Universidad Central "Martha Abreu Estévez" de Las Villas

Facultad de Matemática, Física y Computación

Prototipo de un Almacén de Datos para Emprestur S.A. SUCURSAL CIENFUEGOS.

Tesis presentada en opción al Título Académico de Master en Computación Aplicada

Autor: Lic. Lino H. Rodríguez Acosta
Tutor: Msc. Rosendo Moreno Rodríguez

2002

RESUMEN

El desarrollo de Sistemas de *Data Warehouse* (DWH) se ha tornado en estos tiempos en una gran área de estudio y aplicación de las empresas. La posibilidad de acceder a informaciones confiables de forma rápida y con garantía en la calidad de los datos, está cautivando a los directores de las organizaciones empresariales, que cada vez más necesitan de un control más correcto de los datos de la empresa sin depender de intermediarios para poder tomar decisiones correctas.

Las informaciones contenidas en los sistemas tradicionales orientados a transacciones no suplen las necesidades de consultas de los directivos, que precisan acceder a largos periodos históricos, muchas veces de varios años, los cuales ciertamente no están disponibles en los sistemas utilizados por las organizaciones, que les permitan dirigir las tareas día a día de la empresa.

Hace ya algunos años William H. Immon y Ralph Kimball, entre otros, han venido trabajando en lo que se ha dado en llamar Almacenes de Datos (Data Warehouse), tecnología que brinda una nueva visión del trabajo con datos.

Dado que en Emprestur Cienfuegos, los datos no están disponibles para la correcta ayuda a la toma de decisiones de la alta gerencia, estamos convencidos que el Diseño e implementación de herramientas para la explotación de un Almacén de Datos, ayudará a la dirección de la entidad a la toma de decisiones basadas en los hechos reales de la empresa.

Para el diseño de un Almacén de Datos, se recomienda la puesta a punto de un prototipo que sea capaz de brindar a la alta gerencia de la empresa, ayuda en la decisión de invertir en la construcción de un DWH.

En la Tesis se diseña un DWH y se implementa un prototipo partiendo de un Data Mart para el área de Comercial de la empresa.

SUMMARY

The development of Systems of Data Warehouse (DWH) transformed in these times in a great study area and application of the companies. The possibility to consent to reliable information in a quick way and with guarantee in the quality of the data, it is capturing the directors of the managerial organizations that more and more they need of a more correct control of the data of the company without depending on middlemen to be able to make correct decisions.

The information contained in the traditional systems guided to transactions don't replace the necessities of the directive consultations that specify to consent to long periods historical, many times of several years, those which certainly are not available in the systems used by the organizations that allow them to direct the tasks day by day of the company.

Already some years ago William H. Immon and Ralph Kimball, among other, they have come working in what has been given in calling Data Warehouses, technology that offers a new vision of the work with data.

Since in Emprestur Cienfuegos, the data are not available for the correct help to the taking of decisions of the high management, we are convinced that the Design and implementation of tools for the exploitation of a Data Warehouse, will help to the address from the entity to the taking of decisions based on the real facts of the company.

For the design of a Data Warehouse, the setting is recommended about to a prototype that is able to toast to the high management of the company, helps in the decision of investing in the construction of a DWH.

In the Thesis a DWH is designed and a prototype is implemented leaving of a Mart Dates for the area of Commercial of the company.

INDICE

Introducción

El desarrollo de Sistemas de *Data Warehouse* (DWH) se ha tornado en estos tiempos en una gran área de estudio y aplicación de las empresas. La posibilidad de acceder a informaciones confiables de forma rápida y con garantía en la calidad de los datos está cautivando a los directores de las organizaciones empresariales, que cada vez más necesitan de un control más correcto de los datos de la empresa sin depender de intermediarios para poder tomar decisiones correctas.

Las informaciones contenidas en los sistemas tradicionales orientados a transacciones no suplen las necesidades de consultas de los directivos, que precisan acceder a largos periodos históricos, muchas veces de varios años, los cuales ciertamente no están disponibles en los sistemas utilizados por las organizaciones, que les permitan dirigir las tareas día a día de la empresa.

Los sistemas de apoyo a la toma de decisiones tienen la finalidad de informar a la alta dirección de la empresa de cómo, por qué y dónde ocurren los problemas o las oportunidades de mejora dentro de la entidad, proyectando ciertas características e indicando al usuario las situaciones que muchas veces pasan desapercibidas.

Es común encontrarnos una diversidad de sistemas de información funcionando en las empresas de gran, medio y pequeño porte, sistemas estos que suplen las necesidades de cada uno de los sectores de la organización, Además de estos sistemas que están en uso, existen algunos otros que ya fueron retirados del servicio y que por un largo periodo almacenaron gran cantidad de datos que no se pueden archivar y olvidar para siempre.

En los últimos años, la tecnología que más se adecuó a la creciente necesidad de las empresas de obtener y trabajar con informaciones gerenciales, que dan soporte a la toma de decisión fue la de los DWH, la cual posibilita diversas características que utilizan de forma adecuada y eficiente las herramientas de desarrollo de modernos bancos de datos.

Mediante el uso de distintos recursos de extracción y manipulación de los datos, la tecnología de DWH permite que una gran cantidad de usuarios pueda realizar inferencias en uno o más bancos de datos modelados de forma especial, que agiliza el acceso a las informaciones y también permite la formulación de consultas definidas en cualquier momento, con un simple movimiento de arrastrar y soltar objetos en interfaces gráficas.

A medida que los datos van siendo captados en los sistemas transacionales, los sistemas de apoyo a la toma de decisión van siendo alimentados de estas informaciones en las maneras más diversas, donde estos datos pasan por procesos de agregación, detalle y totalización entre otros. Esta transferencia puede ser realizada automática o manualmente, en todo momento, durante una noche o hasta mensualmente, hacia el sistema DWH.

Emprestur S.A. al igual que otras empresas, se ha dado a la tarea de introducir las tecnologías de la información a su Sistema de Control, pero adolece aún de la integración de todas sus informaciones que permita la toma de decisiones correcta.

En Cuba debido a la introducción acelerada de equipos de procesamiento de información, en algunas empresas no existe una política central en cuanto a la Informatización de la Empresa, por tanto, vemos empresas que tienen sistemas que no comparten información, en los cuales los datos no están orientados a brindar información actualizada a la dirección de la entidad, sino más bien al control.

Emprestur S.A. Sucursal Cienfuegos no escapa de esta situación, como no escapa también de los siguientes problemas:

- A los directivos se le está haciendo cada día más necesario tener su información recopilada, a mano, siempre disponible y actualizada.
- Cada empresa necesita sus datos de diferentes formas con respecto a la toma de decisiones, cada nivel gerencial tiene distintas maneras de consultar sus datos.
- Aumenta la cantidad de informes impresos para ser analizados por la dirección de la empresa, con menos tiempo disponible para su análisis.

- El trabajo de los comerciales se hace engorroso, dificultado por las actividades de cierres contables e impidiendo a ese mismo personal tener los datos a tiempo para los estudios de mercados correpondientes.

Todos estos y otros aspectos pueden presentarse como elementos determinantes en la forma de ver nuestras bases de datos y no relegarlas a archivos pasivos, que solo serán tocados nuevamente, por una auditoría o para solucionar algún que otro problema puntual, más legal que contable.

Hoy en las empresas más avanzadas, se están instalando redes de Microcomputadoras sobre las cuales se han montado los sistemas de contabilidad, finanzas, y otros. Han soportado bases de datos más que distribuidas, dispersas por todos los discos duros de las máquinas de la empresa, sin poder siquiera hacer un estudio de mercado que se acerque medianamente a los hechos registrados en la base de conocimientos de la empresa.

Con la teoría de las Bases de Datos Relacionales, se pensó que este problema sería resuelto, pero seguimos viendo montañas de papeles llenos de datos, en las mesas de los directivos, sin que ellos puedan accionar de una forma acorde al desarrollo de su empresa. Ha llegado el momento en que la informática y los informáticos jueguen su verdadero papel, es decir, utilizar óptimamente los recursos computacionales y brindar información a la dirección de la empresa, para la toma correcta de las decisiones, pero no escrita solamente, confeccionando informes a solicitud del director o del comercial, sino el hecho de ofrecer una herramienta que pueda ser manipulada y controlada por ellos con un simple entrenamiento.

Emprestur S.A. tiene una infraestructura que permite ir integrando paulatinamente, todos los datos que se originan en sus dependencias, con el fin de brindar información actualizada a la alta gerencia.

Hace ya algunos años William H. Immon y Ralph Kimball, entre otros, han venido trabajando en lo que se ha dado en llamar Almacenes de Datos (Data Warehouse), tecnología que brinda una nueva visión del trabajo con datos.

Esta forma de organizar los datos, partiendo de los sistemas que ya existen –desde ahora sistemas legados– permite un ahorro considerable en tiempo y recursos financieros, ya que solo la inversión es en el diseño y construcción del Almacén de Datos.

El Almacén de Datos (DWH) no es en realidad una herramienta. Más bien, es una nueva filosofía en la forma de cómo ver los datos, con el propósito de aprovecharlos óptimamente. Los autores mencionados y otros, han desarrollados métodos de análisis que pueden ayudarnos mucho en la consecución de este objetivo.

En la bibliografía consultada, no se detalla como hacer un DWH, pero se explican conceptos y problemáticas que deben tenerse muy en cuenta para lograr un trabajo meritorio y no malgastar el tiempo inútilmente; No obstante se dan los principales pasos a seguir para su construcción, de forma general, aunque siempre será necesario aportar gran parte de conocimientos sobre la preparación de los datos para estos fines, en el ámbito particular de la entidad para la que se diseñará.

Mediante las aplicaciones de Data Warehousing se concentra e integra la información importante de las organizaciones a través del tiempo, para lograr hacer las consultas hacia la información más accesibles. El objetivo de los Data Warehouse consiste en convertir la información en Utilidades.

Hasta hoy en Emprestur, se han dado algunos pasos para brindar información actualizada a la dirección de la empresa, como son, instalar en la PC del director el subsistema que elabora los informes, Resúmenes de alguna información en formato Access y Excel o FoxPro, entre otras acciones, pero ninguna de estas es capaz de brindar una información correlacionada de forma tal, que el Consejo de Dirección pueda tomar decisiones correctas, basadas en los hechos reales.

Dado que en Emprestur Cienfuegos, los datos no están disponibles para la correcta ayuda a la toma de decisiones de la alta gerencia, estamos convencidos que el Diseño e implementación de herramientas para la explotación de este Almacén de Datos, ayudará a la dirección de la entidad a la toma de decisiones basadas en los hechos reales de la empresa.

Problema Científico.

Si se hace un análisis de todo el organigrama de la empresa, el flujo de datos desde y hacia cada Direcciones, Sucursales, Filiales, Departamentos, los datos que se generan en cada nivel, se puede apreciar que la alta gerencia no tiene a su disposición los datos necesarios y en el momento oportuno a la hora de tomar una decisión

Objeto de esta Investigación:

Diseñar la Base de Datos que permita integrar los datos de los distintos sistemas heredados. El Campo de Acción es Emprestur S.A. Sucursal Cienfuegos.

Objetivo

El Diseño de un prototipo de Base de Datos de Emprestur S.A. utilizando la metodología de Data Warehouse.

Hipótesis

Se obtendrá el Diseño de la Base de Datos General de Emprestur S.A. mediante la integración de las Nuevas Tecnologías de la Información.

Para el desarrollo de la investigación nos proponemos los siguientes **Métodos**: Análisis (síntesis, inducción, deducción, tránsito de lo abstracto a lo concreto) Análisis crítico del contenido de fuente de información.

Por último como resultado de la investigación se espera obtener el siguiente:

Un Prototipo de Data Warehose para Emprestur S.A.

- Análisis teórico del estado en que se encuentra la problemática planteada en la bibliografía contemporánea.
- Estudio de la experiencia acumulada en el desarrollo, aplicación y validación de sistemas informáticos empresariales.
- Diseño de un prototipo de Data Warehouse que permita la integración y almacenamiento de la información generada en Emprestur S.A. Sucursal Cienfuegos.
- Implementar la explotación del prototipo de Data Warehouse mediante una herramienta profesional.
- Análisis de los resultados alcanzados con la implementación del prototipo de Data Warehouse.

