: como resultado del análisis para este requerimiento se obtuvieron cinco dimensiones.

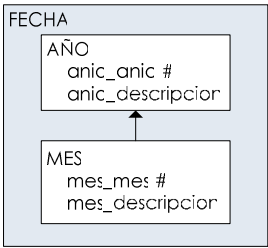


Figura No. 4.20: “Representación Gráfica Dimensión Fecha”

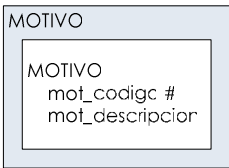


Figura No. 4.21: “Representación Gráfica Dimensión Motivo”

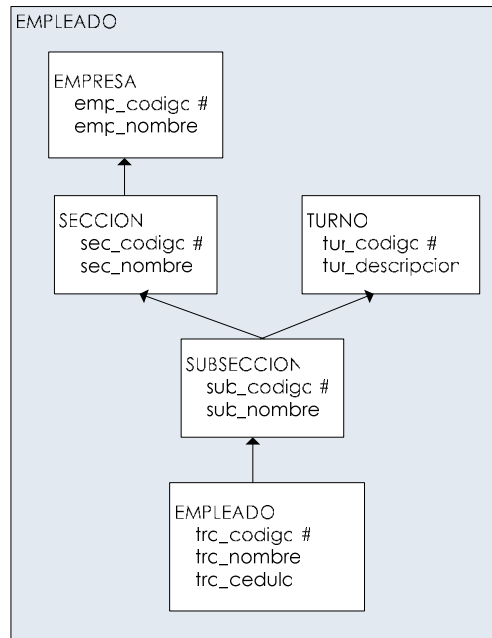


Figura No. 4.22: "Representación Gráfica Dimensión Empleado"

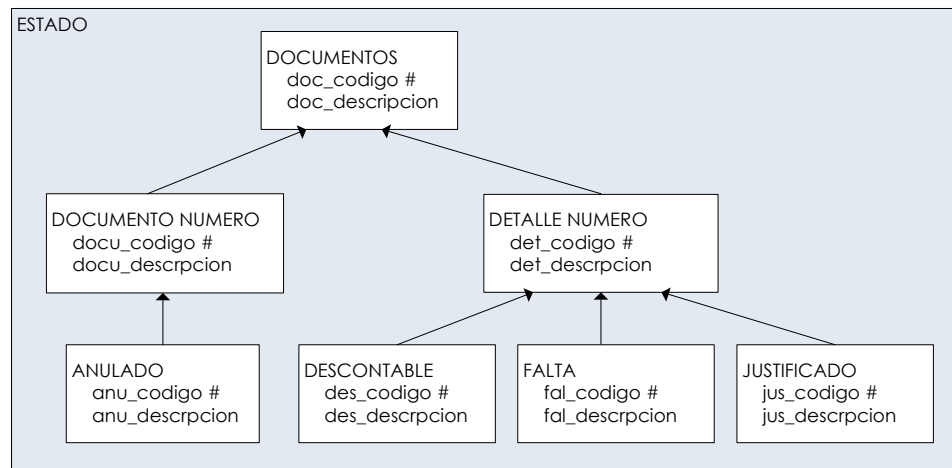


Figura No. 4.23: "Representación Gráfica Dimensión Estado"

Dimensión Permisos: Esta dimensión representa la medida y posee un nivel compuesto por el atributo `num_horas` que representa el total en horas que se han dado en permisos y faltas, y el atributo `cant_permisos` que nos muestra la cantidad de permisos y faltas que se han concedido a los empleados.

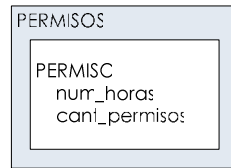


Figura No. 4.24: "Representación Gráfica Dimensión Permisos"

Relación Dimensional Faltas y Permisos: El resultado del cruzamiento de las dimensiones antes definidas nos da la siguiente relación dimensional mostrada en la figura 4.25.

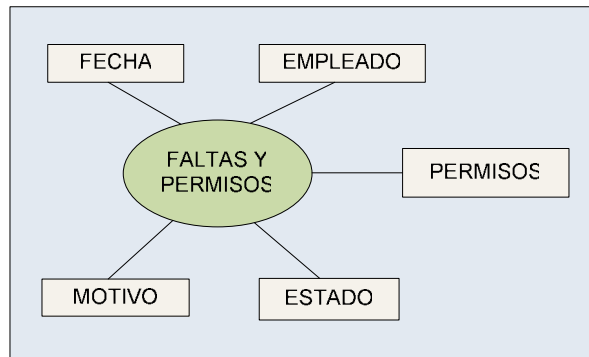


Figura No. 4.25: "Representación Gráfica Relación Dimensional Faltas y Permisos"

Remuneraciones: al analizar la información necesaria para obtener este requerimiento se obtuvieron tres dimensiones.

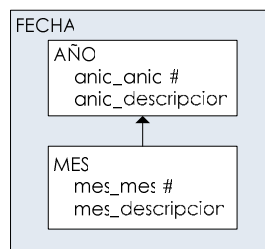


Figura No. 4.26: "Representación Gráfica Dimensión Fecha"

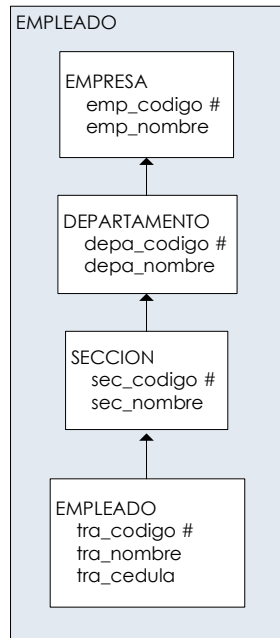


Figura No. 4.27: “Representación Gráfica Dimensión Empleado”

Dimensión Remuneración: en nuestro análisis consideramos esta dimensión como la medida para nuestro requerimiento, y tiene un nivel con el atributo valor_remun que representa el valor en dólares de las remuneración.

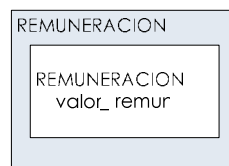


Figura No. 4.28: “Representación Gráfica Dimensión Remuneración”

Relación Dimensional Remuneraciones: La relación dimensional en donde intervienen las tres dimensiones antes definidas se muestra en la figura 4.29

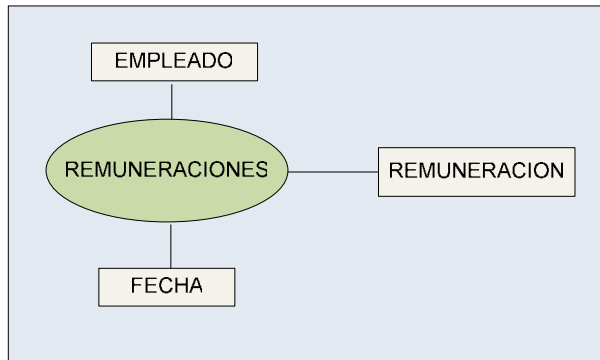


Figura No. 4.29: "Representación Gráfica Relación Dimensional Remuneraciones"

4.2.3. Diseño Lógico Específico

El objetivo de esta etapa es construir el esquema lógico tomando en cuenta el esquema conceptual construido en la fase anterior, y las estrategias para resolver los requerimientos de almacenamiento y performance. El procedimiento a realizar será dividido en dos fases.

- Definición de lineamientos y mapeos a la base fuente: en esta fase el diseñador indica algunos lineamientos que complementan al esquema conceptual, por ejemplo: criterios para fragmentar datos históricos o qué datos almacenar juntos; estos lineamientos más el esquema conceptual conforman el esquema intermedio. Luego de esto, el diseñador establece mapeos o correspondencias entre el esquema intermedio y la base de datos fuente, es decir, indicar de que datos fuentes se tomara la información para cargar al DWH.
- Generación del esquema lógico: se especifica como se almacenará el DWH, ya sea utilizando el esquema copo de nieve o estrella.

4.2.3.1. Definición de lineamientos y mapeos

4.2.3.1.1 Lineamientos

Son pautas que especifica el diseñador, como el estilo que tendrá el DWH (estrella, copo de nieve, etc) y la forma de almacenar la información. Existen

varios tipos de lineamientos, pero describiremos los que utilizaremos en el proyecto.

4.2.3.1.1.1 Materialización de Relaciones

En el modelo conceptual, una relación representa el conjunto de cubos que se pueden obtener del cruce de dimensiones. Por lo tanto, con este lineamiento el diseñador indica que cubos se van a materializar, es decir, cuales se van a implementar y almacenar físicamente en el DWH.

Un cubo esta formado por un nombre, una relación a la cual materializa, un conjunto de niveles que forman su nivel de detalle, u opcionalmente un nivel que es elegido como medida.¹⁶

Para nuestro proyecto materializaremos los siguientes cubos de acuerdo a cada departamento donde se implementa el DWH.

- Ventas: Para la relación dimensional "Ventas", se decide materializar los siguientes cubos.
 - Con detalle de Año, Mes, Documento, Sucursal, Grupo, Línea, Producto
 - Con detalle de Año, Mes, Tipo de Cliente, Ciudad, Activo
 - Con detalle de Año, Mes, Documento, Sucursal, Grupo, Línea, Producto, Tipo de Cliente

La figura 4.30 muestra gráficamente el primer cubo de la relación dimensional Ventas, los demás cubos se muestran en el anexo No. 4.2

¹⁶ Peralta, Verónica. "Diseño Lógico de Data Warehouse a partir de esquemas conceptuales multidimensionales". Noviembre, 2001. <http://www.fing.edu.uy/vperalta.pdf>. [consulta 4 de diciembre de 2007]

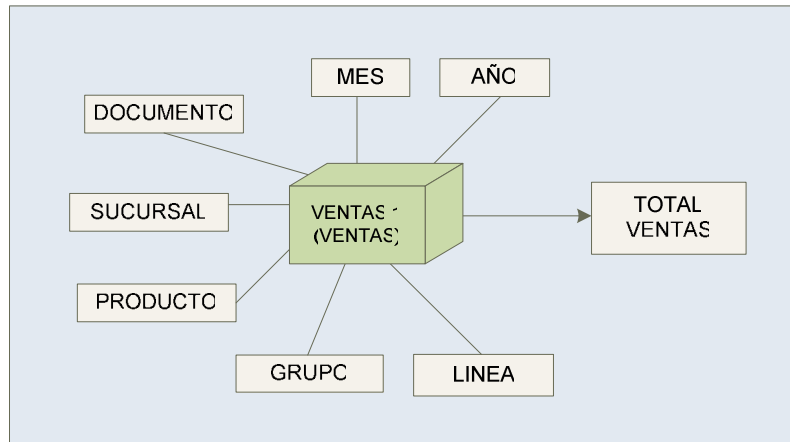


Figura No. 4.30: “Materialización de la Relación Ventas, Cubo 1”

- Logística: En la relación dimensional “Calificación Proveedor” se materializaran los siguientes cubos.
 - Con detalle de Clasificación, Año, Periodo, País, Ciudad, Proveedor
 - Con detalle de Clasificación, Año, Periodo, Calidad Producto, Entrega a Tiempo, Entrega Cantidad, Proveedor

En la figura 4.31 se muestra la representación gráfica de la materialización del primer cubo, el resto de cubos se pueden observar en el anexo No. 4.3.

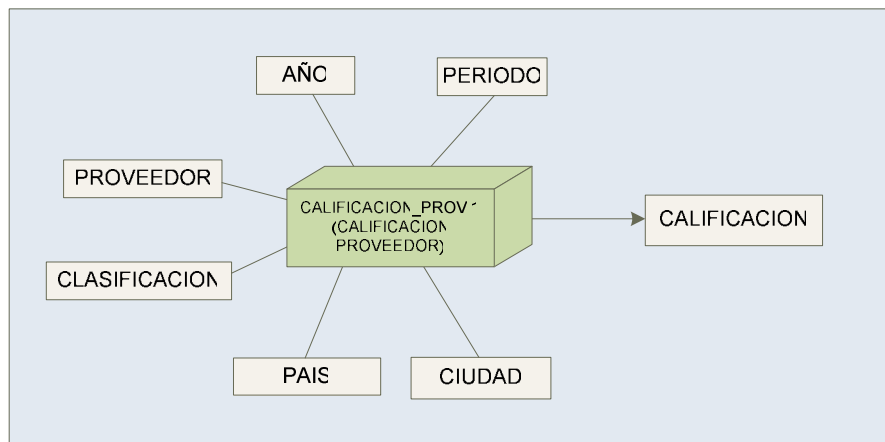


Figura No. 4.31: “Materialización de la Relación Calificación Proveedor, Cubo 1”

- Talento Humano: en este caso tenemos dos relaciones dimensionales, para la primera relación "Faltas y Permisos", se materializan los siguientes cubos:

- Con detalle de Año, Mes, Empresa, Sección, Subsección
- Con detalle de Año, Mes, Motivo, Justificado, Falta
- Con detalle de Año, Mes, Empresa, Sección, Subsección, Motivo, Justificado

A continuación se encuentra el grafico del primer cubo, los cubos restantes pueden observarse en el anexo No. 4.4.

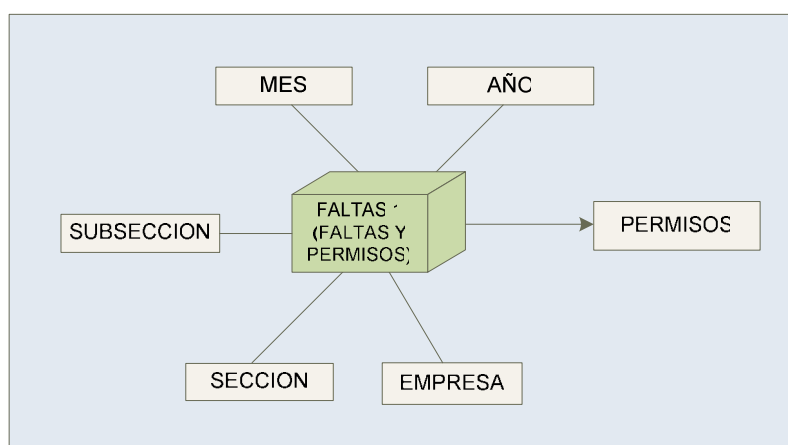


Figura No. 4.32: "Materialización de la Relación Faltas y Permisos, Cubo 1"

Y para la relación dimensional "Remuneraciones" se implementa el siguiente cubo, cuya representación gráfica se muestra en la figura 4.33.

- Con detalle de Año, Mes, Empresa, Departamento, Sección.

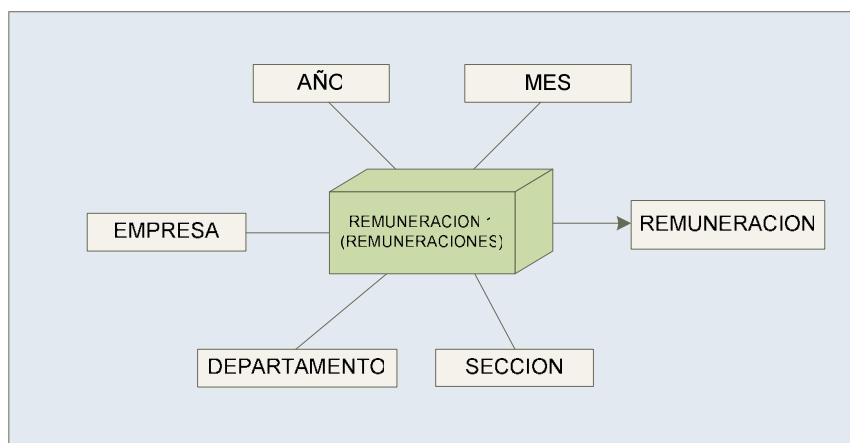


Figura No. 4.33: "Materialización de la Relación Remuneraciones, Cubo 1"

4.2.3.1.1.2 Fragmentación vertical de Dimensiones

En este lineamiento el diseñador decide como se van a almacenar los niveles de las dimensiones especificadas en el esquema conceptual. Se los puede desnormalizar (esquema estrella), normalizar (copo de nieve) o aplicar una estrategia diferente para cada dimensión, ya sea normalizar, desnormalizar o indicar que niveles se guardaran juntos en la tabla.

Un fragmento es un subconjunto de los niveles de la dimensión, por lo tanto, para que la fragmentación se realice correctamente, debemos tomar en cuenta que esta debe ser completa, es decir que todos los niveles deben estar en al menos un fragmento para no perder información.

Gráficamente representamos una fragmentación como una coloración de niveles, los niveles de un mismo fragmento se recuadran con el mismo color. La fragmentación es completa si todos los niveles tienen color¹⁷.