Después de haber realizado las tareas tenidas en cuenta en el diseño, la presente tesis es la expresión de esta investigación. Está conformada por tres capítulos, que se refieren a los siguientes aspectos:

Capítulo 1. Generalidades

Se detalla la situación actual de la empresa para la que se diseña el Almacén de Datos, y se presenta las características de los sistemas de soporte a la toma de decisión, seguida de las variadas posibilidades de arquitectura utilizadas en los proyectos, se podrá apreciar también una comparación entre dos tipos de procesamiento de datos, OLTP y OLAP.

Contiene los tópicos referidos al proyecto y al desarrollo propiamente dicho de sistemas de DWH, identificando inicialmente las funciones de cada uno de los componentes del equipo de proyecto y desarrollo, indicando los pasos que deben seguirse en el proyecto, desde la selección del sector para iniciar el trabajo hasta las reglas que se deben cumplir para que el proyecto tenga un buen final.

También se hace una comparación entre los modelos relacional y dimensional, esta última tradicional en los sistemas de DWH. Algunos de los problemas más comúnmente encontrados durante el desarrollo de sistemas de apoyo a la decisión son citados, también es posible encontrar algunas orientaciones dirigidas a mejorar la ejecución de los sistemas DWH.

Capítulo 2. Diseño del Almacén de Datos

Su propósito principal es explicar de una forma organizada todo el proceso de desarrollo del prototipo que fue implantado en la empresa, utilizando los concepto descritos en el capítulo 1, desde la selección del sector dentro de la empresa, pasando por todas las fases de implantación de un DWH, se analizan las tablas que contienen los datos, pudiendo este análisis servir como base de futuros proyectos dentro de la empresa u otras empresa afines.

Se muestra la metodología que se sigue para el diseño de un Data Warehouse, seguidamente se plantea el diseño del modelo estrella para el Almacén de Datos de Emprestur S.A., además de explicar los módulos que conforman la interfaz operativa de dicho Data Warehouse.

Capítulo 3. Implementación del Almacén de Datos.

Se explica todo lo relacionado con la implantación del prototipo de Data Warehouse, se explica la forma de explotación, las herramientas utilizadas así como el plan de capacitación para la explotación del Almacén de Datos.

Se arriba a conclusiones acerca del diseño del DWH y a la implementación del mismo, se puede apreciar las facilidades que ofrece a la dirección de la empresa el uso de los datos, estructurados según la metodología asumida.

Capítulo 1. Generalidades

1.1.- Descripción del entorno de desarrollo del Almacén de Datos.

Una empresa forma parte del mercado sea regional, nacional o internacional, ella se encuentra siempre compitiendo en una economía globalizada, el aumento del capital intelectual es una obligación para quien desee permanecer competitivo. Empresas que utilizan las informaciones de manera eficiente ganan conocimiento y velocidad de su negocio permitiéndole obtener superioridad en sus mercados.

Es con esta visión que se trata de ofrecer una herramienta, para garantizar que la empresa sea capaz de vencer a sus competidores, para la realización del trabajo, se tiene en cuenta las etapas de planeamiento, proyecto, modelado dimensional e implantación de soluciones de Negocios Inteligentes. La estrategia para construir un sistema de Negocio Inteligente es preparar un Plan de Información. El proceso de planeamiento comienza con la descripción de los objetivos específicos definidos por los directores del negocio. Los factores críticos del suceso de la empresa son consultados con otros profesionales claves y al consejo de dirección, estos pasos garantizan que la gerencia de la empresa esté consciente de las necesidades reales para alcanzar sus objetivos.

La Sucursal de la Empresa del Servicio al Turismo S.A. de la provincia Cienfuegos, es una empresa con carácter territorial, es decir, tiene tres filiales en todo el territorio de la antigua provincia Las Villas, una filial en Cienfuegos, Santa Clara y Trinidad respectivamente. La información de estas filiales, se procesa casi toda, de forma automática con sistemas en diversas plataformas, a continuación detallamos estos.

1.1.1. – Características y explotación de los Sistemas Actuales.

El Sistema de Contabilidad “Account Mate”:
Desarrollado por una entidad extranjera en Clipper para MSDOS, este sistema es el encargado de llevar el mayor de las cuentas de la contabilidad, es solo el núcleo central de

un gran sistema, pero los otros módulos por no conveniencia de la Casa Matriz fueron desechados, dichos módulos fueron sustituidos por algunos productos de factura propia de los especialistas de Emprestur, consta de varias tablas .DBF, pero su tabla principal es el GLMASnn.DBF, donde "nn" significa el número de la entidad. Este Sistema al ser diseñado por una empresa productora de Software, está preparado para el trabajo en Redes, permite la consolidación de múltiples empresas (Filiales o Sucursales según el caso), no posee módulo de reportes en correspondencia a las exigencias del Manual de Procedimientos de Emprestur S.A. Casa Matriz, este sistema tiene una tabla la GLTRSnn.DBF, que es la que guarda todas las transacciones hechas en el mes, pero al efectuar el cierre contable, dichos datos se borran, desde el punto de vista de contabilidad es un sistema muy confiable, se basa en la teoría tradicional de la contabilidad, no se tienen actualizaciones debido a que no es un sistema con licencia.

El Sistema de Submayor:

Desarrollado por los especialistas de la Casa Matriz, en principio, fue diseñado para sustituir los módulos de Cuenta por Cobrar y Cuentas por Pagar del Account Mate, luego se fue convirtiendo en el sistema de control de todas las cuentas, después de crear las obligaciones y efectuar su correspondiente pago o cobro, se actualizan las cuentas del Mayor que existen el Account Mate, está soportado en FoxPro 2.5 para MSDOS, este sistema también se encarga de tomar los datos de los comprobantes de operaciones y registros de contabilidad, tiene en cuenta cuales cuentas son del submayor o no, brinda un conjunto de reportes para el chequeo del estado de las cuentas, no posee una historia de las transacciones, por lo que se hace difícil un análisis del comportamiento de los clientes en el pago, no se puede hacer comparaciones de los clientes en el tiempo y lugar, no se tiene el comportamiento de los proveedores. Este Sistema no está diseñado para el trabajo en redes, se explota en una sola máquina por el personal de contabilidad.

El Sistema REPO:

Se crea por los especialistas de la Casa Matriz, su objetivo fundamental, es obtener reportes de la contabilidad según el Manual de Procedimientos de la Casa Matriz, estos reportes son previamente predeterminados, no existe la posibilidad de crear nuevos, según

necesidades de cada sucursal o filial, este sistema utiliza los datos del Sistema de Contabilidad, su función principal es la presentar los reportes que le interesan a la dirección económica de la Casa Matriz, sus reportes no están vinculados con el Submayor, por tanto, si el director necesita otra información, debe solicitarla al área económica para su impresión. El principal problema de este sistema es la cantidad de información que brinda en algo dado en llamar Estados Financieros, este listado de reportes no satisface las necesidades del Consejo de Dirección, debido al gran volumen de información que contiene, pero de forma diseminada.

El Sistema de Nóminas:

Fue adquirido por encargo a la Empresa de Servicios Informáticos de Cienfuegos, soportado en Clipper para MSDOS. El sistema de nóminas no tiene vínculo alguno con el sistema de contabilidad ni el submayor, solo calcula la nómina de los trabajadores, divididos en sueldistas y jornaleros, emite un comprobante escrito que luego se teclea al sistema de contabilidad, en este comprobante las cuentas que emite no se corresponden con las autorizadas por el área económica, por tanto se debe rehacer dicho comprobante, el sistema no guarda información de los salarios pagados, no permite la consolidación de las distintas filiales de la sucursal, para un análisis posterior del indicador de Salarios total.

El Sistema Automatizado de Inventarios y Contabilidad (SAICON):

Su diseño y programación fue por los especialistas de la Casa Matriz, este subsistema es para el control de los inventarios y genera un comprobante de operaciones, en papel, para la contabilidad, soportado en FoxPro 2.5, este sistema es un gran problema para la empresa, ya que no es confiable, es decir no refleja realmente la situación de los inventarios, frecuentemente se le debe hacer modificaciones a los ficheros de datos, según la existencia real de los almacenes, previo balance con los documentos de entrada y salida, no está diseñado para el trabajo en red, posee un fichero histórico que almacena todas las transacciones efectuadas, pero debido a su inestabilidad no es confiable.

El Sistema de Portadores Energéticos:

Desarrollado por los especialistas de la Casa Matriz, este subsistema, controla los gastos por conceptos de portadores energéticos, se vincula parcialmente con la contabilidad, soportado en FoxPro 2.5, los datos de los portadores energéticos se reflejan en la contabilidad de forma general, es decir todos van a una cuenta determinada, este sistema se encarga de validar el consumo real de cada portador contra la suma que tiene la cuenta en el sistema de contabilidad, emite un balance de la situación real de los portadores energéticos en cada filial y sucursal.

El Sistema de Facturación:

Partiendo de la necesidad de automatizar el proceso de facturación se desarrolló por los especialistas de la Filial de Villa Clara, soportado sobre MSACCESS 97, este sistema su función principal es la confección de las ofertas y facturas, permite que las ofertas se conviertan en facturas, controla las facturas pagadas y las pendientes, brinda la posibilidad de analizar el estado de las facturas, es decir el estado de las cuentas por cobrar por edades, emite un comprobante a contabilidad de las operaciones del mes.

De estos sistemas el único que brinda información medianamente útil a la dirección de la empresa es el Repo, porque es una información de 30 días de atraso.

1.1.2.- Funciones y Características de los ficheros utilizados.

De los sistemas descritos anteriormente, el que más importancia tiene es el Submayor, por la cantidad de información que guarda, la cual presentada correctamente, puede ser útil para la toma de decisiones de la alta gerencia, es en este sistema donde se capta toda la información primaria que luego se procesa y se pasa al Mayor, es decir al Account Mate.

A continuación se ofrece una relación de los principales ficheros del Submayor.

Nombre del Fichero .DBF	Función
DEMASxx	Contiene todos los datos, relativos a los documentos primarios como cobros y pagos, fecha de expiración de las facturas de cobros y pagos, tipo de las cuentas, Deudoras o Acreedoras, importe en ambas monedas, el monto pagado y el saldo de cada cuenta.
DECLIxx	Es un clasificador de clientes y proveedores con los que Emprestur tiene relaciones.
DEUPSxx	Contiene todos los centros de costos de cada filial de Emprestur.
DECTAxx	Clasificador de Cuentas según el Manual de Normas y Procedimientos de Emprestur S.A.

Donde **xx** es el número que identifica a cada filial.

Se expone a continuación la estructura de los ficheros antes mencionados, en el primero, se almacena toda la información primaria de la empresa, es decir, las cuentas por cobrar (estas reflejan las ventas de cada centro de costo) y de las cuentas pendientes de pago (estas reflejan las compras realizadas por cada centro de costo). El resto de los ficheros .dbf son clasificadores del demasxx.dbf

Fichero DEMASxx.DBF.

Campo.	Tipo	Tamaño	Decimales	Descripción
CTA	C	3		Código de la cuenta control
SCTA	C	2		Código de la Subcuenta control
CLIENTE	C	9		Código del Cliente / Proveedor
FACTURA	C	10		Número de la Factura o documento primario
BALMN	N	12	2	Valor correspondiente al Importe inicial del documento primario en Moneda Nacional
BALDIV	N	12	2	Valor correspondiente al Importe inicial del documento primario en USD.
PAGMN	N	12	2	Valor correspondiente al Importe Pagado del documento primario en Moneda Nacional.
PAGDIV	N	12	2	Valor correspondiente al Importe Pagado del documento primario en USD.
SALDOMN	N	12	2	Importe pendiente del campo BALMN
SALDODIV	N	12	2	Importe pendiente del campo BALDIV
UPS	C	4		Código del Centro de Costo emisor o receptor del documento.
SIGLA	C	4		Siglas del reporte que contiene dicho documento
REPORTE	N	4	0	Número del reporte que se emite
DETALLE	C	4		Siglas del documento primario que se emite o recibe.
SUCU	C	2		Código que identifica a la sucursal.
FILIAL	C	2		Código que identifica a cada Filial.
M(*)	C	1		Campo que se utiliza para conocer los documentos que han sido marcados para otro proceso.
M1(*)	C	1		Campo que se utiliza para conocer los documentos que han sido marcados para otro proceso.
ANALV	C	3		Código que relaciona el análisis de venta.
FREPO	D	8		Fecha del reporte que contiene el documento primario.
SALMN(*)	N	12	2	Saldo del cliente o proveedor a la cuenta y subcuenta en Moneda Nacional.
SALDIV(*)	N	12	2	Saldo del cliente o proveedor a la cuenta y subcuenta en USD.