Para las dimensiones creadas en nuestro proyecto decidimos aplicar las siguientes estrategias:

- Departamento de Ventas: en la figura 4.34 se puede observar gráficamente la fragmentación realizada en las dimensiones de acuerdo a lo siguiente:
 - Fecha: Desnormalizado.
 - Producto: Tres fragmentos, grupo, línea y producto como tal.
 - Cliente: Cuatro fragmentos, ciudad, tipo cliente, activo y cliente como tal.
 - Sucursal: Un fragmento, sucursal.
 - Documentos: Tres fragmentos, documento, factura, secuencia detalle.
 - Total Ventas: Al ser la medida no se la fragmenta.

¹⁷ Peralta, Verónica. "Diseño Lógico de Data Warehouse a partir de esquemas conceptuales multidimensionales". Noviembre, 2001. <http://www.fing.edu.uy/vperalta.pdf>. [consulta 13 de diciembre de 2007]

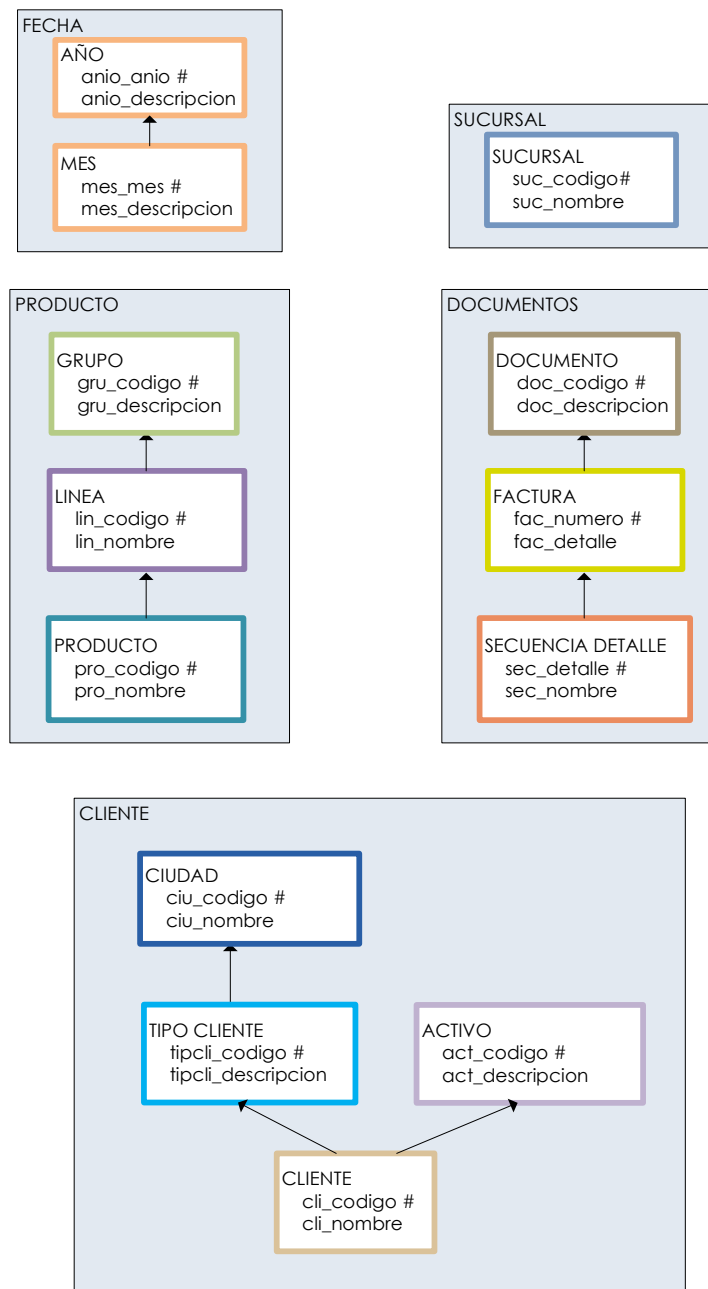


Figura No. 4.34: "Fragmentación Vertical de Dimensiones – Requerimiento Dep. Ventas"

- Departamento de Logística: la representación gráfica de la fragmentación de las dimensiones que participan en este requerimiento se la puede observar en la figura 4.35.
 - Tiempo: Desnormalizado.
 - Clasificación: Un fragmento, clasificación.

- Tipo Calificación: Desnormalizado.
- Proveedor: Tres fragmento, país, ciudad y proveedor como tal.
- Calificación: No se aplica la fragmentación por ser la medida.

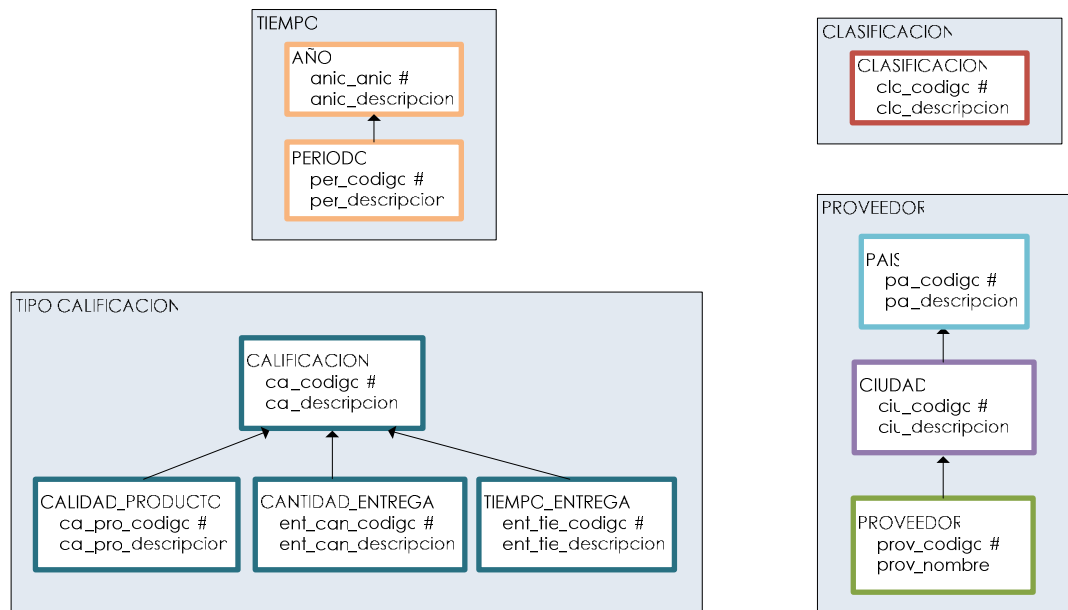


Figura No. 4.35: "Fragmentación Vertical de Dimensiones – Requerimiento Dep. Logística"

- Departamento de Talento Humano: en este departamento tenemos dos requerimientos.

Faltas y Permisos: en la figura 4.36 tenemos la fragmentación vertical realizada en las dimensiones para este requerimiento de acuerdo a las siguientes estrategias.

- Fecha: Desnormalizado.
- Empleado: Cinco fragmentos, empresa, sección, turno, subsección y empleado.
- Estado: Ocho fragmentos, documentos, documento numero, anulado, detalle numero, descontable, falta y justificado.
- Motivo: Un fragmento, motivo.
- Permisos: Será utilizada como la medida, por lo tanto no se la fragmenta.

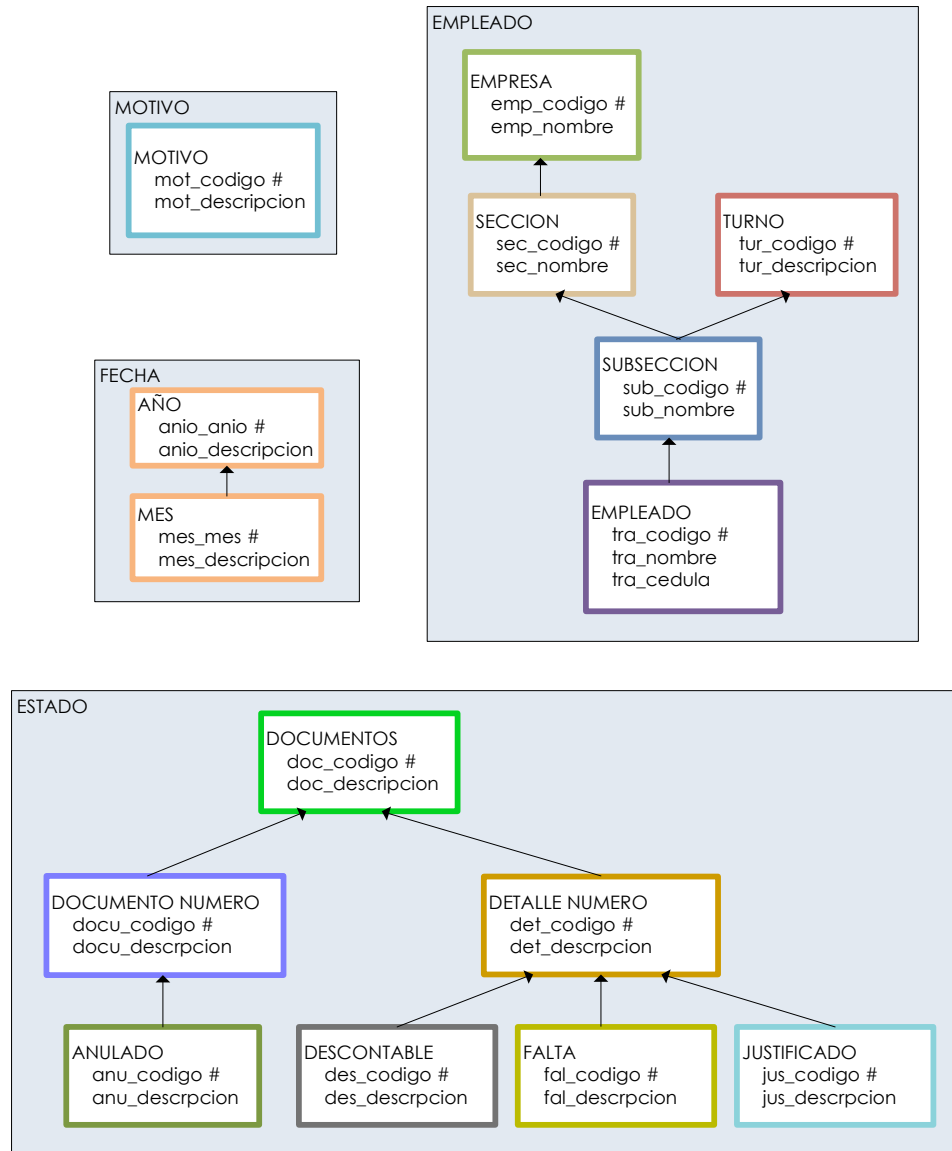


Figura No. 4.36: “Fragmentación Vertical de Dimensiones – Requerimiento Faltas y Permisos Dep. Talento Humano”

Remuneraciones: en este requerimiento las dimensiones fueron fragmentadas de acuerdo a lo siguiente, y su representación gráfica se la puede observar en la figura 4.37.

- Fecha: Desnormalizado.
- Empleado: Cuatro fragmentos, empresa, departamento, sección y empleado como tal.
- Remuneración: No se implementa porque es la medida.

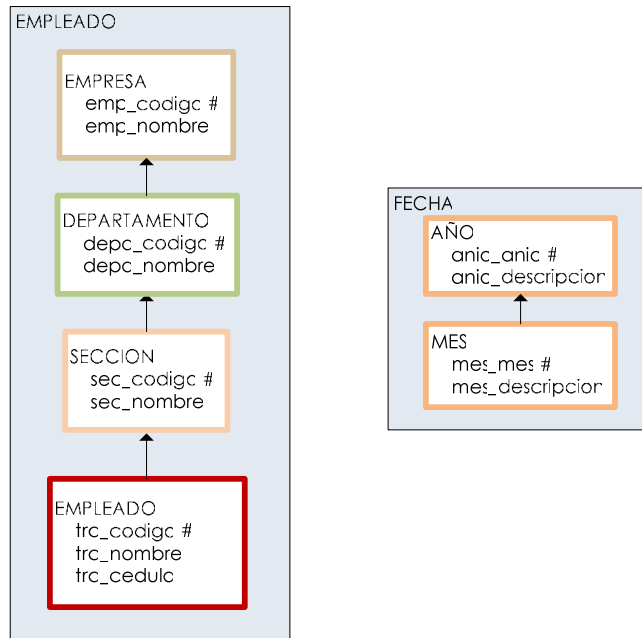


Figura No. 4.37: “Fragmentación Vertical de Dimensiones – Requerimiento Remuneraciones Dep. Talento Humano”

4.2.3.1.1.3 Fragmentación horizontal de Cubos

Para mejorar la performance al momento de realizar consultas en tablas de hechos con gran cantidad de información, se las puede fragmentar horizontalmente, con lo cual se obtienen tablas con la misma estructura y más pequeñas, llamadas franjas o bandas.

Para definir una fragmentación, las condiciones que distinguen una banda de otra, deben ser expresadas en términos de los ítems de los niveles del cubo¹⁸.

Para nuestro proyecto no realizaremos la fragmentación horizontal.

Como resultado de la definición de lineamientos sobre el esquema conceptual, tenemos como resultado el esquema intermedio, que consiste de los ítems, niveles, dimensiones y restricciones del esquema conceptual, e incorpora los cubos, fragmentación de dimensiones y fragmentación de cubos definidos en los lineamientos.

¹⁸ Peralta, Verónica. “Diseño Lógico de Data Warehouse a partir de esquemas conceptuales multidimensionales”. Noviembre, 2001. <http://www.fing.edu.uy/vperalta.pdf>. [consulta 15 de abril de 2007]

4.2.3.1.2 Mapeos

Son funciones que asocian los elementos del esquema intermedio con una expresión construida en base a las tablas y atributos de la base de datos fuente. Una expresión de mapeo puede ser un atributo de una tabla fuente (directo), un cálculo de varios atributos de una tupla (cálculo simple), una totalización de atributos de varias tuplas (cálculo agregado), o un valor externo, como constantes (externo).

Existen dos tipos de funciones de mapeo; los mapeos de dimensiones y los mapeos de cubos.

Gráficamente se puede representar la función de mapeo, mediante links o líneas entre los elementos del modelo conceptual y los atributos de las tablas fuente, cuando el mapeo es directo se representa con una línea corrida, cuando es un cálculo se representa con una línea cortada a cada atributo que interviene en el cálculo y se adjunta la definición del cálculo, y cuando es externo no se utilizan líneas pero se adjunta la expresión a la que mapea.¹⁹

Cabe recordar que al inicio del capítulo se indicó que El DWH de logística no tiene completamente definidas sus tablas origen, ya que se encuentra en fases de pruebas, por lo que es difícil el poder interpretar el mapeo de dimensiones. De igual manera sucede con el DWH de Remuneraciones en el departamento de Talento Humano, puesto que la información la llevan en hojas de Excel y no existen sus tablas origen.

4.2.3.1.2.1 Mapeo de dimensiones

En este caso, se define una función de mapeo para todos los ítems que conforman la dimensión. Estas funciones deben pensarse como consultas SQL para elegir correctamente a que atributo mapear los ítems.

¹⁹ Peralta, Verónica. "Diseño Lógico de Data Warehouse a partir de esquemas conceptuales multidimensionales". Noviembre, 2001. <http://www.fing.edu.uy/vperalta.pdf>. [consulta 21 de julio de 2007]

Para indicar las tablas fuentes se tomo la información que provee el SQL 2000, en la cual se muestran solo atributos que son de nuestro interés para la realización del mapeo.

- Departamento de Talento Humano.
 - Faltas y Permisos: A continuación se muestra uno de los mapeos que componen el DWH de Talento Humano, en el Anexo 4.5 se pueden observar el resto de mapeos de este departamento.

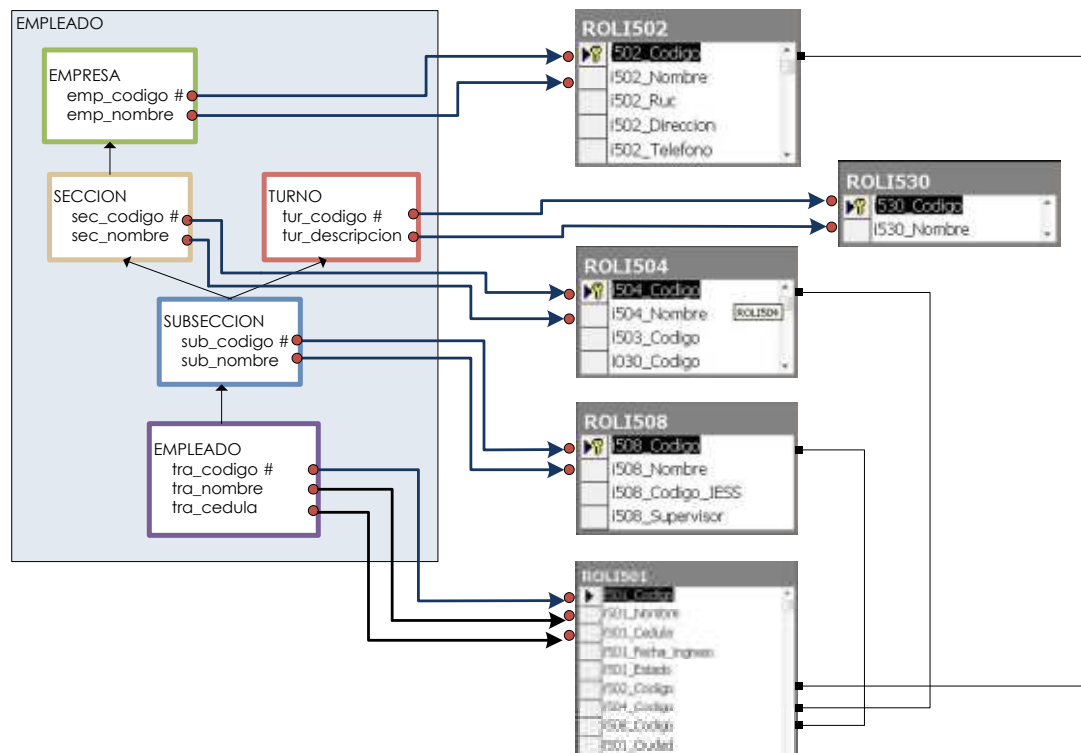


Figura No. 4.38: "Mapeo de la Dimensión Empleado. Dep. Talento Humano"

- Departamento de Ventas. De igual manera que el caso anterior, a continuación se muestra el mapeo de una de las dimensiones que componen este DWH, el resto de dimensiones se las podrá observar en el anexo 4.6

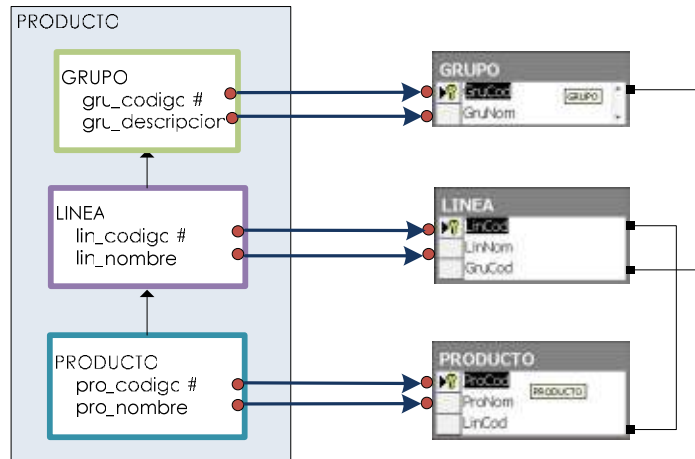


Figura No. 4.39: "Mapeo de la Dimensión Producto. Dep. Ventas"

4.2.3.1.2.2 Mapeo de cubos

Se lo puede realizar de dos formas, mapeo base, que consiste en especificar una función de mapeo y una condición, o con un mapeo recursivo, en donde se definen una secuencia de drill-ups y una función de mapeo base, sin dejar de lado las operaciones roll-up asociadas a las medidas en cualquiera de los dos casos.

- Departamento de Talento Humano. A continuación se muestra el mapeo de cubos del DWH de Faltas y Permisos (Fig. 3.40) y de Remuneraciones (Fig. 3.41). En el anexo 4.6 se muestra el drill-up de Faltas y Permisos.

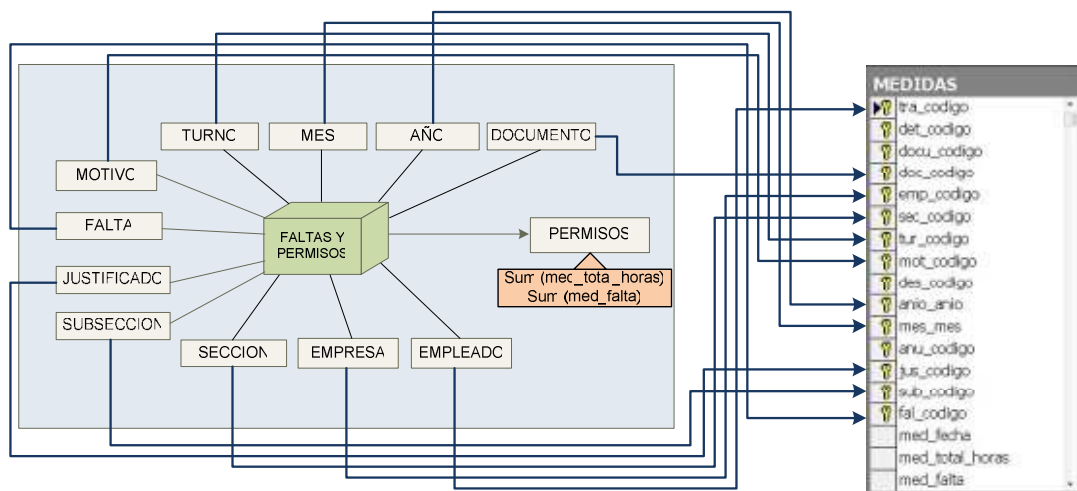


Figura No. 4.40: "Mapeo de Cubos Faltas y Permisos. Dep. Talento Humano"

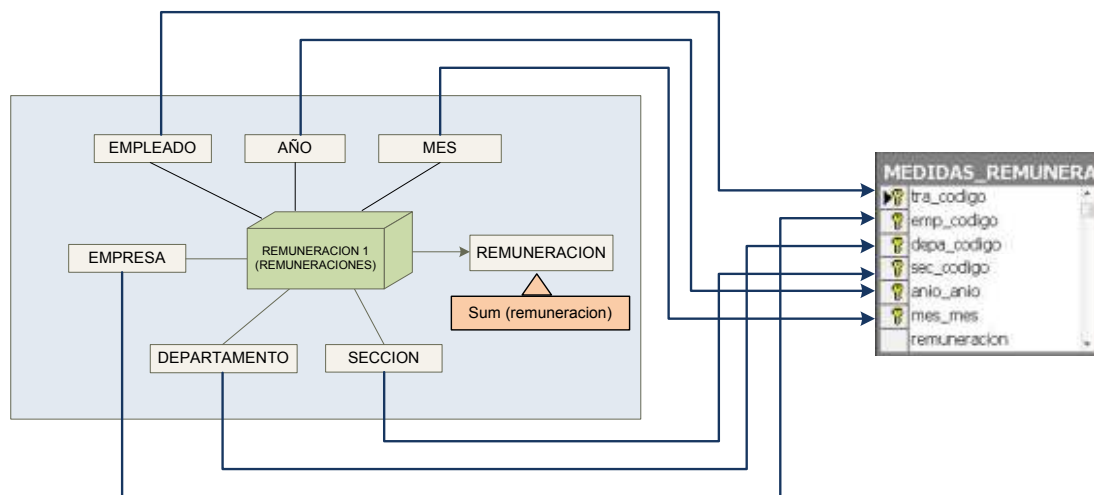


Figura No. 4.41: "Mapeo de Cubos Remuneraciones. Dep. Talento Humano"

- Departamento de Ventas. A continuación se muestra el mapeo de cubos de este departamento. En el anexo 4.8 se puede observar el drill-up de Ventas.

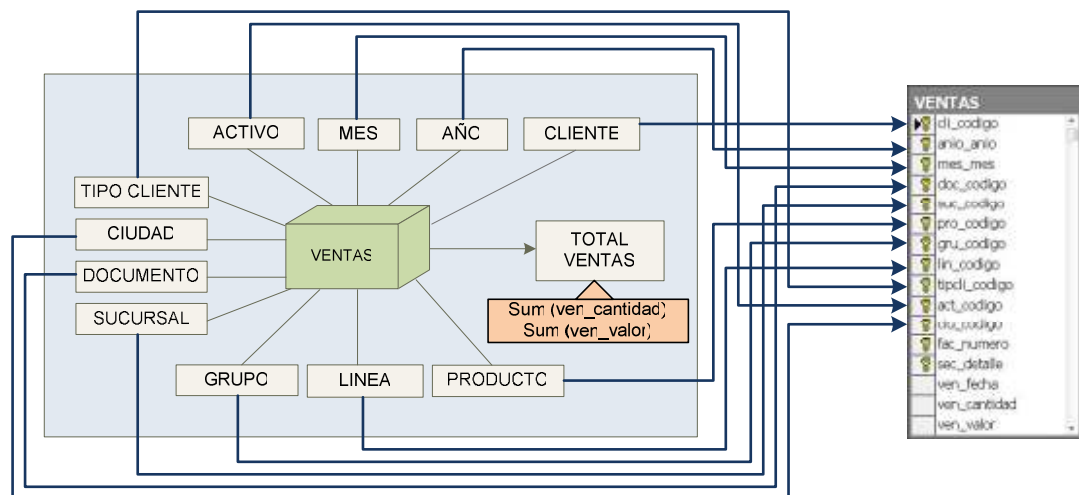


Figura No. 4.42: "Mapeo de Cubos Ventas. Dep. Ventas"

- Departamento de Logística. A continuación se muestra el mapeo de cubos de este departamento. En el anexo 4.9 se puede observar el drill-up de Logística.

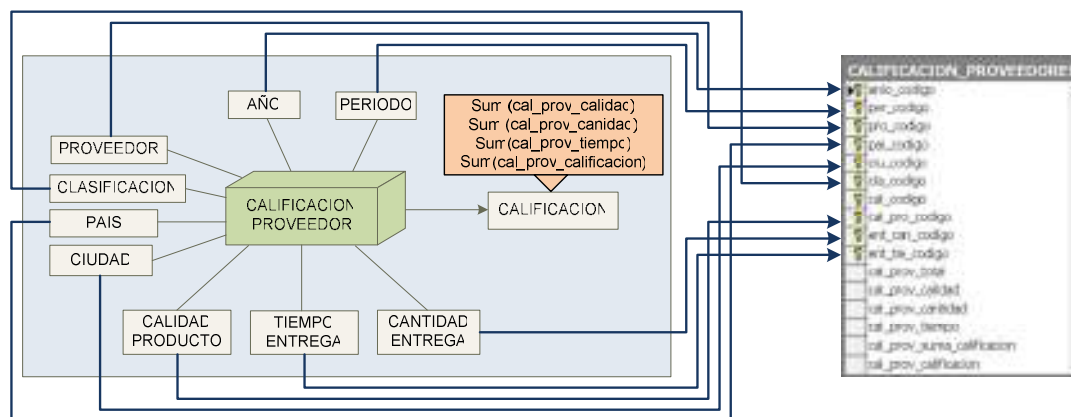


Figura No. 4.43: "Mapeo de Cubos Logística. Dep. Logística"

4.2.3.2. Esquema Lógico

Una vez que se han diseñado los cubos, dimensiones y medidas, se procede a realizar el esquema lógico que muestra como esta formado nuestro DWH.

4.2.3.2.1 Esquema lógico del departamento de Talento Humano: Aquí se puede observar la estructura del DWH de Talento Humano, con todas sus medidas y dimensiones.

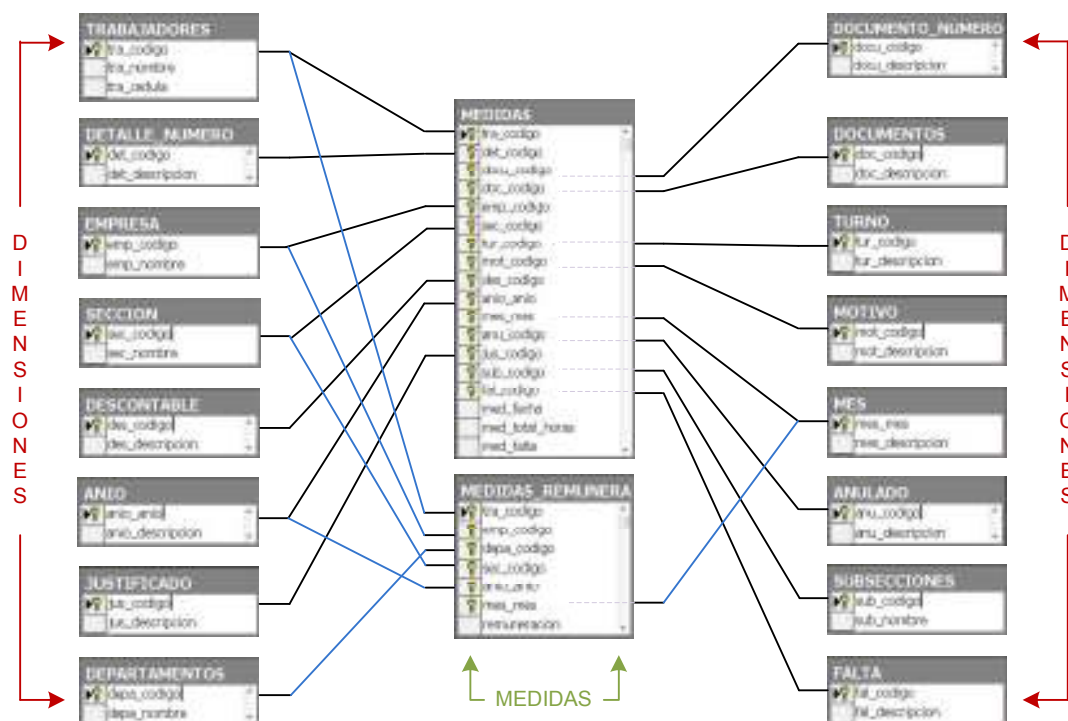


Figura No. 4.44: "Esquema Lógico del DWH del departamento de Talento Humano"

4.2.3.2.2 Tabla de hecho Permisos: en la figura 4.45 podemos observar como esta tabla se relaciona con las dimensiones por medio de una estructura en estrella. Este esquema nos permite realizar todo tipo de reportes especificados en la sección “3.2.3.1 Definición de lineamientos y mapeos” de este capítulo.

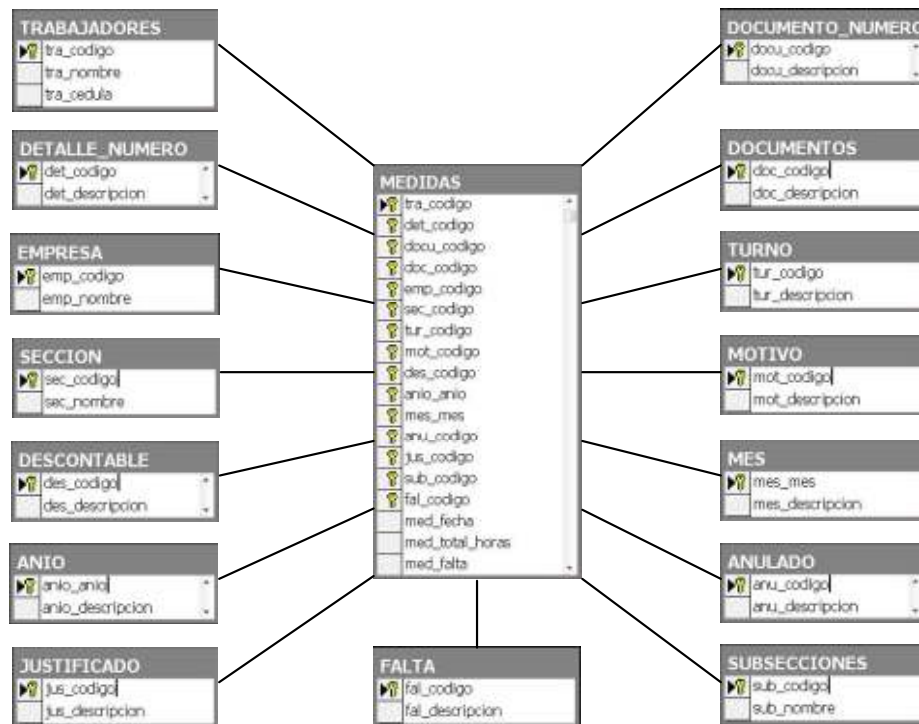


Figura No. 4.45: “Modelo Multidimensional en Estrella, Tabla de hecho Permisos”

4.2.3.2.3 Tabla de hecho de Remuneraciones: con el esquema en estrella que se muestra en la figura 4.46 podemos realizar los reportes solicitados en el punto “4.2.3.1 Definición de lineamientos y mapeos” de este capítulo para el departamento de Talento Humano con respecto a las remuneraciones.

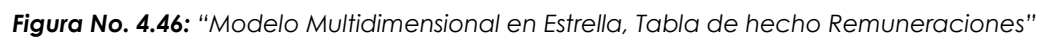
[illegible]

Figura No. 4.47: “Esquema Lógico del DWH del departamento de Ventas”

77

multidimensional de tipo estrella. A continuación se puede observar su medida y sus dimensiones.