Campo	Tipo	Tamaño	Decimales	Descripción
FECHAVEN	D	8		Fecha de vencimiento del documento primario.
CTAC	C	3		Código de la Cuenta Contrapartida de la Cuenta Control
SUB2	C	4		Código de la UPS de la Cuenta Contrapartida

Fichero DECLIxx.DBF

Campo	Tipo	Tamaño	Decimales	Descripción
NOMBRE	C	77		Nombre de los Clientes
CODIGO	C	9		Código de los Clientes
M1(*)	C	1		
M(*)	C	1		
ANAL	C	3		Análisis de las ventas

Fichero DECTAxx.DBF

Campo	Tipo	Tamaño	Decimales	Descripción
CTA	C	3		Número de la Cuenta
SCTA	C	2		Número que identifica la Cuenta
NOMBRE	C	35		Nombre de la Cuenta
SUB(*)	L			
CON(*)	C	1		
M(*)	C	1		
M1(*)	C	1		
NATURALEZA	C	1		Indica si la cuenta es Deudora o Acreedora
SUBMAYOR	L			Indica si la cuenta es de Submayor o No
ANALISIS	L			Indica si la cuenta tiene análisis
SOBREGIRO	L			Indica si la cuenta puede tener sobregiro
FECHAVEN	L			Indica si la cuenta puede vencer

DEUPSxx.DBF

Campo	Tipo	Tamaño	Decimales	Descripción
CODIGO	C	4		Código del Centro de Costo
NOMBRE	C	35		Nombre del Centro de Costo
M(*)	C	1		
M1(*)	C	1		

(*) campos que no se utilizan en la tabla, es decir no contienen información alguna, esto se debe a problemas del diseño inicial y al mantenimiento posterior del sistema.

Es en estas tablas donde el Sistema guarda la información necesaria para la construcción del Almacén de Datos, la cual solo es procesada para actualizar los saldos de las cuentas del Mayor, pero no ofrecen información útil a la dirección de la empresa en el proceso de toma de decisiones.

Por todo lo anterior visto, es que se diseña una herramienta que sea capaz de brindarle a la alta dirección de la Sucursal Cienfuegos, los datos necesarios para una correcta toma de decisiones, el diseño de un Almacén de Datos.

1.2.- Exposición de la teoría del Almacén de Datos.

De acuerdo con W. H. Inmon, quien es considerado como el padre del Data Warehouse: *"Un Data Warehouse es un conjunto de datos integrados, orientados a una materia que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones de una administración"* [1]. Este concepto se traduce literalmente como Almacén de Datos, no obstante si el Data Warehouse fuese exclusivamente un Almacén de Datos, los problemas seguirían siendo los mismos que con las Bases de Datos Relacionales.

El almacenamiento de datos (Data Warehousing) es la parte de la Tecnología Relacional que consiste en el proceso de obtención de datos desde otros ambientes [2].

Debemos tener en cuenta que en los años 80 del siglo XX se llevó a la práctica la Teoría Relacional de Bases de Datos, todos los sistemas se sustentaron en ella y hoy día muchos aún, son fieles seguidores de ella.

A continuación mostramos un resumen de las diferencias entre un Sistema Tradicional de Bases de Datos y el Data Warehouse.

SISTEMA TRADICIONAL	**DATA WAREHOUSE**
Modelo Simétrico	**Tabla de Hechos Central**
Complejidad en la interrelación de las tablas para una consulta	**Fácil diseño de las consultas**
Predomina la actualización.	**Predomina la Consulta.**
La actividad más importante es de tipo operativo (día a día).	**La actividad más importante es el análisis y la decesión estratégica.**
Predomina el proceso puntual.	**Predomina el proceso masivo.**
Mayor importancia a la estabilidad.	**Mayor importancia al dinamismo.**
Datos en general desagregados.	**Datos en distintos niveles de detalle y agregación.**
Importante del tiempo de respuesta de la transacción instantánea.	**Importancia de la respuesta masiva**
Importancia del dato actual.	**Importancia del dato histórico.**
Estructura relacional.	**Visión multidimensional.**
Usuarios de perfiles medios o bajos.	**Usuarios de perfiles altos.**
Explotación de la información relacionada con la operativa de cada aplicación.	**Explotación de toda la información interna y externa relacionada con el negocio.**
Muy complejos de comprender por parte de los usuarios.	**Fácil de comprender por los usuarios**
Muy complejos de recorrer por los software.	**Complejidad baja, para el recorrido de los software.**

El Modelo Entidad Relación (ER) es una poderosa técnica para el diseño de sistemas de procesamiento de transaciones. Ayudado por la automatización, la normalización de datos basado en el modelo ER ha contribuido al éxito fenomenal de conseguir grandes cantidades de datos en bases de datos relacionadas. Sin embargo, el modelo ER no contribuye a la capacidad de los usuarios para hacer consultas a los datos. Yo recomiendo una técnica diferente (llamada Modelo Dimensional), para diseñar la estructura de los datos orientada a la consultas. Ahora que usted ha triunfado en conseguir los datos, es el tiempo para tenerlos fuera [3].

1.2.1 Arquitectura

Podemos definir dos formas de representar la arquitectura de un DWH, una es conceptual y la otra física del modelo relacional que representa el sistema.

1.2.1.1 Visión Conceptual

Una arquitectura conceptual del DWH es basada en los siguientes componentes [4]:

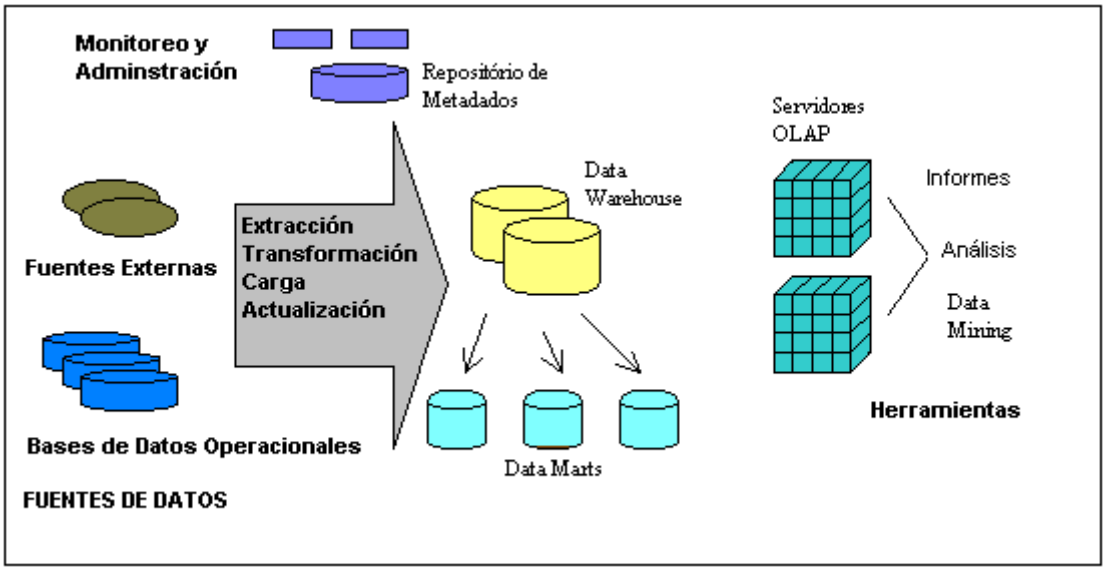

Figura 1.1. Arquitectura de un Sistema DWH

Existen varias herramientas para la extracción de datos de diversas bases operacionales y de fuentes externas, herramientas para la limpieza, transformación e integración de los datos, para la carga de datos en el DWH y otras para la actualización periódica del DWH con el fin de mantener las actualizaciones ocurridas en las fuentes hacia el DWH.

Además del DWH pueden existir varios Data Marts (DMs), que separan los datos por sectores dentro de la organización, dependiendo de los criterios de los departamentos.

Los datos contenidos en el DWH y en los DMs son administrados por uno o más servidores de DWH, los cuales presentan visiones multidimensionales de los datos para una variedad de herramientas *front end.*

La visión multidimensional en forma de cubo de datos indica que las informaciones son visualizadas en líneas y columnas como el formato tradicional de las hojas de cálculo. Esta característica organiza y facilita la consulta de los datos de manera tal, que se puede tener por ejemplo, en una dimensión del cubo los meses del año y en otra dimensión estarían los centros de costos (UPS) y en la tercera dimensión los clientes de estos centros de costos.

Finalmente, existe un repositorio para almacenar y administrar los metadatos, acompañados de herramientas para monitorear y administrar el sistema.

1.2.1.2 Visión en Capas

A continuación explicaremos las capas que componen a esta visión: [5]

- Capas de Bancos de Datos Operacionales y Fuentes Externas: Contienen las bases de datos operacionales y puede contener informaciones de fuentes externas, estos datos reciben un tratamiento especial para poder ser incorporados al DWH
- La Capa de Acceso a los Datos se compone de la mezcla de las herramientas de acceso a la información y los bancos de Datos Operacionales, se comunican con diversos Sistemas de Gestión de Bases de Datos (SGBDs) y sistemas de archivos, siendo este conjunto de características lo que define el nombre de Acceso Universal de Datos

- La Capa de Transporte o Middleware, tiene la función de administrar la transmisión de las informaciones por el ambiente de la RED, que sirve de soporte para el sistema como un todo, separando las aplicaciones operacionales del formato real de los datos, realiza también una colección de mensajes y transacciones y se encarga de entregarlos en los lugares y en el tiempo determinado
- Capa del Data Warehouse, constituye el almacenamiento físico de los datos oriundos de los sistemas operacionales o legados de la empresa y externos, permitiendo un acceso más rápido y seguro a los datos del DWH, además de proveer mayor flexibilidad al tratamiento y facilidades a la manipulación.

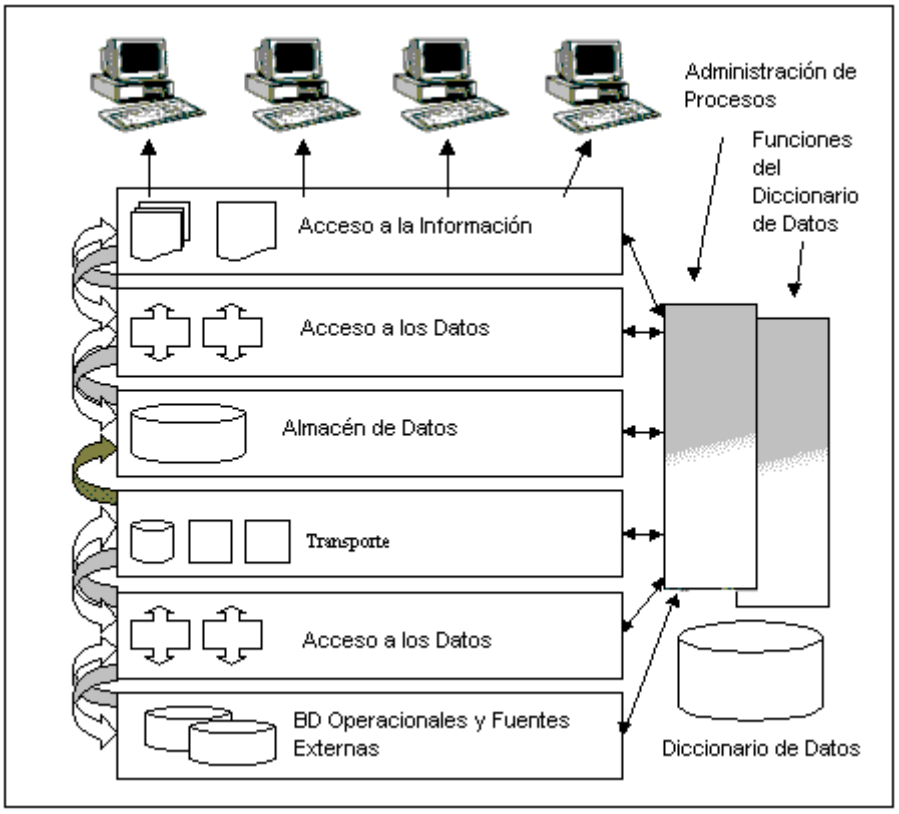

Figura 1.2. Arquitectura en Capas de un DWH

- Capa de Acceso a la Información, proporciona la interacción con los usuarios finales a través de herramientas visuales tradicionales, tales como hojas de cálculo, navegadores, entre otros.

- Capa de los Metadatos (Diccionario de Datos), estos describen los datos y la organización del sistema, pueden ser fórmulas utilizadas para el cálculo, descripciones de las tablas disponibles a los usuarios, descripciones de los campos de las tablas, permisos de accesos, informaciones sobre los administradores del sistema, entre otros.
- Capa de Administración de Procesos, lleva el control de las tareas que mantienen el sistema actualizado y consistente, administrando diversas tareas que son realizadas durante la construcción y mantenimiento del DWH.
- Capa de Administración de Réplica, aunque no se encuentra en la figura, esta capa es una capa superior de la explotación del DWH, su función es seleccionar, editar, resumir, combinar y cargar en el DWH las informaciones a partir de las fuentes externas, en esta capa se debe desarrollar una programación bastante compleja, otra función es la de controlar la calidad de los datos que serán cargados.

1.2.1.3 Estructura Física de los Datos del DWH.

Respecto a la disposición física de los datos, el DWH puede tener una estructura centralizada en único local o ser implementado de forma distribuida. Si optamos por el primer modelo, el centralizado, tenemos un Data Warehouse consolidado y el Banco de Datos (BD) formará un DWH integrado. Definiendo el proyecto de esta forma se puede maximizar el poder de procesamiento y acelerar los procesos de búsqueda de informaciones analíticas.

Si se define una estructura arquitectura federativa, se puede distribuir la información por funciones, separando los datos del sector financiero en un servidor, los datos de marketing en otro local y los datos de producción en un tercer lugar.

Existe una tercera metodología, en la cual se considera una arquitectura de DWH separada por el nivel de granularidad de la información, almacenando los datos más resumidos en un servidor, disponiendo de los datos poco detallados, en un segundo servidor y los datos más detallados (atómicos) en un tercer servidor. La figura 1.3 ejemplifica esta metodología.

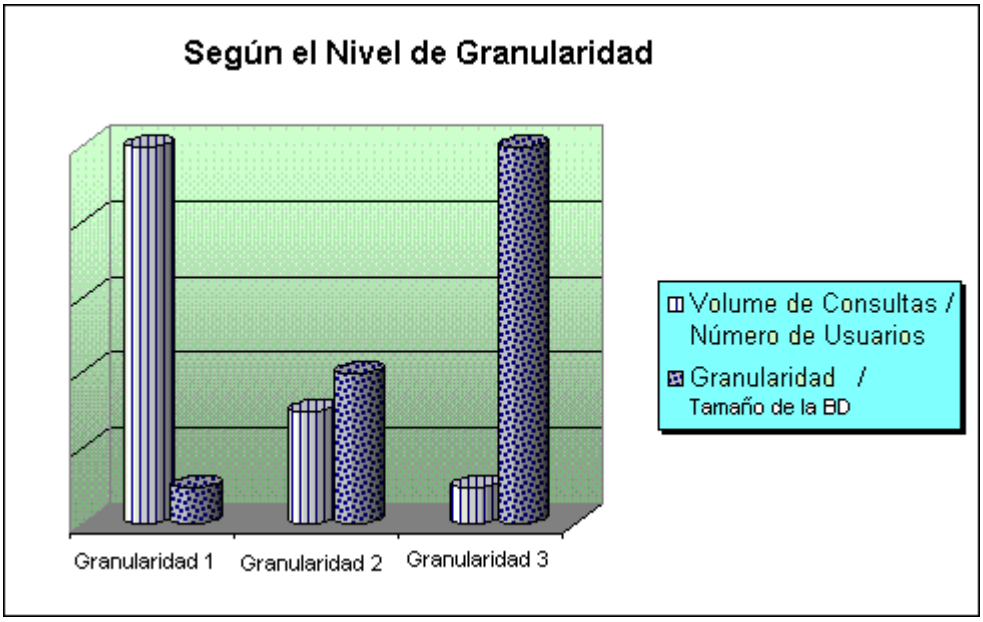

Figura 1.3. Según el grado de Granularidad

El primer servidor generalmente atiende la mayor parte de las consultas, contra una menor cantidad de accesos solicitados para los servidores 2 y 3.

Esta metodología es bastante utilizada, siendo defendida por diversos autores. Es sabido que, además del grado de granularidad del DWH propiamente dicho, se conservan los datos transacionales de donde son colectados los datos atómicos.

1.2.1.4 Arquitectura de Dos Capas

Existe una arquitectura de implantación de sistemas de DWH que consiste en utilizar una computadora de alta capacidad como servidor. Este método dispone de aplicaciones para los usuarios finales en la forma de herramientas *back end*, que sirven para alimentar el DWH con informaciones.

Las organizaciones que pueden crecer con la incorporación de otras empresas del mismo ramo o de otra rama de negocio, gradualmente acumulan diversos sistemas de computación legados, cada uno con sus incompatibilidades de definiciones de los datos. Esta redundancia y falta de consistencia de los datos dificulta la administración de las bases de

datos, convirtiéndose también en una dificultad para desarrollar nuevas aplicaciones *front end.*

Esta arquitectura puede ser llamada como “Paraguas” [6], la cual posee un formato tal que el cabo del paraguas representa el servidor principal y las varillas representan los sistemas de consulta a este servidor.

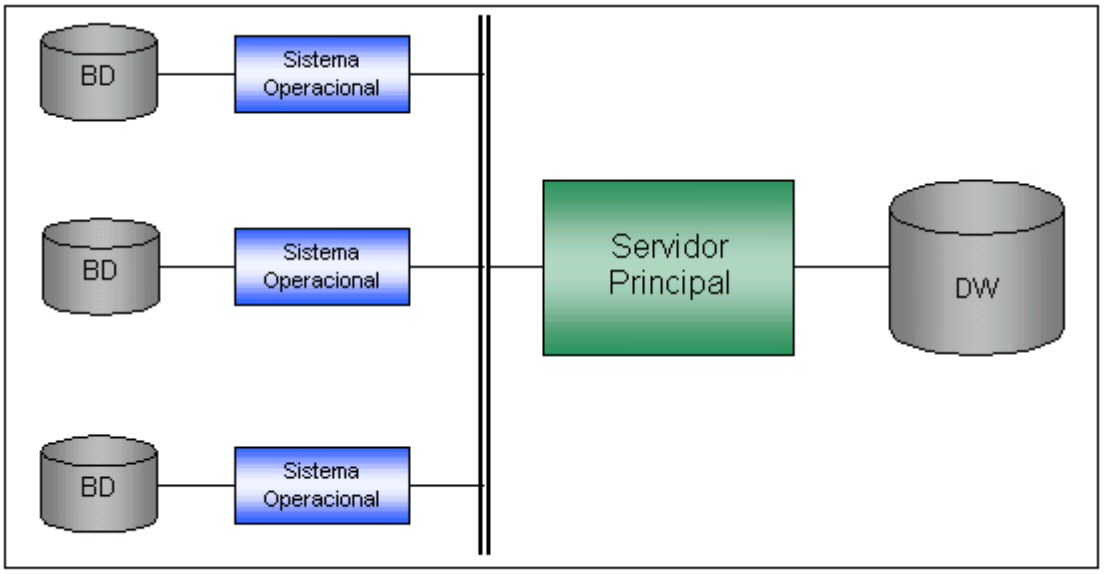

Figura 1.4. Arquitectura “Paraguas”

La arquitectura anterior puede ser usada para construir un sistema de DWH en dos capas, la cual posee los componentes de los clientes (*front end*) y los componentes del servidor (*back end*).

Esta arquitectura es bastante conveniente, una vez que utiliza los sistemas existentes en la empresa y solo requieren de una pequeña inversión en hardware y software.

Uno de los grandes problemas que existen en este tipo de arquitectura es el hecho de que no son escalables, lo que resulta, con el aumento del número de usuarios, en una baja ejecución de las aplicaciones cliente/servidor. Estas anomalías pueden ocurrir por el uso de estaciones clientes muy lentas y con muchos procesos corriendo simultáneamente.

1.2.1.5 Arquitectura en tres capas [7]

Para intentar solucionar los problemas de ejecución resultantes de la arquitectura de dos capas, existe una arquitectura de información en múltiples capas, como lo muestra la figura 2.5. Esta arquitectura es bastante flexible y soporta un número grande de servicios integrados, donde la interfaz del usuario (herramientas *front end*), las funciones de procesamiento del negocio y las funciones de administración de la BD están separadas en procesos, los cuales pueden ser distribuidos a través de la arquitectura de la información.

Este tipo de arquitectura es bastante utilizado, en la primera capa están las aplicaciones de la interfaz con los usuarios, que deben ser gráficas y basadas en redes. Los Datos y las reglas del negocio que pueden ser compartidos por toda la organización, así como la BD para el DWH quedan almacenados en servidores de alta velocidad en la segunda capa, la capa central. En la tercera capa están localizadas las fuentes de datos.

Analizando el ambiente del DWH, los servidores de BD y los servidores de aplicaciones de la capa central, proveen un acceso eficiente y rápido a los datos compartidos.

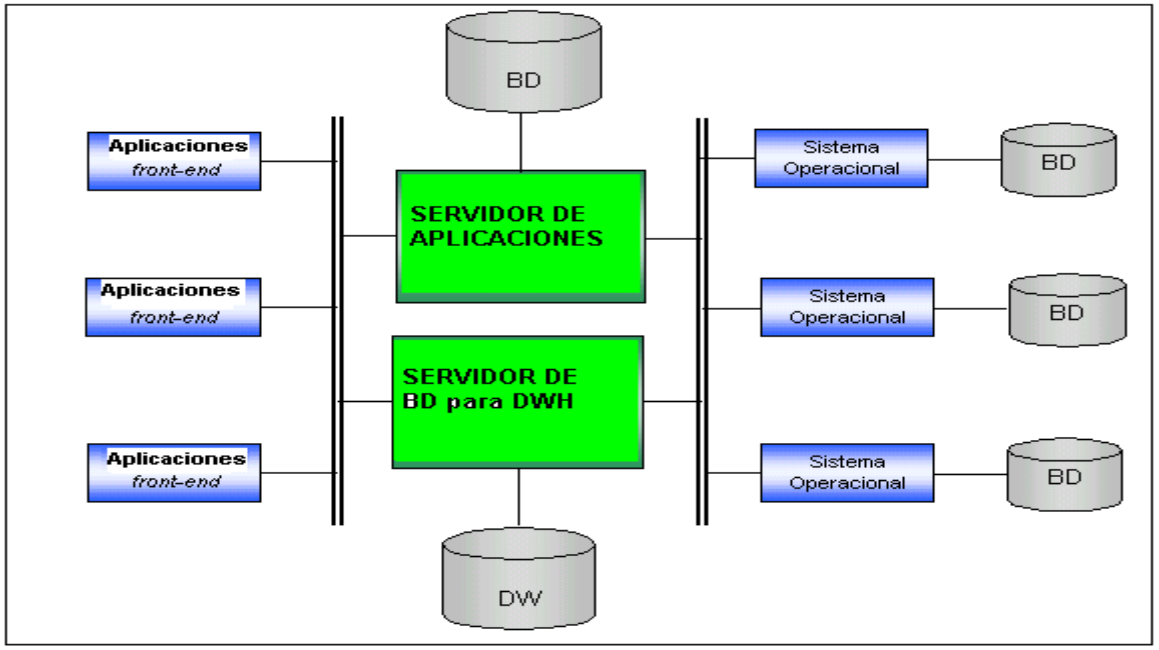

Figura 1.5. Arquitectura en tres capas

1.2.1.6 Un Modelo Alternativo de Arquitectura

Partiendo del modelo de capas descrito anteriormente, es posible definir un modelo alternativo que facilita el entendimiento y reduce el número de componentes a considerar, simplificando su descripción. En la figura 2.6 se puede observar un ejemplo de esta arquitectura [8], en la cual se visualiza en la parte inferior las bases de datos que forman las fuentes internas y externas, arriba de las fuentes de datos, están los módulos extractores que detectan automáticamente las alteraciones ocurridas en las bases de datos. Siempre que ocurre un cambio en el contenido de la base, la información que fue incluida o actualizada es procesada por el integrador que lo refleja en modificaciones en el DWH.

El integrador constituye el módulo responsable de la consolidación del contenido de los datos y del pase posterior para el DWH. Para integrar los datos originales de distintas bases de datos es necesaria la ejecución de algunos procesos, tales como, especificar las diferencias existentes entre las bases de datos, ajustar los datos al modelo conceptual utilizado en el DWH, uniéndolos a los datos ya cargados, solucionar los problemas de duplicaciones e inconsistencias y definir la forma en que las informaciones serán integradas dentro del sistema.