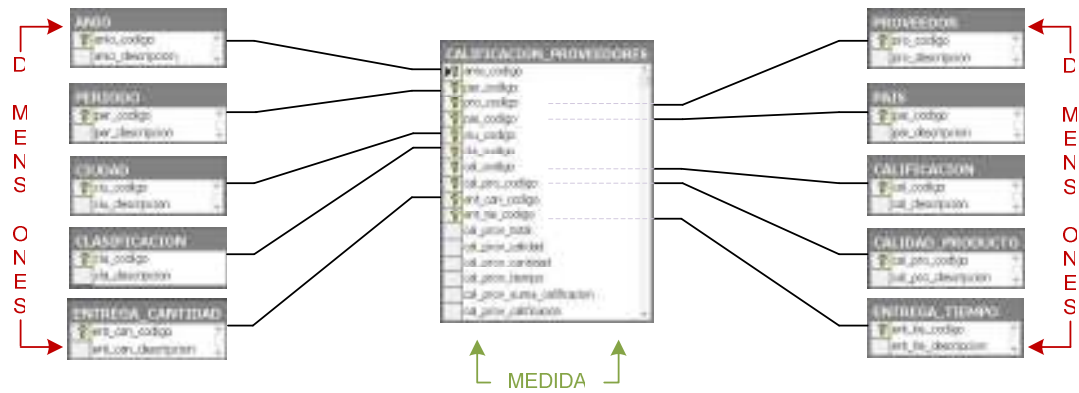


Figura No. 4.48: "Esquema Lógico del DWH del departamento de Logística"

4.2.4. Diseño Lógico General

También llamado diseño físico. En esta etapa del proceso se especifica como se implementara el esquema lógico en el manejador de base de datos, SQL Server 2000, y se preparan las vistas del usuario.

Para nuestro proyecto hemos definido dos etapas que engloban este procedimiento.

4.2.4.1. Diseño de las ETL

Como se explico en el capítulo uno las ETL nos permiten determinar el proceso de extracción, transformación y carga de los datos.

- Extracción: una vez realizado el análisis de las fuentes de información existente y de especificar que datos serán necesarios cargar en nuestro DWH, realizamos la extracción de la información de las fuentes internas utilizando objetos propios de la herramienta Genexus (Data Views), y para el caso de la información externa se desarrollo una aplicación dentro del DHW que permite extraer la información de hojas de Excel a nuestro almacén de datos.

- Transformación: no existió la necesidad de transformar los datos provenientes tanto de las fuentes internas como externas ya que esta información mantiene una estandarización, y nuestro DWH se rigió a las mismas normas que maneja la empresa.
- Carga de Información: para un mejor manejo de la carga de la información, en cada DWH se creo una aplicación que permite especificar un rango de fechas de carga con el propósito de cumplir a cabalidad los requerimientos de la empresa según cada departamento.

4.2.4.2. Herramientas Front End

Para que los usuarios puedan explotar la información almacenada en el DWH se utilizara la herramienta GXplorer, la cual funciona como un complemento de Microsoft Excel y permite de una manera sencilla y dinámica realizar todo tipo de reportes.

4.2.5. Implementación

En esta etapa del desarrollo de nuestro DWH realizaremos la implementación del diseño lógico general especificado anteriormente, cabe recalcar que la implementación será indicada solamente del departamento de Talento Humano, ya que este departamento contiene la mayor cantidad de casos que se pueden encontrar para la recolección de datos.

4.2.5.1. Extracción, transformación y carga:

Como se había explicado en la etapa anterior, para realizar este proceso utilizaremos la herramienta Genexus, la cual nos permitirá generar la aplicación para realizar la carga de información al DWH, a continuación explicamos paso a paso lo que se realizo en este punto.

- Primero debemos crear las dimensiones y medidas para nuestro DWH por medio de la creación de un objeto tipo transacción y especificando el tipo de elemento de la estructura del DWH, como se muestra en las figuras 4.49, 4.50 y 4.51.

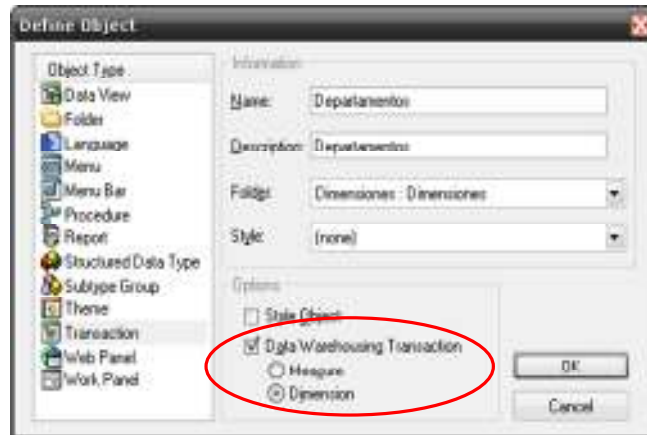


Figura No. 4.49: “Elección del objeto Dimensión o Medida”

GeneXus - [Transaction - Departamentos : Departamentos]

File Edit View Object Insert Advanced Build Tools Knowledge Manager Window Help

Design D'WH_THumano Spanish (default)

Structure	Type	Description	Nulls
Departamentos			
depa_codigo	Character(3)	Codigo Departamento	No
depa_nombre	VarChar(40)	Nombre Departamento	No

Figura No. 4.50: “Estructura de la Dimensión Departamentos”

Structure	Type	Description	Nulls
Medidas			
tra_codigo	Character(13)	Código Trabajador	No
det_codigo	Numeric(9.0)	Número Detalle	No
docu_codigo	Numeric(9.0)	Número Documento	No
doc_codigo	Character(2)	Código Documento	No
emp_codigo	Character(3)	Código Empresa	No
sec_codigo	Character(3)	Código Sección	No
tur_codigo	Character(3)	Código Turno	No
mot_codigo	Character(3)	Código Motivo	No
des_codigo	Character(1)	Descontable	No
anio_anio	Numeric(4.0)	Año	No
mes_mes	Numeric(2.0)	Mes	No
anu_codigo	Character(1)	Código Anulado	No
jus_codigo	Character(1)	Código Justificación	No
sub_codigo	Character(3)	Código Subsección	No
fal_codigo	Character(1)	Código Falta	No
med_fecha	Date	Fecha	No
med_total_horas	Numeric(6.2)	Total en Horas (Permisos)	No
med_falta	Numeric(1.0)	Faltas (Permisos)	No

Figura No. 4.51: “Estructura de la Medida Permisos”

- Como los datos que necesitamos para nuestro DWH se encuentran en diferentes bases de datos, Genexus nos permite utilizar estas tablas como si fueran parte de nuestra base de conocimiento, para lograr esto se utilizan los objetos conocidos como Data View (Vista de una tabla específica en una base de datos específica, la cual permite consultar información externa en nuestra aplicación). Para la creación de un Data View se tienen que realizar 3 acciones:
 - Crear un objeto transacción. Normalmente se usa la misma estructura de la tabla de la base de datos externa (figura 4.52).
 - Crear y especificar la estructura del Data View. Creamos el objeto y luego se indican los índices externos para realizar una mejor conexión con la base de datos externa, de igual manera se indica la plataforma y su estructura como se indica en la Figura 4.53.

- Especificar la conexión del Data Store. Se indica en donde se encuentra la base de datos externa y a que tabla queremos acceder (Figura 4.54).

TRANSACCION PARA EL DATA
VIEW

SISTEMA
OPERACIONAL

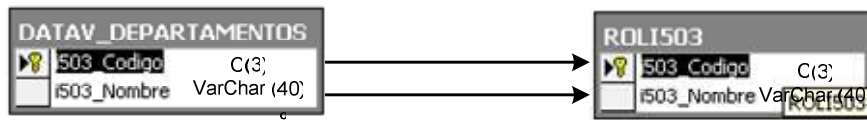


Figura No. 4.52: "Objeto transacción para el Data View y tabla externa"

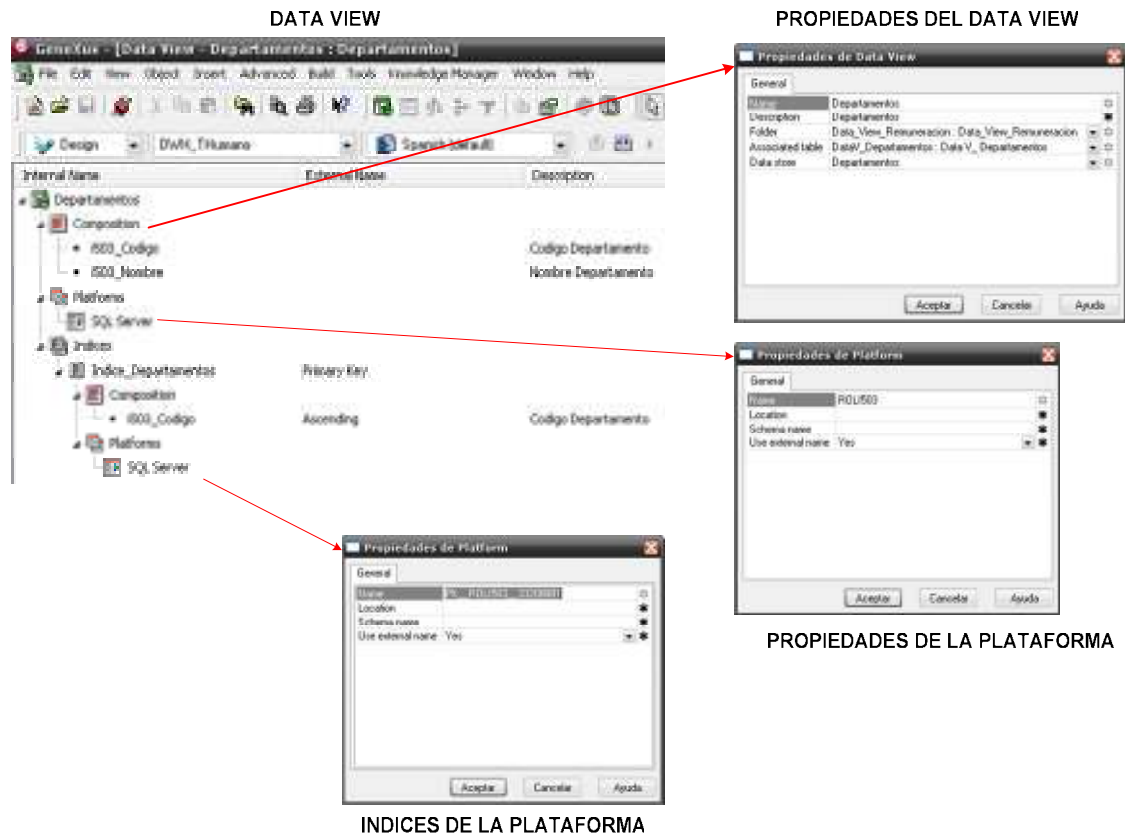


Figura No. 4.53: "Creación y estructura de un Data View"

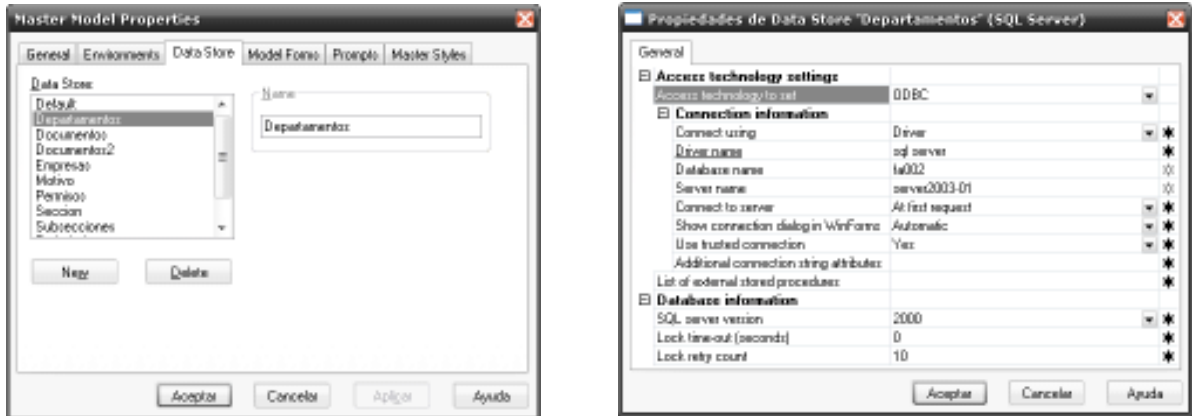


Figura No. 4.54: “Creación y especificación de un Data Store”

- Para el caso de la medida de remuneraciones es necesario a más de la creación de los Data View el crear un algoritmo el cual nos permita importar información desde un archivo, en nuestro caso es un archivo de Excel que contiene los datos ya filtrados para ser cargados al DWH. En la figura 4.55 se muestra la parte del algoritmo para la carga de los datos.

```

archivo.Open(spath)
if archivo.ErrorCode <> 0 //abrimos el archivo
    MsgBox(archivo.Description)
else
    spos = 199
    archivo.SelectSheet("datos") //seleccionamos la hoja que queremos leer
    abanderita = 0
    afila = 4
    Do while abanderita = 0
        afila += 1
        amostiar = archivo.Cells(afila, 0).Text
        if len(trim(amostiar)) > 1 //para asegurarnos que existan datos leamos la primera columna y vemos si tiene mas de 1 caracter
            No
            tra_codigo = atr_codigo
            emp_codigo = emp_codigo
            depe_codigo = archivo.Cells(afila, 05).Text
            sec_codigo = archivo.Cells(afila, 06).Text
            anio_anio = anio
            msc_mes = mes
            remuneracion = archivo.Cells(afila, 03).Number
            MsgBox "INGRESANDO REGISTRO REMUNERACION --> " + amostiar, vbWait
        else
            exit
        end if
    loop
    archivo.Close()
endif
RETURN

```

Figura No. 4.55: “Algoritmo para la importación de información desde un archivo de Excel”

- Una vez creada la base de datos y los objetos necesarios para el DWH, está lista nuestra aplicación para realizar la carga de los datos, en la

figura 4.56 se muestra el menú principal de la aplicación que será implementada.



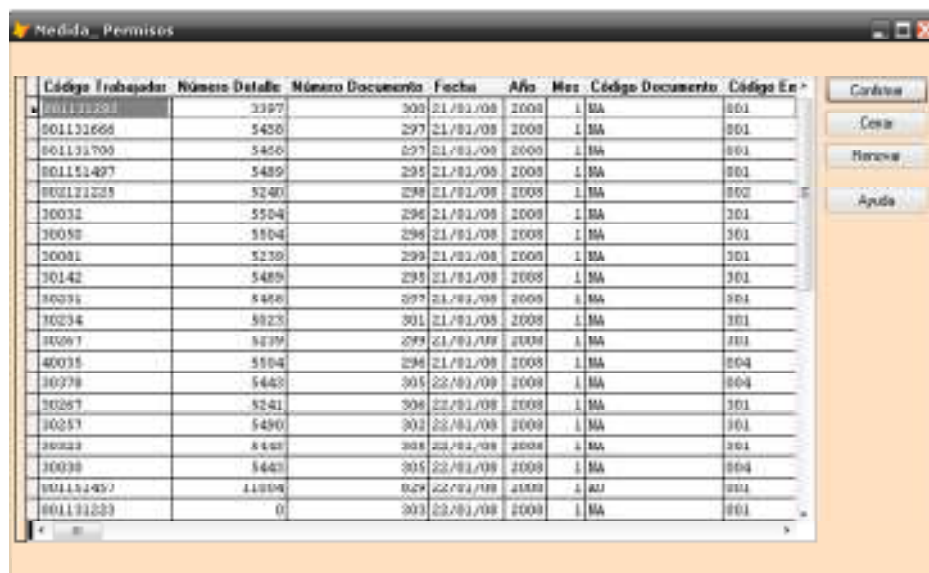
Figura No. 4.56: "Menú principal de la aplicación del DWH"

Una vez escogida la opción de cargar DWH en este caso de permisos, ingresamos el rango de fechas para la carga de información y confirmamos



Figura No. 4.57: "Pantalla para el ingreso de las fechas de carga al DWH"

Al finalizar la carga tendremos una pantalla que nos permite observar mediante una grilla la información que se encuentra ya en nuestro almacén de datos (Figura 4.58).