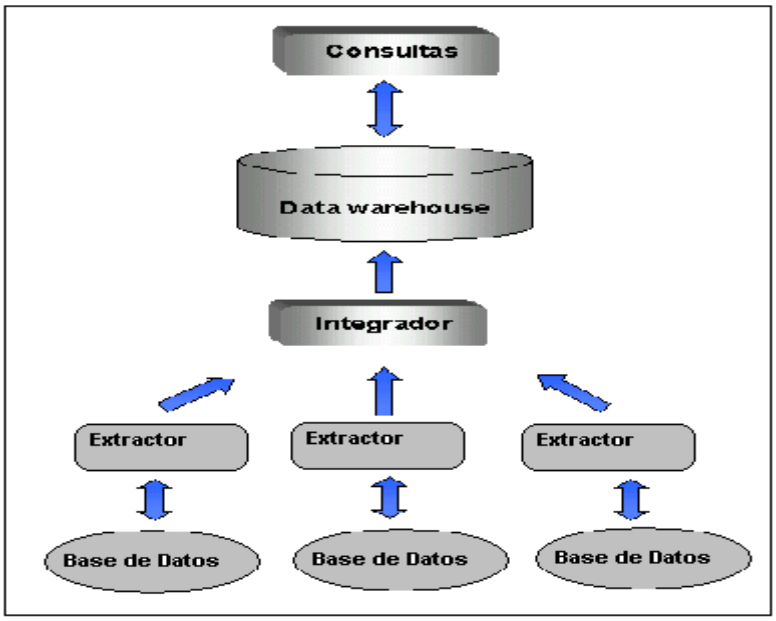

Figura 1.6. Arquitectura Alternativa

1.3. Modelo Dimensional

Existe un gran número de enfoques sobre el modelado de los datos, la mayoría de estos modelos pueden usarse para construir un Data Warehouse. Dentro de estos modelos hay dos que se destacan, uno fue escrito por R. Kimball y divide el modelado de los datos en tres partes: Modelo Empresarial, Modelo Dimensional y Modelo Físico [9]. El otro fue presentado por W.H. Inmon quién también divide el modelado en tres partes: Modelo de Alto Nivel, Nivel Intermedio y Nivel Bajo [10].

El modelo dimensional obtiene datos de un proceso comercial, como marketing o ventas y lo organiza por dimensiones, como servicios o centros de costos. El modelo dimensional se representa por una matriz multidimensional o un hipercubo, los valores de la matriz se llaman dimensiones y funcionan como restricciones o como títulos de líneas.

En un análisis de ventas, los hechos pueden ser cantidades de servicios vendidos o el total del importe de las ventas y las dimensiones pueden ser los nombres de los centros de costos o las Filiales, estas últimas ubicadas en territorios distintos. Una vez que los analistas empresariales tienden a considerar los datos desde punto de vista histórico, normalmente el tiempo también es una dimensión.

Se puede decir que la modelación dimensional, es un nuevo nombre para una técnica antigua, para hacer simples y compresibles las bases de datos.

¿Es necesario modelar?

La modelación de los datos tiene un papel fundamental para el desarrollo de un DW, la modelación dimensional se hace necesaria a la hora de la visualización de los datos, el secreto de comprender esta necesidad, está en la habilidad de mostrar algo como la abstracción de un conjunto de datos, en una manera concreta y tangible.

Ejemplo:

La Sucursal de Emprestur S.A. en Cienfuegos vende servicios en varios territorios y la gerencia mide el desempeño sobre el tiempo.

Entonces los diseñadores del Almacén de Datos, después de analizar la frase anterior la convierten en la siguiente frase, con énfasis en las palabras resaltadas en negrito.

La Sucursal de Emprestur S.A. en Cienfuegos vende **servicios** en varios **territorios** y la gerencia mide el desempeño sobre **el tiempo**.

Es fácil imaginarse esta afirmación como un Cubo de datos de tres dimensiones, donde cada palabra resaltada, representa un eje cartesiano.

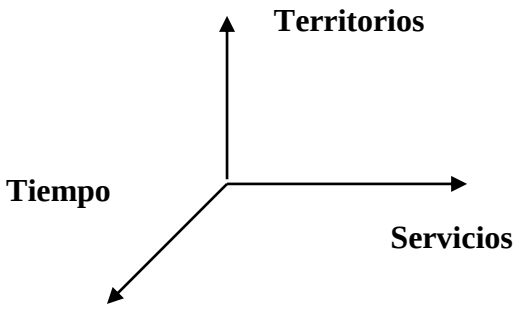

Gráfico No.1: Ejemplo de Dimensiones

En este gráfico, podemos suponer que los puntos dentro del cubo, están donde las mediciones del negocio para esa combinación Servicio, Territorio y Tiempo son almacenados. Este es el modelo dimensional, donde las variables Servicios, Territorio y Tiempo son las dimensiones de este modelo.

La clave fundamental que posibilita el éxito del modelo dimensional es su simplicidad, esta simplicidad permite a los usuarios comprender las bases de datos y posibilita al software que se utilice, navegar eficientemente por ellas.

En la mayoría de las formas el proceso de diseño dimensional conlleva a adoptar la **posición del fuerte,** además de atacar la simplicidad. Ahora si renunciamos al principio de la compresión del usuario y el desempeño del software, se puede mantener un diseño coherente de la base de datos que sirva a las necesidades de un Almacén de Datos.

En su libro "The Data Warehouse Toolkit", Ralph Kimball, deja bien claro que *"La distinción entre el modelo dimensional y el modelo de dependencia de los datos, está en el mismo centro del diseño del almacén de datos"* [11].

La ventaja principal de este tipo de sistema se basa en su concepto fundamental, la estructura de la información. Este concepto significa el almacenamiento de información relativa a un tema en común y confiable, en una estructura basada en la consulta y el procesamiento jerarquizado de la misma, además de un entorno bien diferenciado de los sistemas legados.

Según Bill Inmon, el Almacén de Datos se caracteriza por ser: [12].

Integrado: los datos almacenados en el Data Warehouse deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas legados deben ser eliminadas. La información debe estructurarse también en distintos niveles de detalle para adecuarse a las exigencias del usuario.

Temático: los datos para que se genere el conocimiento, se integran partiendo de los sistemas legados conocidos también como sistemas operacionales, estos datos se organizan por su naturaleza, es decir al tema que reponden, para facilitar su acceso y compresión por parte de los usuarios finales, quien en definitiva son los que toman las decisiones. Por ejemplo todos los datos de los proveedores pueden ser consolidados en una única tabla del Data Warehouse, haciendo más fácil la petición de información de los proveedores, ya que toda la información radica en un solo lugar.

Histórico: de la información contenida en un Almacén de Datos el tiempo es una parte implícita, a diferencia de los sistemas operacionales, en el Data Warehouse, la información almacenada sirve, entre otras cosas, para realizar análisis de tendencias, por tanto, el Almacén de Datos se carga con los distintos valores que toma una variable en el tiempo, para permitir comparaciones.

No volátil: la información en un Almacén de Datos, existe para ser leída y no modificada, es permanente, actualizando siempre con la incorporación de los nuevos valores, sin la modificación de los ya existentes.

Contiene datos relativos a los datos: este concepto viene asociado al término de metadatos, estos permiten mantener información de la procedencia de los datos, la periodicidad de recarga, su fiabilidad, forma de cálculo, entre otras, todas relativas a los datos del Almacén de Datos.

La función de los metadatos es la de simplificar y automatizar la obtención de la información desde los sistemas legados a los sistemas informacionales, estos tienen los siguientes objetivos.

- Soportar al usuario final, ayudandole a acceder al DataWarehouse con su propio lenguaje de negocio, indicando la información existente y su significado, también ayuda a construir consultas, informes y análisis mediante las herramientas de navegación.
- Permitir a los administradores del Data Warehouse, realizar auditorias, gestionar la información histórica, administración del Almacén de Datos, elaboración de programas para la extracción de la información.

Ralph Kimball por su parte plantea los siguientes objetivos de un Almacén de Datos, los cuales apoyan lo ya dicho por Bill Inmon [13].

1.- *El Almacén de Datos provee acceso a los datos empresariales.*

Como acceso entendemos las siguientes situaciones: El personal necesitado de estos datos, en una organización, tiene que ser habilitado a conectarse con el almacén de datos, desde sus puestos de trabajos mediante computadoras personales. Esta conexión tiene que ser inmediata en demanda y con un grado de desempeño aceptable.

2.- *Los datos en un almacén de datos son consistentes.*

La consistencia se basa en, que, cuando dos personas requieren un dato, ellos obtienen ese mismo dato, sin importar el tiempo en que lo soliciten, es decir si el jefe del centro de costo de Carpintería de Aluminio, necesita saber cual fue su producción en el mes de enero del año 2000, y está corriendo el mes de julio, sea el mismo que el Director de la Sucursal Cienfuegos solicite, para saber la producción de la Carpintería de Aluminio en el mes de enero del año 2000, desde el mes de enero del año 2001.

3.- *Los datos en el almacén de datos pueden ser separados y combinados por medio de cualquier dimensión en el comercio (los requerimientos clásicos de* ***cortar*** *y* ***formar cubos****).*

Este requermiento de cortar y formar cubos, habla directamente del enfoque dimensional.

4.- *El almacén de datos no es solamente dato, sino además un conjunto de herramientas para consultas, analizar y presentar la información.*.

El hardware central del almacén de datos, el software de la base de datos relacional y los datos en si (denominados como el "cuarto trasero"), según Kimball en su libro "The Data Warehouse Toolkit", forman sólo el 60 % de lo que se necesita para un almacén de datos, el 40 % restante es el conjunto de herramientas del "cuarto delantero" que consultan, analizan y presentan los datos.

5.- *El almacén de datos es el lugar donde publicamos los datos usados*.

La responsabilidad de publicar está en el mismo centro del almacén de datos. El dato no solo se procesa, se acumula y luego se deja perder, en lugar de eso el dato es celosamente ensamblado, partiendo de una variedad de fuentes de información (sistemas legados)

relativas a la organización, adquirido, asegurando su calidad y luego publicado para su uso adecuado. El dato inseguro o incompleto es responsabilidad del jefe de la calidad de dato, quien no debe permitir que este sea publicado al resto de los usuarios.

6.- *La calidad de los datos en el almacén de datos es una guia a utilizar en la reingeniería de la empresa.*

El almacén de datos no puede asegurar datos de mala calidad, si un empresario X no necesita un dato Y, y luego por una situación Z, necesita del dato Y, entonces no hay nada que hacer con el almacén de datos, en la busqueda de este dato Y, para explicar la situación Z.

Kimball define el proceso de diseño e impletanción de un Almacén de Datos como el proceso de montaje de los datos y dice *"es un proceso importante que incluye, entre otros subprocesos los de, Extración, Transformación, Carga y Garantía de Calidad de los Datos*" [14].

Para comprender el concepto de Data Warehouse, es importante tener presente los procesos que lo conforman.

- **Extracción**: obtención de información de las distintos sistemas legados.
- **Elaboración**: filtrado, limpieza, depuración, homegenización y agrupación de la información.
- **Carga**: organización y actualización de los datos y metadatos en la base de datos.
- **Explotación**: extracción y análisis de la información en los distintos niveles de agrupación.

En la figura 1.7, se muestran estos pasos.

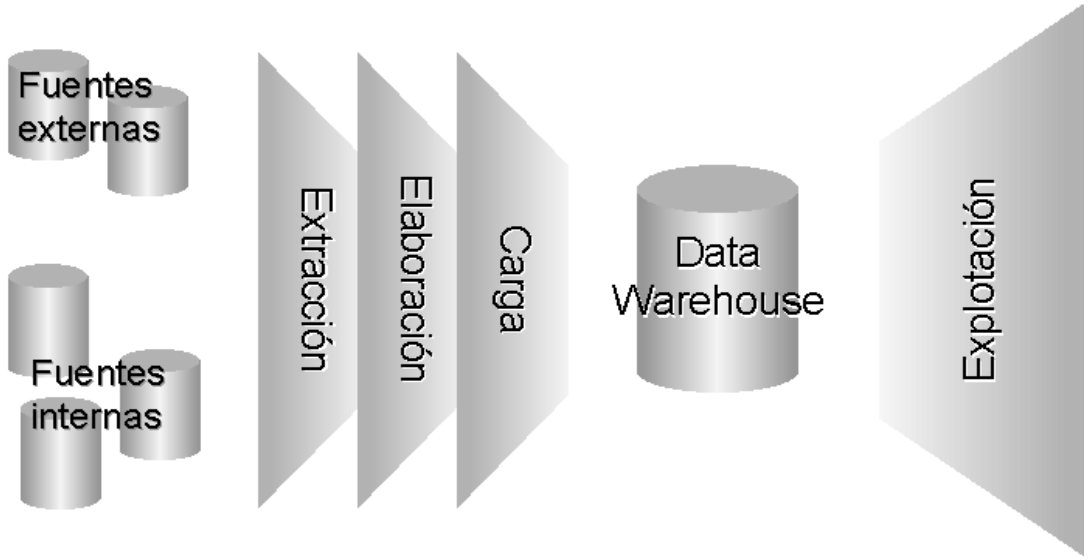

Figura 1.7. Pasos para el Diseño y Explotación de un DW.

Para el usuario el único proceso visible es el proceso de explotación del Data Warehouse, aunque el éxito del Almacén de Datos está en que los tres procesos iniciales son los que alimentan la información del mismo y suponen el mayor porciento de esfuerzo a la hora de desarrollar el Data Warehouse.

Una de las claves del éxito en la construcción de un Almacén de Datos, es el desarrollo de forma gradual, seleccionando a un departamento como piloto y expandiendo progresivamente el Data Warehouse a los demás usuarios, este proceso es conocido como cosntrucción de Data Marts, un DWH es la suma de todos los DM departamentales o de procesos.

Para la construcción de los DMs, se deben cumplir los pasos siguientes:

- Selección del Proceso
- Análisis de las necesidades de los usuarios
- Determinación del Nivel de Agregación
- Definición de las dimensiones del Modelo Estrella

- Definición de los hechos
- El Almacenamiento de valores calculados en la tabla de hechos
- Diseño de las tablas de dimensiones
- Selección de la durabilidad de los datos.

Selección del Proceso.

Es necesario definir en que sector de la empresa será iniciado el proceso de implantación del DWH, en este sector será implantado un DM que contendrá las informaciones sectoriales o departamentales que luego serán integradas a otros DMs de la empresa, este proceso debe ser gradualmente y sin perder la esencia del DWH empresarial.

Análisis de las necesidades de los usuarios.

Una de las actividades del Análisis de Sistema es la definición del problema, con base a la realización de una serie de entrevistas y reuniones con los colaboradores de la empresa que están implicados en el proyecto del sistema de información que se está desarrollando.

En la implantación de sistemas de DWH este tipo de actividad funciona de la misma forma, los diseñadores deben buscan en los distintos departamentos de la empresa las necesidades de información de sus directivos, además de esto, se debe buscar en las bases de datos transaccionales, las informaciones que están disponibles, para que las preguntas que podrán hacer los directivos puedan tener respuestas. Las expectativas de los usuarios finales va aumentando en medida que estos tengan una mayor visión del poder estos sistemas, convirtiendo mucho más importante la tarea de los diseñadores.

Determinación del Nivel de Agregación.

El entendimiento del nivel de agregación es dado por la definición de la tabla de hechos, las relaciones entre las tablas de hechos y la dimensiones es lo que permite la implementación del modelo en estrella. Físicamente, cada una de las relaciones entre las dimensiones y la tabla de hechos se realiza a través de una llave múltiple.

Definición de las dimensiones del Modelo Estrella

Las dimensiones constituyen la base para definir los encabezados de las filas de los reportes a los usuarios finales, estas componentes del modelo estrella deben estar de acuerdo con las denominaciones tradicionales, utilizadas por los usuarios, con el fin de que sean rápida y ampliamente comprendidas.

Es posible dividir los modelos implicados en el proceso de transformación de los datos de ambientes operacionales para ambientes de apoyo a la toma de decisión en tres tipos: el Modelo Corporativo, el Modelo de Datos del DWH y el proyecto Departamental, más conocido como Data Mart. El Modelo de datos Corporativo representa una visión de los datos de la empresa y de sus relaciones, constituyendo la base del desarrollo de los sistemas transacionales y analíticos, el modelo de DWH difiere del corporativo por el hecho de que es orientado a los negocios, siendo más integrado y basado en una estructura que permita el soporte a la decisión. En este tipo de modelo existe una arquitectura en las informaciones extraídas de las bases operacionales deberán pasar por varios procesos antes de integrarse al DWH global de la empresa.

El DM posee informaciones relativas a un sector de la empresa, estos datos pueden no estar organizados de la misma forma que estarían definidos en la visión general del DWH de toda la organización, aunque es recomendable que los DMs deben extraer las informaciones del DWH global. El punto de partida para el proyecto y construcción de un DWH es el modelo corporativo de la empresa, independientemente de que su estructura sea relacional o multidimensional. [15]

Uno de los grandes problemas de las técnicas clásica de modelación es la falta de distinción entre un modelado de un ambiente transacional y uno de ambiente de soporte a la toma de decisión. Existen métodos a seguir, como el de W. Inmon, enunciados en el *Modelo Dimensional* que definió tres modelos: Un modelo de Alto Nivel, definido por el diagrama Entidad Relación (DER), un Nivel Intermediario, definido por el conjunto de artículos de datos (*data information set*) y un Nivel Bajo correspondiente al modelo Físico.

La finalidad del DER es la misma para un DWH y los sistemas relacionales, o sea, el modelo ER es basado en una percepción del mundo real que consiste en un conjunto de objetos básicos llamados entidades y la relación entre estos objetos. De esta forma, el DER promueve el registro del mundo real de tal forma que propicia al usuario una visión de cómo se comporta el sistema. El DIS es un modelo más complejo, el cual puede unir conjuntos de tablas de diferentes sistemas transacionales. El modelo físico es definido como la forma en que el DWH será implementado. [16]

Una dimensión es la representación de los Items de tipos semejantes y normalmente se presentan a lo largo del tiempo. Existe un consenso de que la parte más importante de las dimensiones es la composición de sus atributos y la mejor forma de manipularlos. En la fase del diseño de un DWH se debe poner atención a la definición de las dimensiones y de sus atributos, ya que la calidad del banco de datos es proporcional a los atributos de la dimensión.

Durante la definición de la granularidad de las tablas de hechos es donde se conocen por primera vez las dimensiones del DWH. De acuerdo con la definición de las informaciones a analizar por los usuarios es posible seleccionar cuales serán los atributos de estas dimensiones. Las informaciones son los elementos que forman la tabla de hechos y las formas de presentar esta información son las dimensiones.

En el detalle de las dimensiones que serán utilizadas en el DWH se deben seguir estos dos pasos:

- **Fijar las Dimensiones**

La forma más simple e inmediata de definir las dimensiones de un DWH se hace a través del análisis del grado de granularidad de las tablas de hechos, otra forma bastante útil es intentar descubrir con el propio usuario aquellas informaciones que le son necesarias, para poder hacer una interpretación correcta de los datos del sistema.

- **Búsqueda de informaciones en los sistemas transacionales.**

La selección de los datos a partir de los sistemas legados puede convertirse en una tarea muy dura, una buena estrategia es la de intentar identificar las fuentes de los datos, definiendo los procesos que se relacionan con cada dimensión, verificando donde es que ubicados y cuales son los atributos que realmente interesan para el DWH.

La dimensión Tiempo

Uno de los factores más importantes para la implementación de sistemas de DWH es el cuidado que se debe tener al definir una dimensión de tiempo. Una vez que por definición los sistemas de análisis de datos se basen en informaciones históricas, tendrán almacenados en sus bases de datos varios períodos de informaciones originadas de las bases transacionales, una definición adecuada de estos períodos de tiempo es de vital importancia.

En caso de que sea implementado un DWH sin la dimensión tiempo, no será posible procesar factores importantes al analizar los datos, como el mes del año donde el nivel de producción es bajo, o en que mes del año un cliente puede ejecutar una inversión donde requiere de nuestros servicios. Para este tipo de análisis es necesario almacenar atributos relativos a las ventas por cada cliente contratado.

- **Integrando los Datos**

Después de la definición de las informaciones y de las tablas que serán utilizadas para componer las dimensiones del modelo, es necesario integrar estos datos, validando sus fuentes en cuanto a la disponibilidad de los datos contenidos en estas fuentes y verificando también su calidad. Las distintas formas en que estas informaciones son consultadas también es una característica importante de la integración, además de la visión general de los sistemas que están funcionamiento en la empresa, se debe tener presente aquellos sistemas que ya no están en explotación o que fueron trasladados a otras plataformas, tanto de hardware como de software, ya ellos pudieran contener informaciones útiles a los usuarios. El tratamiento de las informaciones es útil para que se pueda seleccionar los items

que realmente interesan tener en el DWH. La selección de las informaciones se hace a través de los siguientes pasos:

Analizando los Datos Operacionales

Los sistemas transacionales necesitan de una serie de informaciones de control que no son útiles al DWH y que deben ser ignoradas durante la carga del DWH. Algunos de los datos referentes a los códigos numéricos o siglas, que posibilitan una relación optimizada entre las diferentes tablas, pueden ser sustituidos por textos legibles alterando los códigos existentes en las dimensiones.

Existen también informaciones operativas que son constantemente actualizadas y que dificultan la interpretación de sus contenidos en consultas, una vez que se encuentran incompletas, dentro de estas informaciones podemos citar, las banderas, los campos de estado, totalizaciones parciales de valores.

Desnormalización de Relaciones

El proceso de normalización sirve para modelar el diagrama ER en el sentido de definir entidades que son desmembradas de una tabla, tales como clientes que tienen una especialidad y la ciudad donde viven. En el caso de los sistemas de apoyo a la toma de decisión la característica de organizar las informaciones en tablas distintas y relacionadas entre si, puede ser una desventaja, pues la complejidad para el tratamiento de estas informaciones que ahora están dispersas en varias tablas puede comprometer la ejecución del sistema. Se debe entonces aplicar un proceso inverso al de la normalización, una vez que los sistemas de DWH deben ser lo más desnormalizados posibles. En los DWH es reducida la necesidad de utilizar consultas del tipo *joins,* para eliminar la normalización se pueden unir tablas relacionadas en el proceso transacional. Si por un lado se pierde la normalización, por otro se gana una mejor ejecución en las consultas por el hecho de que la estructura de los datos fue simplificada. [17]

Los siguientes tipos de tablas que deben ser desnormalizadas:

- ✓ Tablas que comparten una llave común o parcial
- ✓ Tablas diferentes, en las cuales sus datos son frecuentemente utilizados juntos.
- ✓ Tablas donde el patrón de inserción es el mismo

Estableciendo el patrón para los atributos.

Los patrones de los datos en sistemas DWH normalmente se refieren al tipo y al tamaño. Existen campo del tipo Fecha, con diversos formatos, dependiendo del banco de datos que se utilice como fuente de datos. Un determinado sistema puede adoptar las Fechas como una cadena(string) con el formato DDMMAAAA, donde las **A** representan a los años, las **M** a los meses y las **D** a los días, hay otro sistema que puede adoptar el formato de barras de división "/" y la fecha se representa de la siguiente forma DD/MM/AAAA. Como se puede notar en el ejemplo, es necesaria la conversión de los datos a solo un formato, este formato es predefinido por los diseñadores del DWH.

Estableciendo valores implícitos

Por la propia característica de los sistemas de apoyo a la decisión, contienen datos de periodos relativamente grandes en el tiempo, es posible que algunos datos no estuvieran definidos en determinados momentos representados en los archivos fuentes de los datos.

En caso de que algún atributo no posea valores para cargar el DWH, el sistema debe prever valores predefinidos que serán insertados para suplir su falta. Estos valores son los llamados implícitos, o sea, en informaciones que no sea posible determinar cual es el valor y que dato relativo al atributo en cuestión, entonces se inserta un valor implícito. Vale la pena resaltar que se pudieran insertar valores Nulos o ceros, aunque no se recomienda esta práctica, solo en casos extremos, ya que atenta contra la calidad de los datos.

Modelando las Dimensiones

Las llaves de entrada de las tablas de los sistemas transacionales raramente son utilizadas en los sistemas de apoyo a la decisión. En casos más simples se inserta una nueva

dimensión de tiempo, pero en los casos más complejos existe la necesidad de insertar claves basadas en procesos *Hashing* que facilitan el acceso a los datos.