Código Trabajador	Número Detalle	Número Documento	Fecha	Año	Mes	Código Documento	Código Em
801131291	3397	300	21/01/08	2008	1	MA	801
801131666	5426	297	21/01/08	2008	1	MA	801
801131700	5426	297	21/01/08	2008	1	MA	801
801131497	5489	295	21/01/08	2008	1	MA	801
801131223	5240	298	21/01/08	2008	1	MA	802
30032	5504	296	21/01/08	2008	1	MA	301
30058	5504	296	21/01/08	2008	1	MA	301
30081	5238	299	21/01/08	2008	1	MA	301
30142	5489	295	21/01/08	2008	1	MA	301
30231	8488	297	21/01/08	2008	1	MA	804
30234	5823	301	21/01/08	2008	1	MA	301
30267	3279	299	21/01/08	2008	1	MA	301
40035	5504	296	21/01/08	2008	1	MA	804
30278	5443	305	22/01/08	2008	1	MA	804
30287	8241	308	22/01/08	2008	1	MA	301
30287	5490	302	22/01/08	2008	1	MA	301
30323	8488	308	22/01/08	2008	1	MA	301
30338	5443	305	22/01/08	2008	1	MA	804
801131497	41808	829	22/01/08	2008	1	MA	804
801131223	0	303	22/01/08	2008	1	MA	801

Figura No. 4.58: "Información del DWH"

4.2.5.2. Herramientas Front End;

Para un mejor entendimiento de cómo fue utilizado gexplorer en nuestro proyecto explicaremos brevemente los cuatro módulos que esta herramienta maneja:

- Gexplorer Manager: Es el modulo con el cual creamos la metadata que luego va a ser utilizada para la explotación de la información del DWH
- Gexplorer Settings: Este modulo nos permite configurar los parámetros generales que van a utilizar las herramientas Gexplorer OLAP, como son seteos de las medidas y dimensiones, creación de usuarios y los privilegios que tienen cada uno de estos sobre las metadatos creados anteriormente.
- Gexplorer Olap para MExcel: Funciona como un complemento de Microsoft Excel el cual permite agregar comandos y funciones necesarios para manipular el DWH.

- Gxplorer Olap para Acceso Web: permite realizar las mismas consultas que se hacen desde EXCEL pero desde Internet, pudiendo consultar en forma remota la información real del negocio.

A continuación explicamos como fueron utilizados los tres primeros módulos en la implementación de nuestro DWH.

- Una vez creada nuestra base de conocimiento con sus respectivas dimensiones y medidas es momento de crear la metadata y esto lo hacemos con ayuda del modulo Gxplorer Manager.
 - El primer paso es crear una metadata (Figura 4.59) y asignarle un nombre y una ubicación (Figura 4.60)

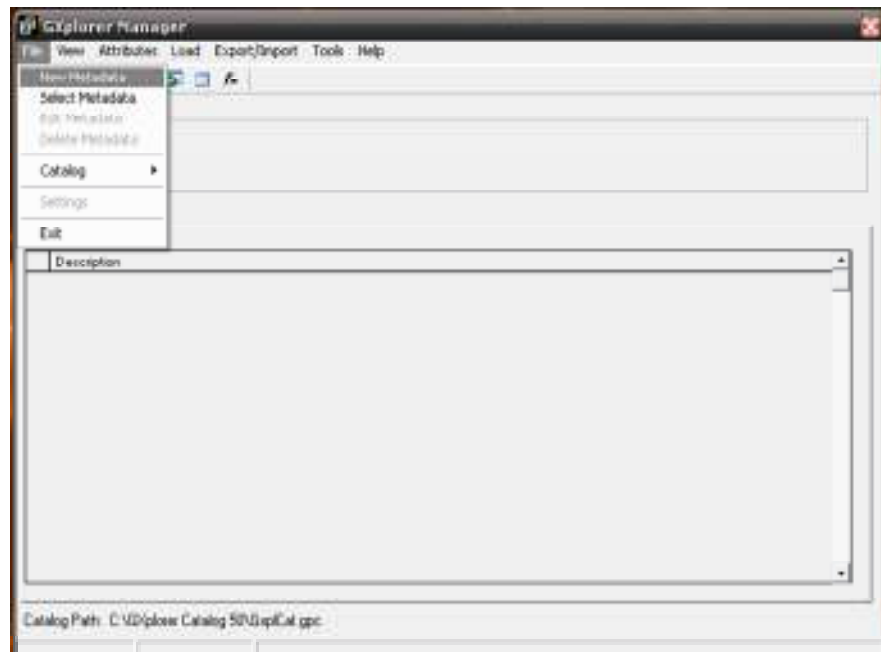


Figura No. 4.59: "Creación del Metadata"

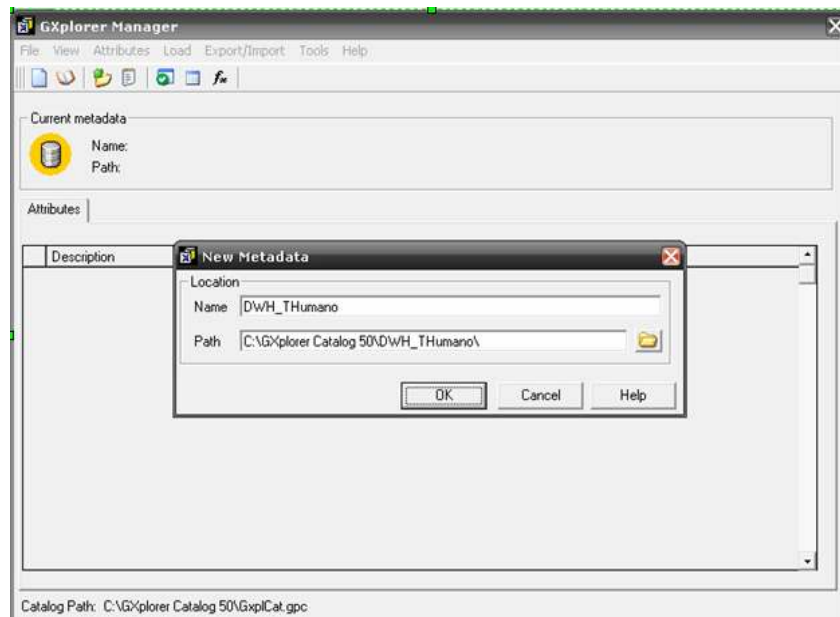


Figura No. 4.60: “Asignación de Nombre y Ubicación del Metadata”

- Una vez creada la metadata, el siguiente paso es cargarla (figura 4.61), primero indicamos el tipo de paradigma a ser cargado (para nuestro caso es de dimensiones y medidas figura 4.62), segundo elegimos la ubicación de nuestra base de conocimiento (figura 4.63) y luego elegimos el modelo creado en nuestra aplicación Genexus (figura 4.64); por ultimo debemos crear y asignar un ODBC para podernos conectar a nuestro DWH (figura 4.65) y de esta manera nuestra metadata está completamente creada (figura 4.66).

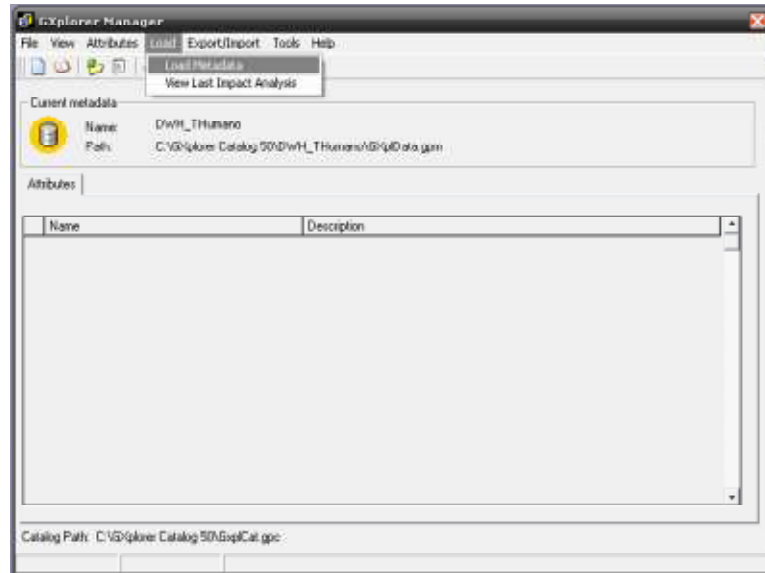


Figura No. 4.61: "Menú de carga del Metadata"

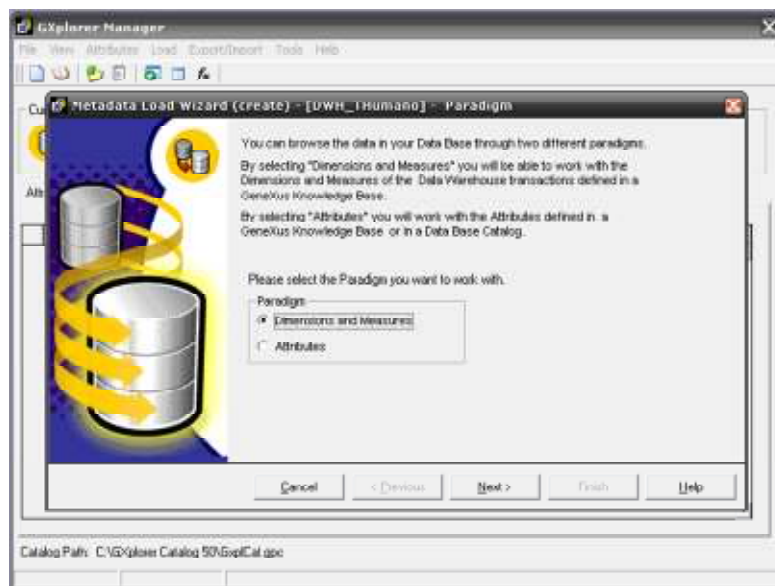


Figura No. 4.62: "Elección del paradigma"

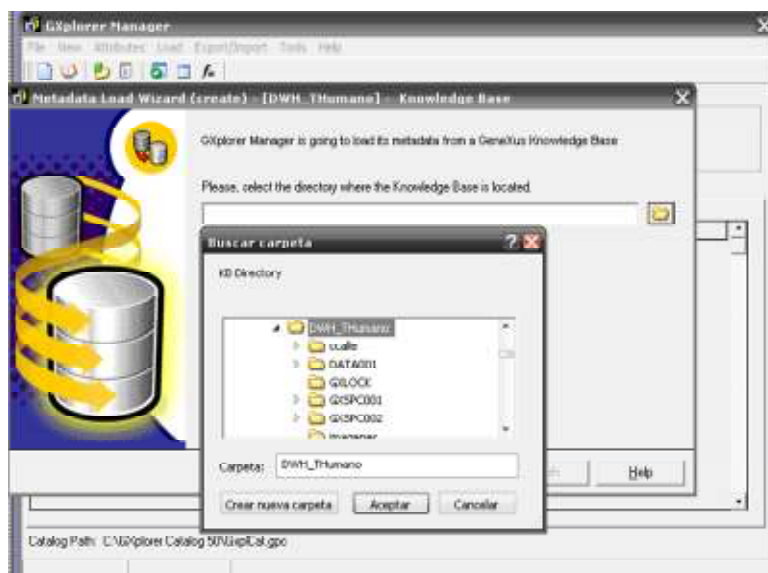


Figura No. 4.63: "Elección de la ubicación de la base de conocimiento"

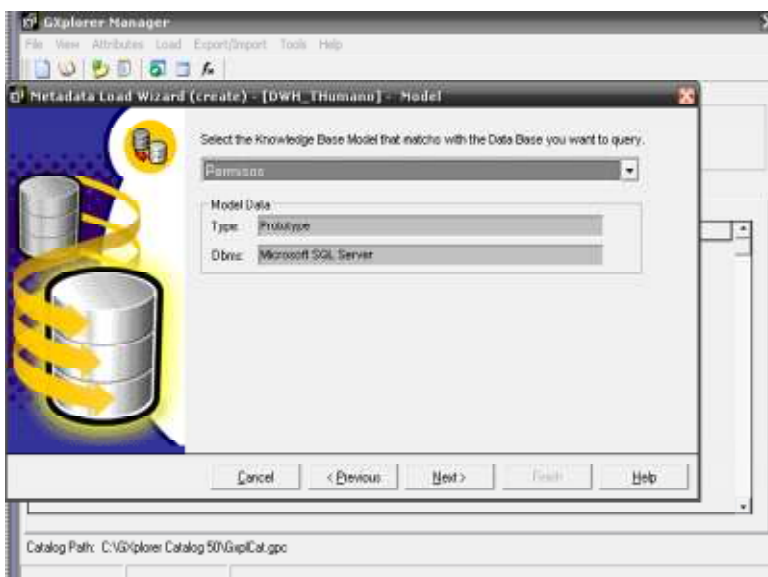


Figura No. 4.64: "Elección del modelo de la base de conocimiento Genexus"

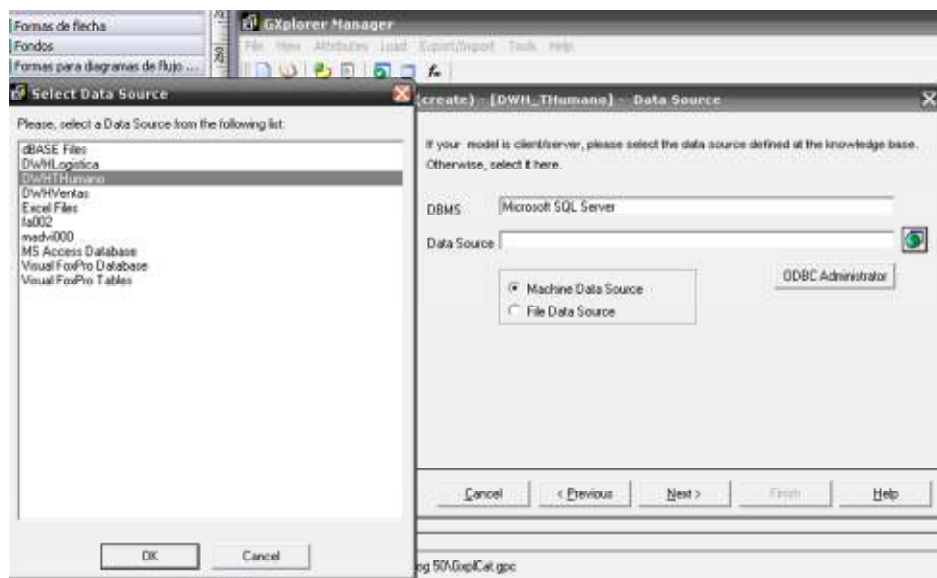


Figura No. 4.65: "Creación y asignación del ODBC para el Metadata"

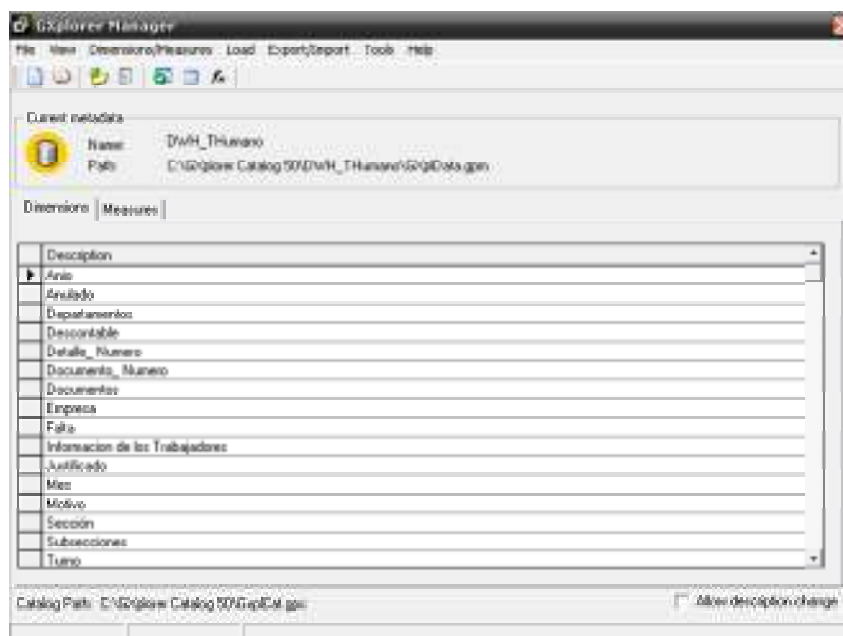


Figura No. 4.66: "Visualización de la Metadata con sus dimensiones y medidas cargadas"

- Una vez listo nuestro metadata es momento de crear los respectivos usuarios y dar permisos a cada uno de ellos (si el DWH lo necesita), para efectos de demostración se creó un usuario adicional para el uso del DWH. Primero ingresamos al modulo de Gxplorer Settings e ingresamos con la clave que viene por defecto e indicando la metadata que

queremos utilizar (figura 4.67) y luego creamos un nuevo usuario (figura 4.68). A más de esto se pueden crear grupos, nuevos catálogos, dar opciones y permisos a cada usuario según la metadata que se elija, etc.