Según W. Inmon se debe reestructurar las llaves en las siguientes situaciones [18]:

- ✓ Existe la posibilidad de que se produzca una alteración en la llave y no es deseado su reutilización.
- ✓ Es necesario rastrear modificaciones de los datos.
- ✓ Es necesario crear Items para describir otros Items agregados.

La reestructuración puede hacerse adicionando una llave de tiempo, esta tarea no es tan fácil como parece. Como ya se había descrito anteriormente, debe hacerse un análisis con respecto a los periodos de tiempo que se van a consultar posteriormente, un análisis incompleto puede llevar a resultados de consultas no satisfactorios para los usuarios finales. Es interesante notar también que las informaciones de fecha no deben ser asociadas a las tablas de hechos.

Los términos operacionales contenidos en la base de dato transacional deben ser sustituidos preferentemente por otros más vinculados a las actividades administrativas, en vez de ser tratados con términos técnicos de un área especifica de la empresa, una vez que los usuarios finales son los directivos y técnicos encargados de tomar decisiones.

- **Relación entre las Tablas de Dimensiones y de Hechos**.

El tipo de relación entre las dimensiones y la tabla de hechos puede ser de Mucho a Mucho, visto que el factor temporal es importante en sistemas de DWH. Se debe tener establecida la necesidad del uso de llaves foráneas entre estas las tablas de dimensiones y de hechos, además de controlar la integridad entre las tablas, ya que si una tupla de la tabla de dimensión fuera eliminada, se debe resolver los problemas causados por la exclusión a través de la tabla de hechos.

- **Definición de las jerarquías en las Dimensiones.**

Una de las características encontradas en las dimensiones es la posibilidad de existencia de jerarquías. Algunos autores consideran que una jerarquía es un atributo de una dimensión. Por ejemplo, la dimensión de tiempo puede ser divida en Horas, días, semanas, meses, años, o la dimensión de localidad o territorio puede ser dividida en ciudades, municipios, provincias, etc. La jerarquización de las dimensiones es hecha de acuerdo con los análisis hechos durante el diseño del sistema de apoyo, pues los diseñadores deben analizar con mucho tino cuales serian los beneficios y perjuicios que se pueden obtener de una adopción incorrecta de la dimensión. La definición de jerarquías puede traer ventajas en el momento en que los usuarios deseen hacer totalizaciones y visualizaciones resumidas de los datos.

Definición de los hechos

Los hechos constituyen el conjunto de registros de la tabla central del modelo estrella. De acuerdo con el nivel de granularidad de la tabla de hechos es posibles determinar cuales serán los hechos almacenados. Los nuevos hechos pueden ser adicionados en cualquier momento del modelo, sin comprometer las características ya implementadas.

Los sistemas DWH deben prever e incorporar las modificaciones de las características que van surgiendo a través del tiempo.

Granularidad.

Uno de los puntos fundamentales que debe ser profundamente analizado en el proyecto de un sistema de apoyo a la decisión es la cuestión de la granularidad.

Una definición del nivel de los sistemas de apoyo se basa principalmente en el grado de granularidad de los datos, está puede traer por un lado un costo muy elevado e innecesario, en caso de que el nivel de granularidad sea bajo ya que son muchos los datos almacenados de forma no resumida.

Por otro lado un alto nivel de granularidad puede resumir de tal forma los datos que no será posible realizar las consultas mas detalladas en las informaciones contenidas en la base de datos del DWH.

La definición del nivel de granularidad del sistema depende fundamentalmente de los datos que están disponibles en las fuentes de datos, que serán cargadas al DWH. Para sortear este problema pueden ser adoptadas algunas de las siguientes metodologías, adaptar las fuentes de datos para el mismo nivel de granularidad, procurar trabajar con el menor nivel posible que sea común entre las fuentes distintas, siendo esta ultima la que más se adopta por los diseñadores, ya que con ella son descartados muchos datos a analizar.

El punto principal de la definición de un DWH está en definir un equilibrio entre la necesidad real de los niveles de granularidad para el usuario y el costo de la implementación

El costo relacionado al tiempo de vida útil del DWH debe ser considerado, porque el costo de almacenamiento de estos datos puede ser prácticamente despreciado, teniendo en cuenta las constantes reducciones de los costos de los medios de almacenamiento físico.

Algunas técnicas de definición de la granularidad son sugeridas; Confección de prototipos, Reuniones de Retroalimentación, Análisis de los datos disponibles, entre otras. Aunque se puede definir un nivel granularidad dual, decir podemos tener un nivel de detalle básico, y podemos tener resumen, por ejemplo, tener todos los movimiento del mes de cada cuenta y al final de mes el saldo mensual.

El Almacenamiento de valores calculados en la tabla de hechos

Existen algunos valores que deben ser almacenados de forma redundante, ya que esto representa algunos gigabyte de más en la base de datos. Pero agilizan las consultas que puedan hacer. Estos valores pueden ser por ejemplo los precios de acuerdo con el margen

de ganancia obtenida en la venta, para poder confeccionar un informe sobre las pérdidas y las ganancias.

La definición de un campo calculado es muy importante, debido a que se debe tener en cuenta el costo beneficio de las consultas con respecto a los datos almacenados.

Diseño de las tablas de dimensiones

En este punto se indica al diseñador que puede adicionar el mayor número de textos posibles con el fin de que el usuario entienda los datos.

La forma de explicar esto respecto a los datos, debe ser más objetiva y clara posible, utilizando los términos entendidos por los usuarios de forma amplia, ya que son ellos en definitiva los que manipulan la información. El diseñador debe tener presente que los datos siempre son utilizados para la consultas y que estos deben estar escritos en los términos del usuario.

Selección de la durabilidad de los datos.

La definición del intervalo de tiempo que estarán los datos a la disposición de las consultas y comparaciones, constituye un factor importante para el la realización del diseño de un sistema de apoyo a la toma de decisión. Conocer el comportamiento de los datos en el transcurso del tiempo puede tornarse en un triunfo para las empresas que utilizan este tipo de sistema, una vez que las consultas podrán ser confeccionadas para confirmar determinadas alteraciones de algunas ventas.

Muchas empresas almacenan varios años en sus sistemas de DW, ciertamente invirtiendo en equipos y tecnología de procesamiento masivo de datos a fin de obtener tiempos de respuestas adecuadas a bases de datos de algunos Gigabyte o en algunos acasos de Terabytes.

Un factor que debe observa el diseñador es la posibilidad de obtener efectivamente los datos de tiempos anteriores, los cuales pueden estar almacenados en medios que muchas

veces no disponibles. Pueden también estar almacenados en otros formatos que no son utilizados por la empresa o en equipos que no están en uso en la actualidad.

Otro factor importante es la posibilidad de que los datos contenidos en las dimensiones estén configurados con situaciones antiguas, como por ejemplo servicios que no se ofrecen o centros de costos cerrados por su improductividad.

Beneficios que puede aportar un Almacén de Datos.

- Proporciona una herramienta para la toma de decisiones, en cualquier área funcional, basandose en información integrada y global del negocio.
- Facilita la aplicación de técnicas estadísticas de análisis y modelación, para encontrar relaciones ocultas entre los datos del Data Warehouse, obteniendo un valor añadido para el negocio de dicha información.
- Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.
- Simplifica dentro de la empresa, la implantación de sistemas de gestión integral con relación al cliente.
- Supone una optimamización tecnológica y económica en la generación de informes, con retornos de la inversión imprevistos.

Data Warehousing es el proceso de extraer y filtrar datos de las operaciones comunes de la empresa, procedentes de los distintos subsistemas operacionales, para transformarlos, integrarlos, totalizarlos y almacenarlos en un depósito o repositorio, para poder acceder a ellos cada que vez que se necesite. Podemos concebir un Data Warehouse como un almacén-factoría de datos o información, que concentra la información de interés para toda la organización y distribuye dicha información por medio de diversas herramientas de consulta y de creación de informes orientadas a la toma de decisiones. Con esta tecnología se convierten los datos operacionales de una organización en una herramienta competitiva, que permite a los usuarios finales examinar los datos de modo más estratégico, realizar análisis y detección de tendencias, seguimiento de medidas críticas, producir informes con

mayor rapidez, un acceso más fácil, más flexible y más intuitivo a la información que se necesite en cada momento. Frecuentemente, datos que son difíciles de interpretar, desde varias fuentes, se convierten en información lista para el usuario final, otorgando así una mayor ventaja competitiva a la organización.

El Almacén de Datos sustenta su teoría en la del Modelo Dimensional, a continuación, brindamos una descripción de este modelo, y algunas de sus características más importantes.

El Modelo Dimensional (esquema de enlace estrella).

Es un modelo dimensional muy asimétrico, existe aquí una gran tabla en el centro del esquema quien domina todas la relaciones, es decir con múltiples enlaces conectándola a otras tablas, de ahí su nombre de modelo estrella, el resto de las tablas tienen todas un enlace simple con la tabla central, a esta tabla central se le denomina **tabla de hechos** y a las otras **tablas dimensionales**.

Los sistemas de DWH pueden tener varias tablas de hechos, cada representando un proceso diferente dentro de la empresa, constituidos los DM, estos se pueden enlazar unos a otros dependiendo de la necesidad y también de la posibilidad de que esto acontezca.

Las tablas de hechos son enlazadas a través de las relaciones de las tablas de dimensiones utilizando llaves. Estas tablas son en mucho menores en tamaño y número de registros que la tabla de hechos a la que ellas se enlazan. Cada tabla de dimensión tiene una única llave y sus campos son típicamente de texto, vale señalar, que en el DM diseñado, los campos llaves son numéricos, esto se debe a la comodidad que supone el uso del Analysis Service, las dimensiones se utilizan como fuente de encabezamiento de los informes que se brindan.

Un esquema estrella se basa en dos tipos de consultas: *browse* y *join multitablas*

Una consulta del tipo *browse* es definida para aplicarse en una sola tabla, sin que sea necesario utilizar los comando *join*, por ejemplo, ver un clasificador de Clientes.

Las consultas con *joins multitablas* son precedidas por una serie de *browses* que hacen uso de la estructura del modelo estrella a través de diversas uniones entre tablas de hechos y las dimensiones.

En el modelo dimensional, primero son identificados los procesos empresariales que serán la base para el diseño de las tablas de hecho. Un modelo Entidad Relación habitual las tablas deben sufrir un proceso de normalización, ahora si las tablas de dimensión fueran normalizadas en estructuras de "copo de nieve", donde las dimensiones están compuestas por más de una tabla, pueden surgir dos problemas. Primero, el modelo de datos es bastante complejo para se presentado a los usuarios. Segundo, la unión entre las diversas partes del "copo de nieve" comprometerá el desempeño del sistema como un todo.

La Tabla de Hechos.

La tabla de hechos es donde las mediciones numéricas del negocio son almacenadas, cada una de las mediciones se toman como la intersección de todas la dimensiones. Los mejores y más útiles hechos son numéricos, valorados continuamente y aditivos, la razón para afirmar esto, es que virtualmente en cada consulta hecha contra esta tabla de hechos, estamos preguntando cientos, miles o millones de artículos a ser usados por el Sistema Gestor de Bases Datos, para construir el conjunto de respuesta, todos estos registros son comprimidos a unas pocas filas del conjunto respuesta del usuario, e irremediablemente, la única forma realmente válida para compactar estos artículos en el conjunto de respuesta es adicionándolos, por tanto, si las mediciones son numéricas y si ellas son aditivas, podemos fácilmente construir el conjunto de respuesta del usuario [19].

Las Tablas Dimensionales.

Las tablas dimensionales son aquellas donde las descripciones textuales de las dimensiones del negocio son almacenadas, estas descripciones ayudan a detallar cada miembro de la dimensión que se trate. En una base de datos bien diseñada, la tabla dimensional del servicio tendrá muchos campos, los mejores campos son los de tipo texto para usuarios como la fuente de las restricciones y encabezados de filas en el conjunto respuesta del

usuario, debido al papel que juegan los campos de la dimensión, el de describir uno de los elementos de una dimensión, son más útiles si son textos. Pueden existir campos en estas tablas de dimensión que no sean texto, sino, numéricos, pero son complementos de la descripción del servicio.

Componentes a tener en cuenta a la hora de construir un Almacén de Datos.

Antes de diseñar un proyecto para la construcción de un Almacén de Datos, es bueno estimar el tamaño de la base de datos, para poder planificar el hardware y software a utilizar en la implementación y explotación de un Data Warehouse [20].