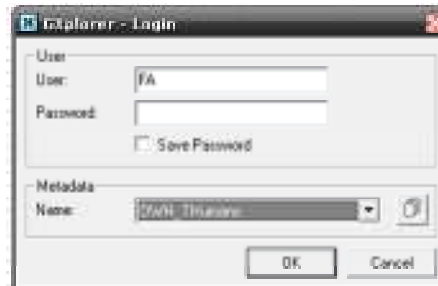


Figura No. 4.67: "Ingreso al modulo GXplorer Settings"

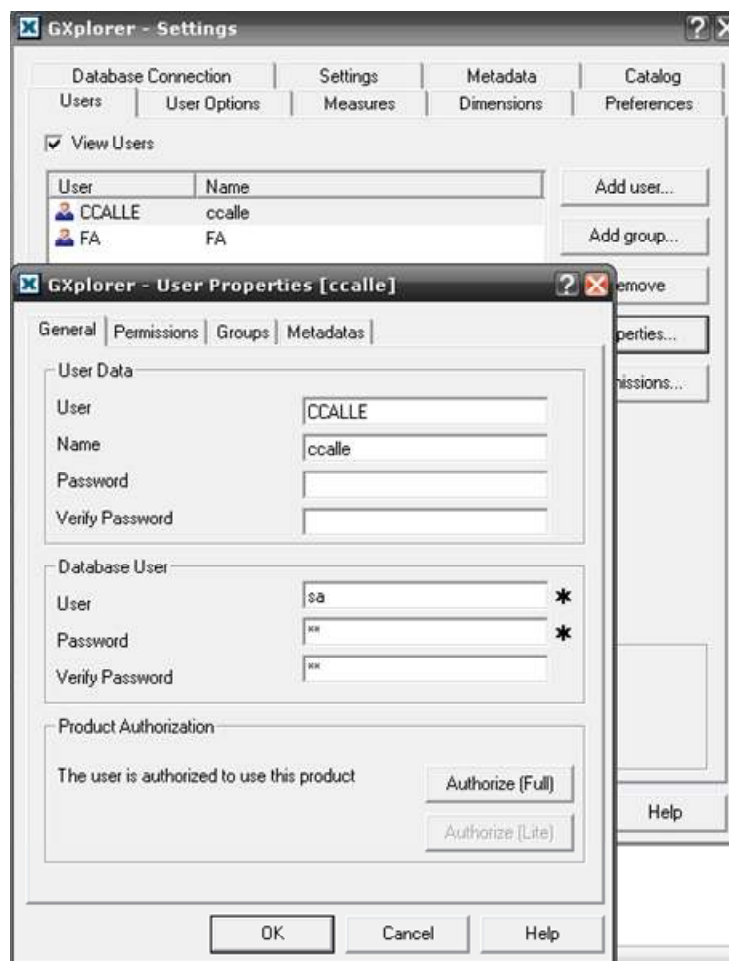


Figura No. 4.68: "Creación de un nuevo usuario"

- Por ultimo tenemos el modulo Gxplorer Olap para M\$Excel que lo encontraremos como complemento de Microsoft Excel y nos permitirá explorar la información de nuestro DWH (figura 4.69).

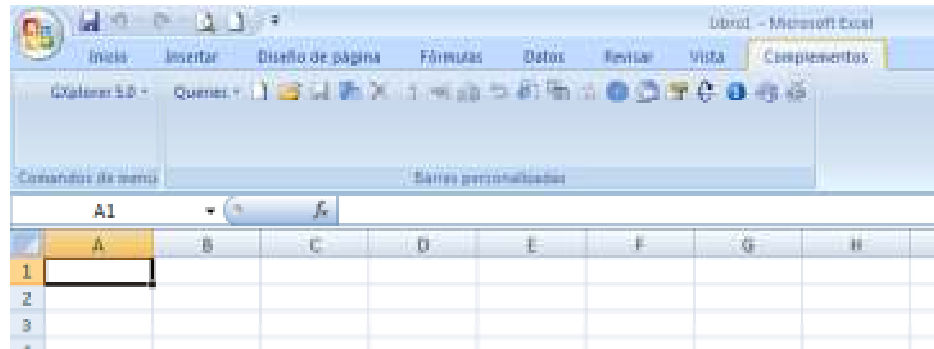


Figura No. 4.69: "Ubicación de la barra de Gxplorer Olap para M\$Excel"

- Una vez ubicada la barra del Gxplorer, procedemos a crear una nueva consulta (figura 4.70), indicando el usuario y la metadata a utilizar.

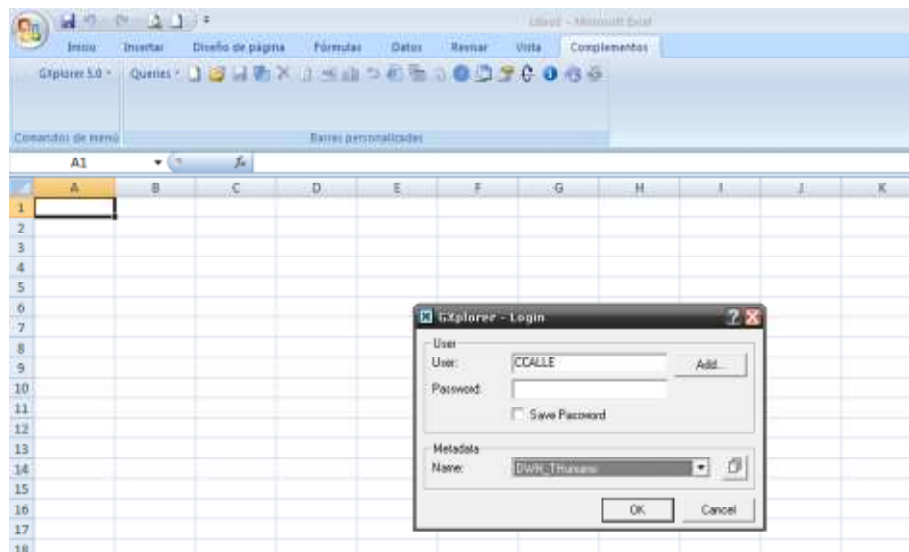


Figura No. 4.70: "Nueva Consulta Gxplorer"

- A continuación nos aparece una pantalla la cual nos muestra todas las dimensiones y medidas de nuestro metadata, aquí elegimos que dimensiones queremos visualizar y cuál va a ser la medida resultante (figura 4.71). De igual manera se pueden configurar varias opciones ya sea a las dimensiones, medidas, o la consulta completa.

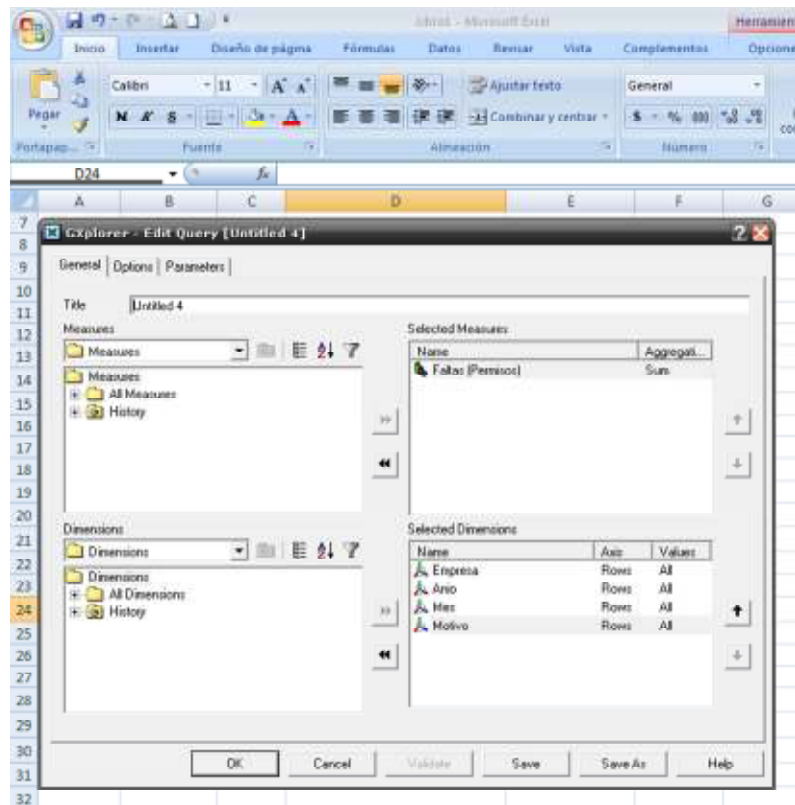


Figura No. 4.71: "Selección de las dimensiones y medidas a consultar"

- Una vez creado nuestro reporte los resultados se muestran de la siguiente manera (figura 4.72):

Libro1 - Microsoft Excel					
Inicio Insertar Diseño de página Fórmulas Datos Referencias Vista Complementos					
Calibre 11 A A [iconos] Ajustar texto General					
N B I [iconos] Combinar y centrar \$ % &					
Portapapeles Fuente Alineación Idioma					
Escribe aquí lo que buscas [icono]					
A	B	C	D	E	F
7					
8	Empresa	Año	Mes	Motivo	Faltas (Permisos)
9	FIBRO ACERO S.A.	AÑO 2008	ENERO	ASUNTO DE EMPRESA	0
10				ASUNTO PATICULAR	0
11				ATRAZO	2
12				CALAMIDAD DOMESTICA	0
13				CERTIFICADO MEDICO FABRICA	0
14				FALTA	5
15				VACACION	0
16			Total ENERO		7
17			FEBRERO		15
18			MARZO		8
19			Total AÑO 2008		30
20	Total FIBRO ACERO S.A.				30
21	FUNDICIONES Y T.T.S.A.	AÑO 2008	ENERO	ASUNTO PATICULAR	0
22				CALAMIDAD DOMESTICA	0
23				CERTIFICADO MEDICO FABRICA	0
24			Total ENERO		0
25			FEBRERO		0
26			MARZO		0
27			Total AÑO 2008		0
28	Total FUNDICIONES Y T.T.S.A.				0
29	# SOTEM C				30
30	# TEMPOR				70
31	Total general				130

Figura No. 4.72: “Resultado de la consulta con Gxplorer Olap para MExcel”

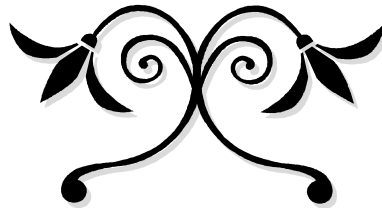
- A mas de los resultados mostrados anteriormente, este complemento también nos permite graficar los resultados (figura 4.73)

4.3. Conclusiones

Al ser de vital importancia contar con información precisa, confiable, segura y que pueda ser consultada en cualquier momento, es importante desarrollar el DWH de la mejor manera posible. En este capítulo se ha expuesto una metodología efectiva y fácil de seguir, que en nuestro caso nos permitió cumplir el objetivo de desarrollar el DWH para la empresa Fibro Acero S.A. utilizando la herramienta Genexus y base de datos SQL Server 2000.

Basándonos en un esquema de diseños pudimos conocer a fondo todos los detalles de cada requerimiento que los usuarios necesitan y cumplirlos a cabalidad y con calidad.

Además, al tener una comunicación continúa con el personal de sistemas de la empresa se pudo interpretar y desarrollar de mejor manera cada requerimiento, resultando un desarrollo completo que engloba las necesidades de los usuarios que les permite de una manera sencilla utilizar la información almacenada en el DWH a través de la herramienta Gxplorer como componente de Excel.



CAPITULO 5

PRUEBAS SOBRE UN DATA WAREHOUSE

5. PRUEBAS SOBRE UN DATA WAREHOUSE

5.1. Introducción

Dentro de la realización y desarrollo de un software, una de las partes más importantes es la planificación de una estrategia para la detección y corrección de errores, ya que, al tener un mal funcionamiento de un programa los datos, resultados y análisis que nos entregan pueden llevarnos a tomar decisiones incorrectas y por ende, desconfiar del software y en el peor de los casos prescindir de dicha aplicación.

Para que esto no ocurra, se han creado diferentes tipos de pruebas según el escenario en el que se trabaje y los diferentes tipos de errores que se necesiten corregir. Estas pruebas se diferencian según el tiempo en las que se apliquen y se dividen en 2 grupos: Pruebas durante la programación y pruebas después de la programación, cada uno de estos grupos tienen subgrupos los cuales serán descritos en el desarrollo de este capítulo y también se analizarán las ventajas y desventajas que presentan cada uno de estos.

5.2. Tipos de pruebas y características

Existen dos tipos de pruebas que pueden ser realizadas para la comprobar el correcto funcionamiento de un programa, estas son: Pruebas durante la Programación y Pruebas después de la programación ²⁰, y cada una de estas tienen diferentes subtipos que serán indicados y explicados a continuación:

- Pruebas durante la programación
 - Prueba de Integración.
- Pruebas después de la programación
 - Prueba de Caja Negra
 - Prueba de Caja Blanca
 - Prueba de Aceptación
 - Prueba de Rendimiento
 - Prueba de Robustez
 - Prueba de Resistencia
 - Prueba de Transformación

5.2.1. Prueba de Integración

Tiene como objetivo el verificar el correcto funcionamiento entre dos módulos que interactúan entre sí. En el caso que se necesitara de un tercero para comprobar su funcionamiento, es necesario crear módulos simuladores y ver si los datos entregados son los esperados por cada uno de los módulos. Se recomienda que este tipo de pruebas se realicen de manera ascendente

²⁰ <http://www.masterdisseny.com/master-net/librecom/index.php3>. [consulta 24 de febrero de 2008]

(Down to Top), ya que las pruebas se realizarán con los módulos existentes y la generación de módulos simuladores será casi nula.

5.2.2. Prueba de Caja Negra

Este tipo de prueba nos muestra si el programa hace lo que tiene que hacer. Esta se enfoca en la parte externa del modulo o programa sin importar su codificación y lo que trata de encontrar es fallas en su interfaz²¹, apariencia del menú, codificación de teclas, etc. De igual manera se debe realizar la prueba de sus entradas aplicando 3 clases de rangos: por debajo, dentro y por encima del rango de los límites; ya que estos rangos nos permiten verificar que el programa puede reaccionar ante cualquier escenario de la manera más adecuada.

5.2.3. Prueba de Caja Blanca

Se basa en el diseño de un plan de pruebas en el código fuente, el mismo que se va ejecutando sistemáticamente hasta que haya corrido todo o la mayoría del código, en otras palabras, lo que trata de hacer es verificar que el código fuente se ejecute como estaba planeado. Esta prueba no siempre llega a ser completada en un 100%, ya que al tener bucles sus probabilidades se vuelven infinitas. Para no hacer de los bucles un problema mayor, se puede llevar un control indicando un valor de cobertura, el cual nos permitirá conocer si el porcentaje de código revisado es suficiente para que el programa sea considerado como confiable.

²¹ Molina, Marco. "Metodología y tecnología de la programación". Marzo, 1997. [consulta 5 de marzo del 2008]

Si se desea realizar la comprobación de los bucles, es necesario realizar tres tipos de pruebas: cero ejecuciones, una ejecución y varias ejecuciones. Cabe indicar que estas pruebas se realizan cuando el software esta completo y esta prueba no remplaza a la prueba de Caja Negra, sino que la refuerza.

5.2.4. Prueba de Aceptación

Estas pruebas son las realizadas al cliente. Aquí se ve si el software cumple con el objetivo y los requisitos preestablecidos por la empresa y los usuarios. Esta prueba puede ser realizada al finalizar cada módulo o antes de que se entregue como la versión final.

5.2.5. Prueba de Rendimiento

Son aquellas que determinan los tiempos de respuesta, el espacio que ocupa el módulo en disco o en memoria, el flujo de datos que genera a través de un canal de comunicaciones, etc.

5.2.6. Prueba de Robustez

La prueba de robustez se encarga de verificar la capacidad que tiene el programa de soportar entradas incorrectas, por ejemplo, ingresar datos numéricos en campos alfanuméricos, o ingresar en una factura la cantidad de un producto en el lugar donde se ingresa el código de un producto, etc.

5.2.7. Prueba de Resistencia

Las denominadas pruebas de resistencia se utilizan para saber hasta donde puede soportar el programa en condiciones extremas. Por ejemplo, los tiempos de respuesta con el procesador a un 95% de su utilidad o con muy poco espacio en disco.

5.2.8. Prueba de Transformación

Esta se la realiza cuando el software esta terminado, y tiene como objetivo el permitir conocer a fondo el programa realizado. El equipo de trabajo se divide en dos grupos e introducen errores para que el otro equipo encuentre los errores utilizando los métodos que se han utilizado para la corrección de errores del software. Esta prueba tiene un alto costo de tiempo y dinero.

5.3. Ventajas y desventajas

Como todo tipo de elementos que intervienen en un proyecto tienen sus ventajas y desventajas, los tipos de pruebas no están libres de ellos. A continuación se indicarán las ventajas y desventajas más relevantes:

Pruebas de integración.

- Ventajas
 - Facilita la detección y corrección de errores, ya que se prueba lo creado a medida que se avanza el proyecto.
 - Se ahorra tiempo y dinero cuando se utiliza este tipo de método de prueba.
- Desventajas

- Si no se planea adecuadamente la manera de creación de los módulos puede ser contraproducente, ya que nos puede llevar a la generación de varios módulos simuladores y la pérdida de tiempo y dinero se incrementaría.

Pruebas de Caja Negra

- Ventajas
 - Verifica la ejecución correcta de la interfaz del módulo finalizado y esto ahorra tiempo, dinero y esfuerzo al final del proyecto.
- Desventajas
 - No debe ser hecha por el usuario final, ya que si el módulo tiene varios errores este puede llevarse una desilusión, generar un rechazo y desconfianza del software al final del proyecto.

Prueba de Caja Blanca

- Ventajas
 - Nos asegura el correcto funcionamiento del software.
 - Permite ver errores de codificación.
- Desventajas
 - No siempre es posible completar el 100% de esta prueba.
 - Solamente se lo realiza al finalizar el software.
 - Si el diseño de pruebas no es generado de manera correcta se perdería tiempo y dinero.

Prueba de Aceptación

- Ventajas
 - El cliente se siente parte del desarrollo del sistema y permite un avance más rápido y efectivo.
 - Se conoce si todos los requisitos han sido satisfechos.

Prueba de Rendimiento

- Ventajas
 - Permite conocer ciertos módulos que no funcionan correctamente y también mejorar los procesos para futuros programas.
 - Al realizar estas pruebas el desarrollador puede incrementar sus conocimientos y corregir los diferentes errores que se hayan cometido en el software y aplicarlos en futuros proyectos.
- Desventajas
 - Este tipo de pruebas no son muy utilizados por los desarrolladores de software. Los diferentes casos que se pueden observar con estas pruebas se las va desarrollando una vez que el software se encuentra en producción y a medida que el mismo lo necesite.

Prueba de Robustez

- Ventajas
 - Al ingresar datos errados podemos ver que el programa funcione correctamente en distintos escenarios.

Prueba de Resistencia

- Ventajas
 - Permite conocer si nuestro software es estable según diferentes condiciones que se presenten.

Prueba de Transformación

- Ventajas
 - Permite tener un conocimiento amplio del software.
 - Al utilizar las herramientas para detectar errores nos familiarizamos más con estas y es posible detectar errores de una manera más rápida y eficiente.
- Desventajas
 - Se tiene costos muy grandes de tiempo, dinero y personal.

5.4. Conclusión

Como se indicó al inicio de este capítulo, la detección y corrección de errores es una de las partes más importantes dentro del desarrollo de aplicaciones ya que nos permite conocer los defectos de nuestro sistema y arreglarlos. A lo largo de la historia se han creado diferentes maneras de corregir a los errores dependiendo del escenario en el que se encuentren: Verificar el código fuente, verificar la interfaz con el usuario e inclusive se ha pensado ciertas características que eran dejadas en segundo plano como la robustez, el rendimiento, la resistencia, etc.

Con estas diferentes técnicas o métodos, se puede llegar a realizar un software o proyecto con un porcentaje mínimo de errores, permitiendo que los resultados entregados por los programas sean más precisos y confiables, dando pautas a que los empleados puedan tomar mejores decisiones y de esa manera incrementar la productividad de una empresa.

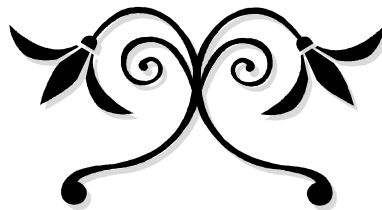
Una vez concluido el desarrollo del DWH y en caso de ser implementado, sería necesario aplicar ciertas pruebas descritas en este capítulo, según nuestro criterio las más adecuadas podrían ser:

- Prueba de Integración: Cuando se cargan los datos desde una base de datos externa estas se registran tanto en dimensiones como medidas, una de las maneras de verificar que el software estaba bien integrado

es ver si la información de las dimensiones coincidían con la que encontrábamos en la medida.

- Prueba de Caja Negra: El momento de realizar un reporte de análisis de los datos se pudo observar que los datos resultantes eran eficientes y confiables, esto nos indicó que todas las operaciones que se realizan internamente están correctos.
- Pruebas de Aceptación: Con esta prueba verificamos si la interfaz y los requisitos realizados por el usuario son los correctos y aceptados por el mismo, esto es de vital importancia ya que la persona que utiliza el sistema es el que hará perdurar el sistema o en algunos casos será el primer opositor del mismo.

Utilizando estos tres tipos de pruebas en la implementación del DWH podremos estar seguros que nuestra aplicación tendrá éxito y será de calidad.



CAPITULO 6

CONCLUSIONES

6. CONCLUSIONES

Al término del desarrollo de esta tesis hemos podido analizar las ventajas de implementar un DWH, como son:

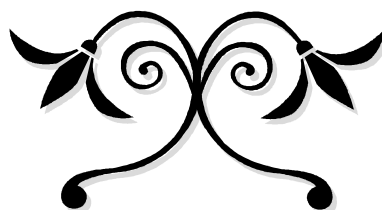
- Versatilidad para obtener informes por parte de los gerentes departamentales de una manera rápida y confiable.
- Generación de informes multidimensionales de una manera simple y amigable, algo que no podría obtenerse con los sistemas operacionales.
- La información obtenida da a conocer a fondo lo que sucede en la empresa, permitiendo a los gerentes mejorar sus estrategias.
- Nos permite ser más eficaces en nuestras labores diarias, ya que ahorramos recursos en tiempo y esfuerzo, haciendo que esos recursos se enfoquen en la productividad de la empresa.

Es por esto que en Fibro Acero S.A. se decidió crear un DWH en el departamento de Ventas, Talento Humano y Logística, no se lo va a realizar en el departamento de Producción ya que en este momento se está desarrollando su sistema. Este desarrollo se lo hará con la herramienta Genexus, utilizado por las facilidades que brinda para la creación, administración y explotación del DWH a través de Gxplorer, que al ser un complemento de Microsoft Excel da la facilidad a los gerentes de trabajar en un ambiente muy conocido para ellos.

Los resultados de este proyecto tanto para los desarrolladores como para la empresa fueron muy satisfactorios, debido a que a través de todos los análisis realizados en cada una de las fases que se han presentado, se logró obtener un diseño muy ajustado a las necesidades y requerimientos de cada departamento como son:

- Ventas: obtener información de las ventas totales de la empresa en general, como de cada una de las sucursales.
- Talento Humano: poder tener información que les ayude a controlar las faltas y permisos del personal, así como obtener informes sobre las remuneraciones de sus empleados.
- Logística: manejar información sobre los proveedores de la empresa para escogerlos de una manera eficiente.

Y para nosotros como desarrolladores el mostrar no solo el desarrollo de un DWH para la empresa Fibro Acero S.A. sino también los conceptos y fundamentos necesarios para servir de guía en la creación de un DWH sin importar el ámbito en el que se desenvuelva la empresa.



CAPITULO 7

BIBLIOGRAFIA

7. BIBLIOGRAFIA

INMON, W.H. "Building the Data Warehouse". Tercera Edición, Jhon Wiley & Sons, 2002.

SEPÚLVEDA CASTRO, Jorge. URRUTIA SUÁREZ, Luis. "Diseño e Implementación de un Data Warehouse para la gestión de ventas de la empresa vitivinícola Miguel Torres Chile". Diciembre, 2004.

MOLINA, Marco. "Metodología y tecnología de la programación". Universidad Politécnica de Valencia 1997.

CARPANI, Fernando. "CMDM: Un modelo Conceptual para la Especificación de Bases Multidimensionales". Agosto, 2000.

MASTER-NET. Probando Software y números de versión.

<http://www.masterdisseny.com/master-net/librecom/index.php3>

WIKIPEDIA. Pruebas sobre un software. http://es.wikipedia.org/wiki/Beta_testing

INSTITUTO NACIONAL DE ESTADÍSTICA E INFORMÁTICA. Manual para la construcción de un Data Warehouse.

<http://www.inei.gob.pe/web/metodologias/attach/lib619/INDEX.HTM>. 1997.

ASESORÍA, DISEÑO Y DESARROLLO DE APLICACIONES.

<http://www.navactiva.com/web/es/atic/aseso/desarrollo/asesor1/2005/32086.jsp>

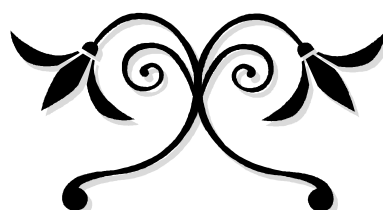
WIKIPEDIA. Modelo Relacional. http://es.wikipedia.org/wiki/Modelo_relacional.

PERALTA, Verónica. "Diseño Lógico de Data Warehouse a partir de esquemas conceptuales multidimensionales". Noviembre, 2001.

<http://www.fing.edu.uy/vperalta.pdf>.

GENEXUS. <http://www.technologies.cl/hxwec20.exe?1,1,64>.

ASOCIACIÓN COLOMBIANA DE INGENIEROS DE SISTEMAS.
<http://www.acis.org.co/index.php?id=623>. Noviembre 2005.



ANEXOS

ANEXOS

Anexo 4.1: Extracto de los archivos en Excel que contienen la información a ser migrada.

ID	PRODUCTO	CANTIDAD	PRECIO	MONEDA	ORIGEN	DESTINO	ESTADO
1	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
2	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
3	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
4	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
5	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
6	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
7	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
8	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
9	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
10	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
11	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
12	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
13	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
14	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
15	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
16	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
17	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
18	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
19	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
20	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
21	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
22	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
23	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
24	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
25	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
26	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
27	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
28	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
29	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
30	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
31	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
32	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
33	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
34	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
35	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
36	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
37	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
38	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
39	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
40	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
41	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
42	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
43	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
44	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
45	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
46	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
47	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
48	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
49	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
50	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
51	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
52	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
53	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
54	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
55	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
56	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
57	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
58	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
59	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
60	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
61	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
62	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
63	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
64	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
65	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
66	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
67	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
68	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
69	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
70	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
71	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
72	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
73	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
74	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
75	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
76	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
77	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
78	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
79	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
80	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
81	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
82	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
83	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
84	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
85	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
86	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
87	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
88	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
89	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
90	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
91	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
92	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
93	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
94	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
95	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
96	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
97	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100
98	TUBO DE PLASTICO	100	12.50	USD	MANUFACTURA	TUBOS	100
99	TUBO DE ALUMINIO	100	12.50	USD	MANUFACTURA	TUBOS	100
100	TUBO DE CEMENTO	100	12.50	USD	MANUFACTURA	TUBOS	100

Anexo 4.2: Representación gráfica de la materialización de cubos de la Relación dimensional Ventas.

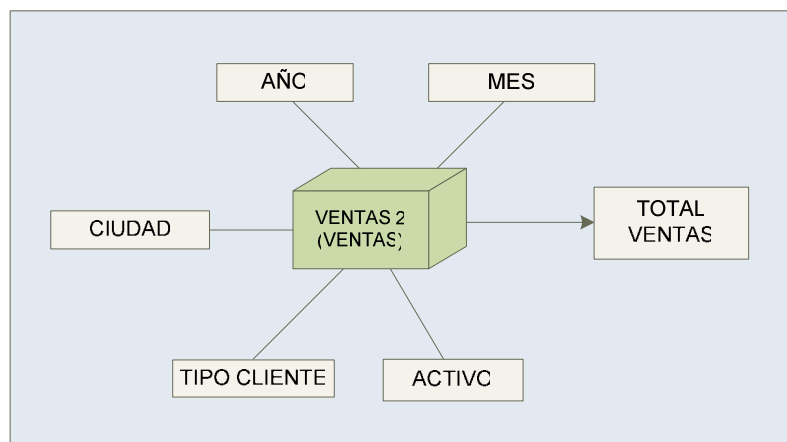


Figura No. 4.74: “Materialización de la Relación Ventas, Cubo 2”

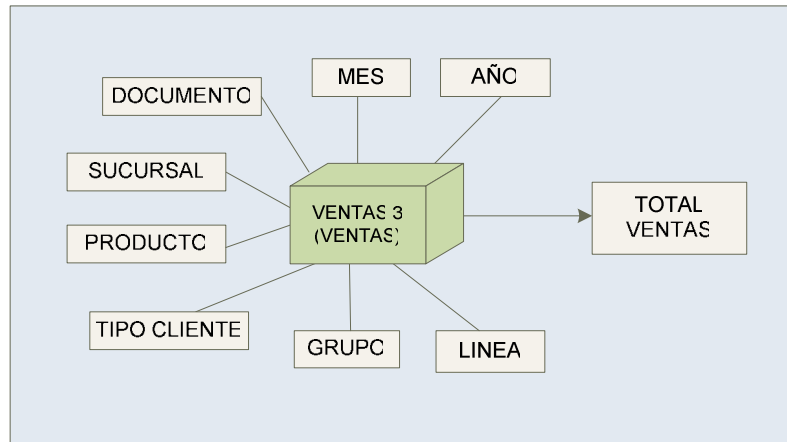


Figura No. 4.75: "Materialización de la Relación Ventas, Cubo 3"

Anexo 4.3: Representación gráfica de la materialización de cubos de la Relación dimensional Logística.

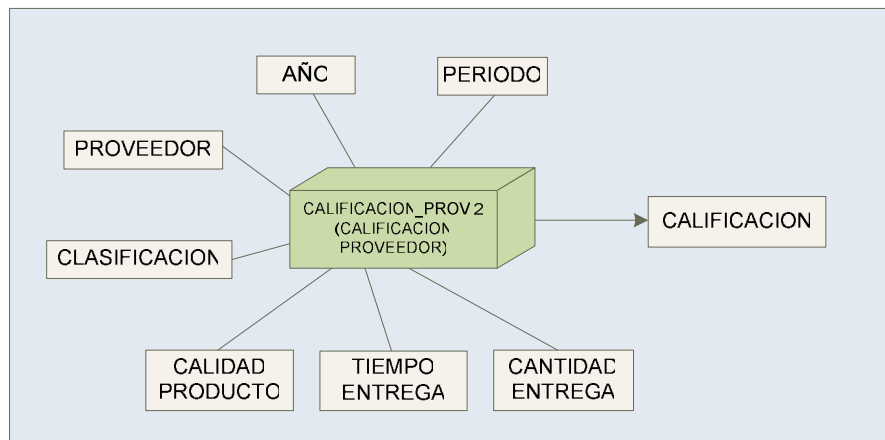


Figura No. 4.76: "Materialización de la Relación Calificación Proveedor, Cubo 2"

Anexo 4.4: Representación gráfica de la materialización de cubos de la Relación dimensional Talento Humano.

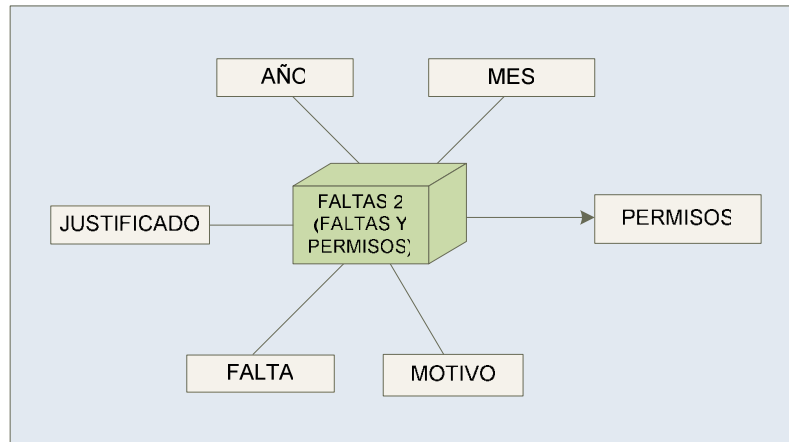


Figura No. 4.77: “Materialización de la Relación Faltas y Permisos, Cubo 2”

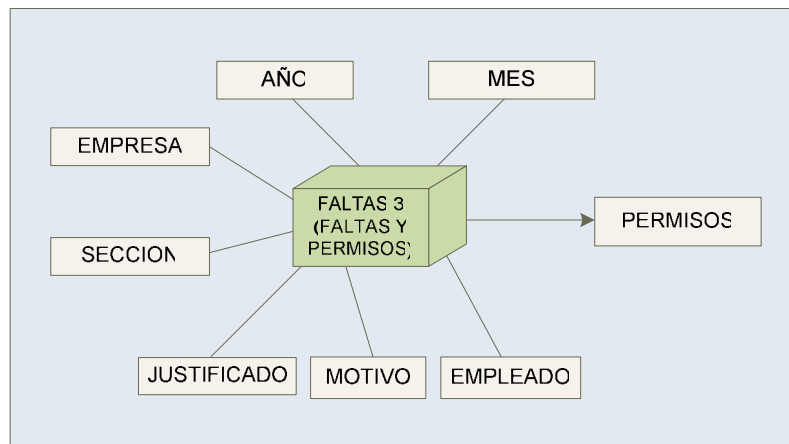


Figura No. 4.78: “Materialización de la Relación Faltas y Permisos, Cubo 3”

Anexo 4.5: Representación gráfica del mapeo de dimensiones del departamento de Talento Humano.

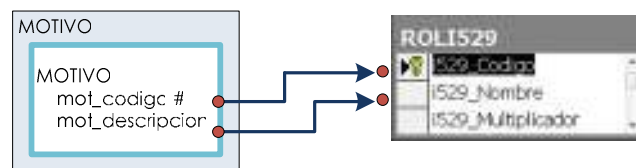


Figura No. 4.79: “Mapeo de la Dimensión Motivo. Dep. Talento Humano”

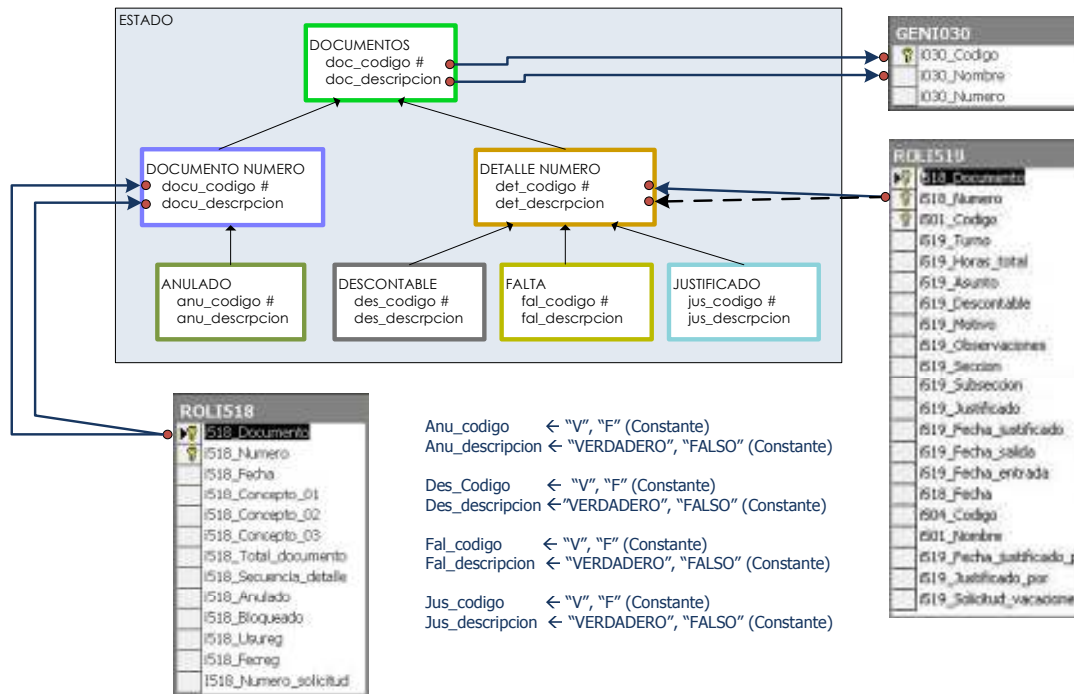


Figura No. 4.80: "Mapeo de la Dimensión Estado. Dep. Talento Humano"

Anic_anic ← Year(ROLI518. i518_Fecha)
Anic_descripcion ← "AÑO" ++ Year(ROLI518. i518_Fecha)

Mes_mes ← Month(ROLI518. i518_Fecha)
Mes_descripcion ← "MES" ++ Month(ROLI518. i518_Fecha)

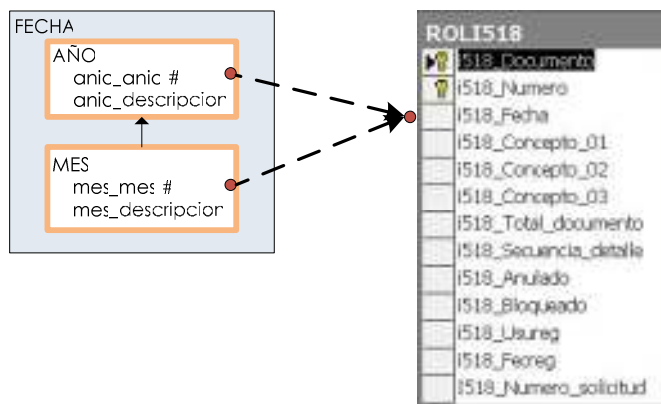


Figura No. 4.81: "Mapeo de la Dimensión Fecha. Dep. Talento Humano"

Anexo 4.6: Representación gráfica del mapeo de dimensiones del departamento de Ventas.

Anic_anio ← Year(DETALLE.DetFec)
 Anic_descripcion ← "AÑO" ++ Year(DETALLE.DetFec)

Mes_mes ← Month(DETALLE.DetFec)
 Mes_descripcion ← "MES" ++ Month(DETALLE.DetFec)

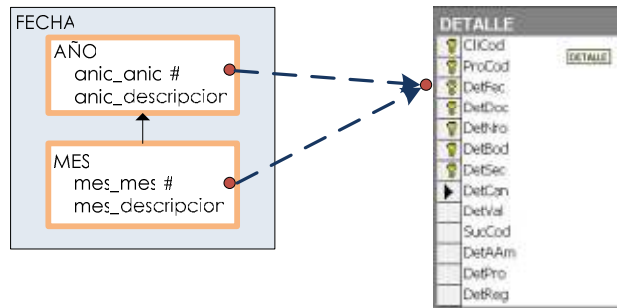


Figura No. 4.82: "Mapeo de la Dimensión Fecha. Dep. Ventas"

Act_codigo = "A ", "I" (Constante)
 Act_descripcion = "ACTIVO", "INACTIVO" (Constante)

Tipcli_codigo = "104" (Constante)
 Tipcli_descripcion = "Consumidor Final" (Constante)

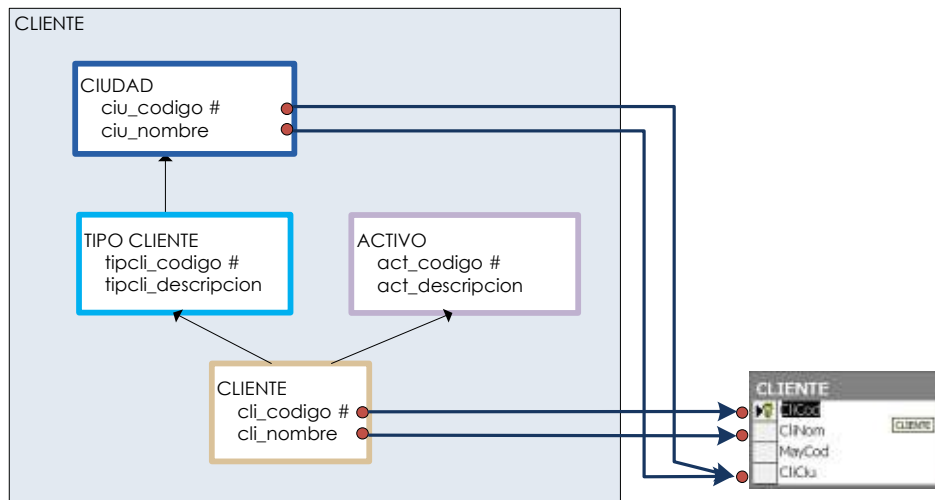


Figura No. 4.83: "Mapeo de la Dimensión Cliente. Dep. Ventas"

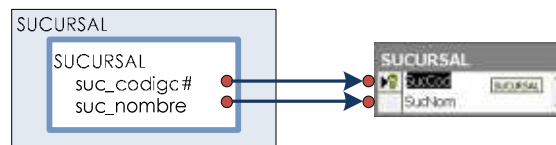


Figura No. 4.84: "Mapeo de la Dimensión Sucursal. Dep. Ventas"

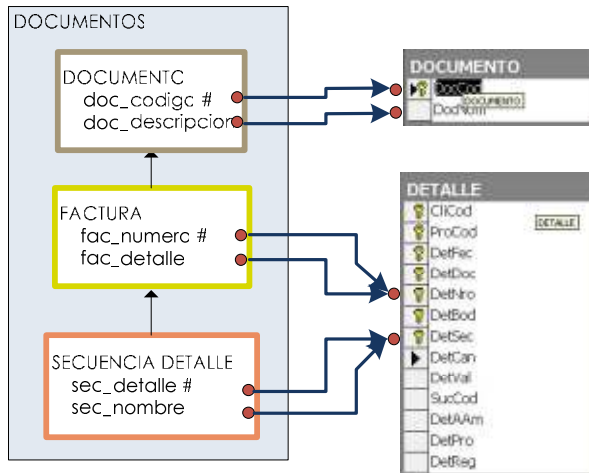


Figura No. 4.85: "Mapeo de la Dimensión Documentos. Dep. Ventas"

Anexo 4.7: Representación gráfica del mapeo de cubos del departamento de Talento Humano.

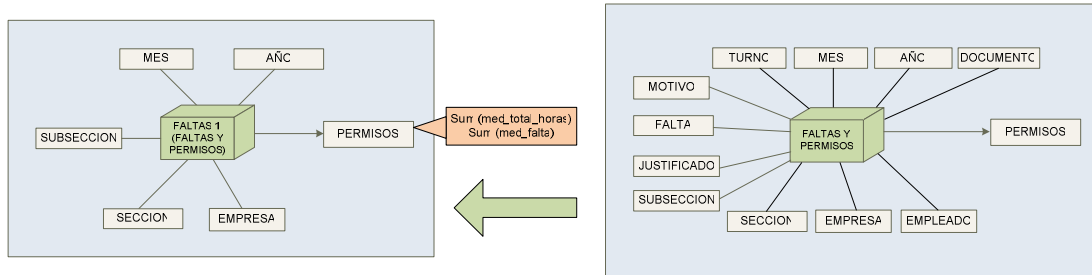


Figura No. 4.86: "Drill-up del Cubo Faltas1. Dep. Talento Humano"

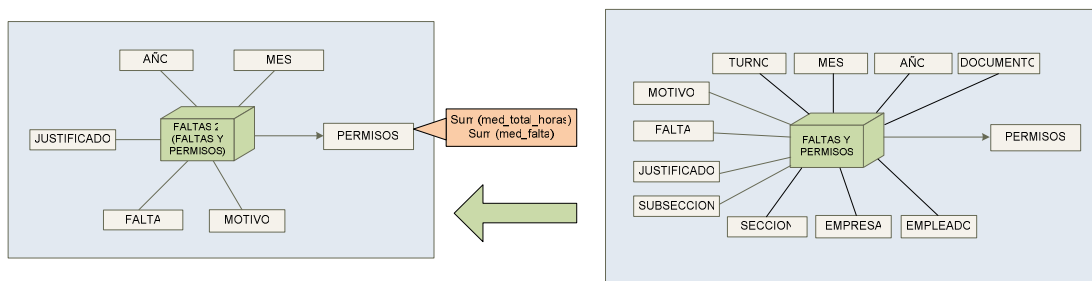


Figura No. 4.87: "Drill-up del Cubo Faltas2. Dep. Talento Humano"

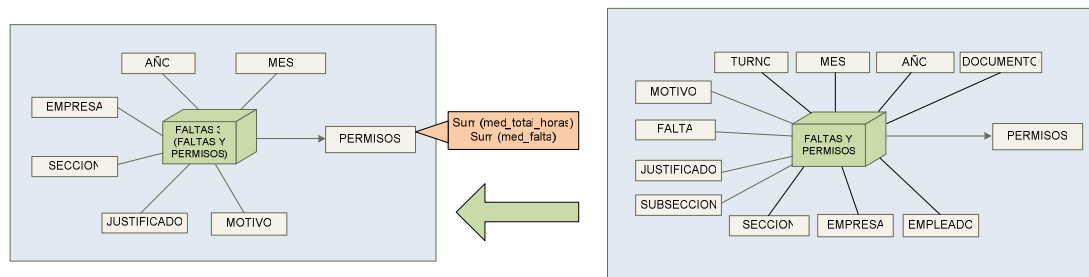


Figura No. 4.88: "Drill-up del Cubo Faltas2. Dep. Talento Humano"

Anexo 4.8: Representación gráfica del mapeo de cubos del departamento de Ventas.

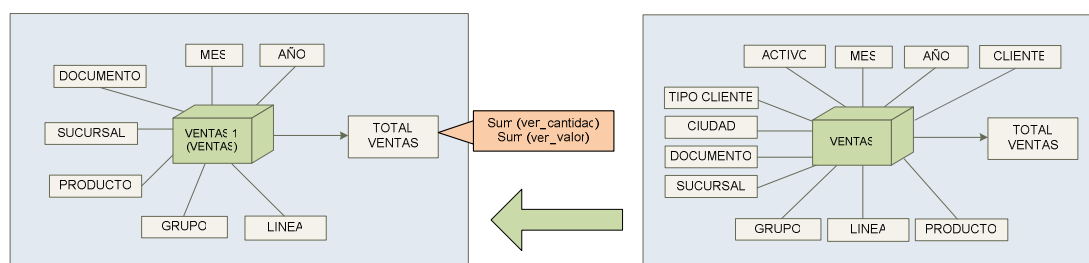


Figura No. 4.89: "Drill-up del Cubo Ventas1. Dep. Ventas"

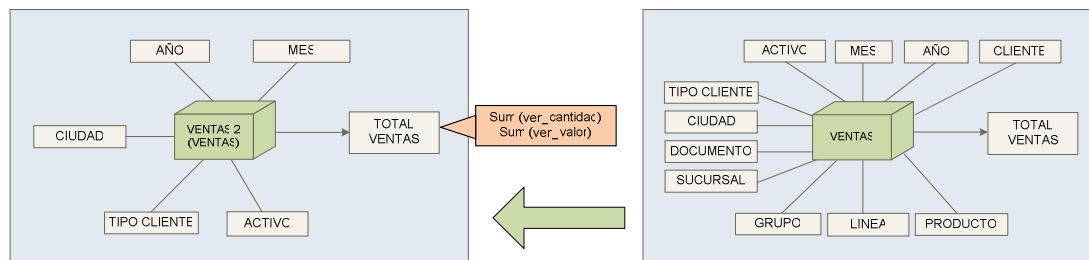


Figura No. 4.90: "Drill-up del Cubo Ventas2. Dep. Ventas"

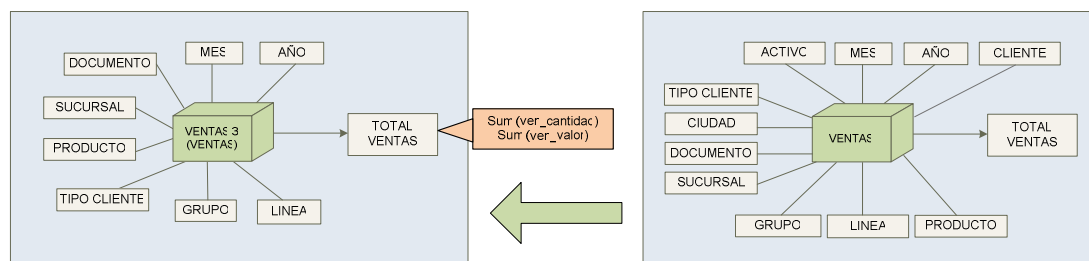


Figura No. 4.91: "Drill-up del Cubo Ventas3. Dep. Ventas"

Anexo 4.9: Representación gráfica del mapeo de cubos del departamento de Logística.

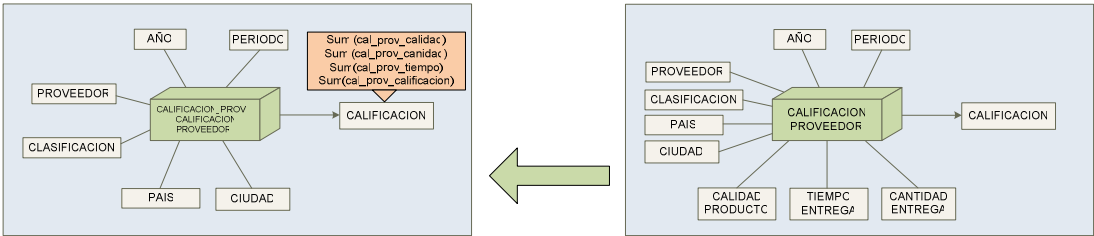


Figura No. 4.92: "Drill-up del Cubo Calificación Proveedor1. Dep. Logística"

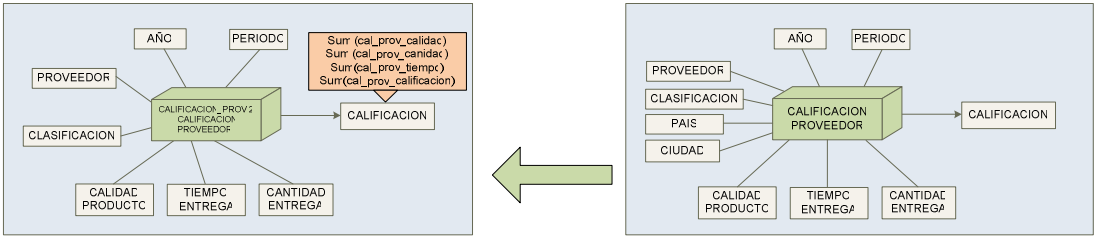


Figura No. 4.93: "Drill-up del Cubo Calificación Proveedor2 . Dep. Logística"