Hardware

Un componente esencial a la hora de poder contar con un Data Warehouse que responda a las necesidades analíticas avanzadas de los usuarios, es el poder contar con una infraestura hardware que la soporte, en este sentido son críticas, a la hora de evaluar uno otro hardware, dos características principales:

Por un lado, a este tipo de sistemas suelen acceder pocos usuarios con unas necesidades muy grandes de información, a diferencia de los sistemas operacionales, con muchos usuarios y necesidades puntuales de información. Debido a la flexibilidad requerida a la hora de hacer consultas complejas e imprevistas, y al gran tamaño de información manejada, son altas prestaciones de la máquina.

Por otro lado, debido a que estos sistemas suelen comenzar con una funcionalidad limitada, que se va expandiendo con el tiempo (situación aconsejada), es necesario que los sistemas sean escalables para dar soporte a las necesidades crecientes de equipamiento. En este sentido, será conveniente optar por una arquitectura abierta, que nos permita aprovechar lo mejor de cada fabricante.

Software de almacenamiento (SGDB)

El sistema que gestione el amacenamiento de la información (Sistema de Gestión de Bases de Datos o simplemente SGBD), es otro elemento clave en un data warehouse, independientemente de que la información almacenada en el Almacén de Datos se pueda analizar mediante visualización multidimensional, el SGBD puede estar realizado utilizando tecnologías de Bases de Datos Relacionales o Multidimensionales.

Las bases de datos relacionales, se han popularizado en los sistemas operacionales, pero se han visto incapaces de enfrentarse a las necesidades de información de los entornos DW, como ya hemos visto anteriormente, también debido a que las necesidades de información suelen atender a consultas multidimensionales, todo indica, que las Bases de Datos Dimensionales llevan ventaja. En este sentido son de aplicación los comentarios hechos en el hardware, por requerimientos de prestaciones, escalabilidad y de consolidación tecnológica.

Software de extracción y manipulación de datos.

En este apartado analizaremos un componente esencial a la hora de implantar un Almacén de Datos, la extracción y manipulación. Para esta labor que entra dentro del ámbito de los profesionales de tecnologías de la información, es crítico el poder contar herramientas que permitan controlar y automatizar las necesidades de actualización del DW.

Estas herramientas deberán proporcionar las siguientes funcionalidades:

- Control de la extracción de los datos y su automatización, disminuyendo el tiempo empleado en el descubrimiento de procesos no documentados, minimizando el margen de error y permitiendo mayor flexibilidad.
- Acceso a diferentes tecnologías, haciendo un uso efectivo del hardware, software, datos y recursos humanos existentes.

- Proporcionar la gestión integrada del *Data Warehouse* y los *Data Marts* existentes, integrando la extracción transformación y carga para la construcción del Almacén de Datos y de los Mercados de Datos.
- Uso de la arquitectura de los metadatos, facilitando la definición de los objetos de negocio y las reglas de consolidación.
- Acceso a una gran variedad de fuentes de datos diferentes.
- Manejo de exepciones.
- Planificación, *logs*, interfaces a esquemas de terceros.
- Interfaz independiente del Hardware.
- Soporte en la explotación del Almacén de Datos.

A veces, no se suele prestar la suficiente atención a esta fase de la gestión del Data Warehouse, aún cuando supone una gran parte del esfuerzo en la construcción de un Almacén de Datos.

Dentro de las técnicas de explotación de la implantación de un DW mencionamos las siguientes:

1. El uso que se puede realizar de las utilidades OLAP del Almacén de Datos para análisis multidimensionales.
2. Las facilidades de obtención de información mediante consultas e informes libres, y el uso de técnicas de Minería de Datos que nos permitan descubrir información oculta en los datos, mediante el uso de técnicas estadísticas.

OLAP

La explotación del DW mediante información de gestión, se fundamenta básicamente en los niveles agrupados o calculados de información, la información de gestión se compone de conceptos de información y coefcientes de gestión, que los cuadros directivos de la empresa pueden consultar según las dimensiones de negocio que se definan. Dichas dimensiones se estructuran a su vez en distintos niveles de detalle.

Los sistemas de soporte a la decisión usuando tecnologías de Almacenes de Datos, se llaman sistemas OLAP (**On Line Analytical Processing**), estos deben cumplir con los siguientes requerimientos:

- Soportar requermientos complejos de análisis.
- Analizar datos desde diferentes perspectivas.
- Soportar análisis complejos contra un volumen ingente de datos.

La funcionalidad de los sistemas OLAP se caracteriza por ser un análisis multidimensional de datos corporativos, que soportan los análisis del usuario y unas posibilidades de navegación, seleccionado la información a obtener.

Existen dos arquitecturas diferentes para los sistemas OLAP:

- OLAP multidimensional (MOLAP)
- OLAP relacionales (ROLAP)

La arquitectura MOLAP usa bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente.

La arquitectura ROLAP por el contrario cree que las capacidades OLAP están perfectamente implantadas sobre bases de datos relacionales.

Actualmente se nota más apertura para la creación de los reportes sofisticados, donde se permiten usar varios criterios de comparación con una presentación visual amigable para el usuario, como son las herramientas OLAP (on-line analytical processing, procesamiento analítico en línea) [21].

Consultas y Reportes Libres

Las consultas o informes libres trabajan tanto sobre el detalle como sobre las agregaciones de la información. Realizar este tipo de explotación en un DW supone una optimización del tradicional entorno de informes, dado que el Almacén de Datos mantiene una estructura y una tecnología más apropiada para este tipo de solicitudes.

Los sistemas de "Query and Reporting", no basados en almacenes de datos, se caraterizan por la complejidad de las consultas, los altisimos tiempos de respuestas y la interferencia con otros procesos informáticos que compartan su entorno.

La explotación del Almacén de Datos mediante "Query and Reporting" debe permitir una gradación de la flexibilidad de acceso, proporcional a la experiencia y formación del usuario. A este respecto, se recomienda el mantenimiento de tres niveles de dificultad.

1. Los usuarios pocos expertos podrán solicitar la ejecución de informes o consultas predefinidas según unos parámetros predeterminados.
2. Los usuarios con cierta experiencia podrán generar consultas flexibles mediante una aplicación que proporcione una interfaz gráfica de ayuda.
3. Los usuarios altamente experimentados podrán escribir, total o parcialmente, la consulta en un lenguaje de interrogación de datos.

Data Mining o Minería de Datos

El Data Mining es un proceso que, a través del descubrimiento y cuantificación de relaciones predictivas en los datos, permite transformar la información disponible en conocimiento útil de negocio. Esto es debido a que no es suficiente "navegar" por los datos para resolver los problemas de negocio, sino que se hace necesario seguir una metodología ordenada que permita obtener rendimientos tangibles de este conjunto de herramientas y técnicas de las que dispone el usuario. Constituye por tanto una de las vías clave de explotación del Data Warehouse, dado que es este su entorno natural de trabajo. Se trata de un concepto de explotación de naturaleza radicalmente distinta a la de los sistemas de

información de gestión, dado que no se basa en coeficientes de gestión o en información altamente agregada, sino en la información de detalle contenida en el almacén. Adicionalmente, el usuario no se conforma con la mera visualización de datos, sino que trata de obtener una relación entre los mismos que tenga repercusiones en su negocio.

Técnicas de Data Mining

Para soportar el proceso de Data Mining, el usuario dispone de una extensa gama de técnicas que le pueden ayudar en cada una de las fases de dicho proceso, las cuales pasamos a describir:

Análisis estadístico: Utilizando las siguientes herramientas: ANOVA: o Análisis de la Varianza, contrasta si existen diferencias significativas entre las medidas de una o más variables continuas en grupo de población distintos.

Regresión: define la relación entre una o más variables y un conjunto de variables predictoras de las primeras. Ji cuadrado: contrasta la hipótesis de independencia entre variables. Componentes principales: permite reducir el número de variables observadas a un menor número de variables artificiales, conservando la mayor parte de la información sobre la varianza de las variables.

Análisis cluster: permite clasificar una población en un número determinado de grupos, en base a semejanzas y diferencias de perfiles existentes entre los diferentes componentes de dicha población.

Análisis discriminante: método de clasificación de individuos en grupos que previamente se han establecido, y que permite encontrar la regla de clasificación de los elementos de estos grupos, y por tanto identificar cuáles son las variables que mejor definan la pertenencia al grupo.

Métodos basados en árboles de decisión: El método Chaid (Chi Squared Automatic Interaction Detector) es un análisis que genera un árbol de decisión para predecir el

comportamiento de una variable, a partir de una o más variables predictoras, de forma que los conjuntos de una misma rama y un mismo nivel son disjuntos. Es útil en aquellas situaciones en las que el objetivo es dividir una población en distintos segmentos basándose en algún criterio de decisión.

El árbol de decisión se construye partiendo el conjunto de datos en dos o más subconjuntos de observaciones a partir de los valores que toman las variables predictoras. Cada uno de estos subconjuntos vuelve después a ser particionado utilizando el mismo algoritmo. Este proceso continúa hasta que no se encuentran diferencias significativas en la influencia de las variables de predicción de uno de estos grupos hacia el valor de la variable de respuesta. La raíz del árbol es el conjunto de datos íntegro, los subconjuntos y los subconjuntos conforman las ramas del árbol. Un conjunto en el que se hace una partición se llama nodo. El número de subconjuntos en una partición puede ir de dos hasta el número de valores distintos que puede tomar la variable usada para hacer la separación. La variable de predicción usada para crear una partición es aquella más significativamente relacionada con la variable de respuesta de acuerdo con test de independencia de la Chi cuadrado sobre una tabla de contingencia.

Algoritmos genéticos: Son métodos numéricos de optimización, en los que aquella variable o variables que se pretenden optimizar junto con las variables de estudio constituyen un segmento de información. Aquellas configuraciones de las variables de análisis que obtengan mejores valores para la variable de respuesta, corresponderán a segmentos con mayor capacidad reproductiva. A través de la reproducción, los mejores segmentos perduran y su proporción crece de generación en generación. Se puede además introducir elementos aleatorios para la modificación de las variables (mutaciones). Al cabo de cierto número de iteraciones, la población estará constituida por buenas soluciones al problema de optimización.

Redes neuronales: Genéricamente son métodos de proceso numérico en paralelo, en el que las variables interactúan mediante transformaciones lineales o no lineales, hasta obtener unas salidas. Estas salidas se contrastan con los que tenían que haber salido, basándose en

unos datos de prueba, dando lugar a un proceso de retroalimentación mediante el cual la red se reconfigura, hasta obtener un modelo adecuado.

Lógica difusa: Es una generalización del concepto de estadística. La estadística clásica se basa en la teoría de probabilidades, a su vez ésta en la técnica conjuntista, en la que la relación de pertenencia a un conjunto es dicotómica (el 2 es par o no lo es).

Si establecemos la noción de conjunto borroso como aquel en el que la pertenencia tiene una cierta graduación (¿un día a 20ºC es caluroso?), dispondremos de una estadística más amplia y con resultados más cercanos al modo de razonamiento humano.

Series temporales: Es el conocimiento de una variable a través del tiempo para, a partir de ese conocimiento, y bajo el supuesto de que no van a producirse cambios estructurales, poder realizar predicciones. Suelen basarse en un estudio de la serie en ciclos, tendencias y estacionalidades, que se diferencian por el ámbito de tiempo abarcado, para por composición obtener la serie original. Se pueden aplicar enfoques híbridos con los métodos anteriores, en los que la serie se puede explicar no sólo en función del tiempo sino como combinación de otras variables de entorno más estables y, por lo tanto, más fácilmente predecibles.

WEBHOUSING

Todos coincidimos en afirmar que Internet se ha convertido en el acontecimiento más revolucionario del mundo de la informática, y ya algunos vaticinan que los cambios más significativos en el ámbito de los sistemas de información corporativos vendrán con la aplicación de la tecnología Internet, concretamente redes privadas vía Internet [22].

La popularización de Internet y la tecnología WEB, ha creado un nuevo esquema de información en el cual los clientes tienen a su disposición unas cantidades ingentes de información. La integración de las tecnologías Internet y Data Warehouse tienen una serie de ventajas como son:

1. Consistencia: toda la organización accede al mismo conjunto de datos y ve los informes que reflejan sus necesidades. Hay una única versión de la verdad.
2. Accesibilidad: la empresa accede a la información a través de un camino común (el navegador de Internet), simplificando el proceso de búsqueda de la información.
3. Disponibilidad: la información es accesible en todo momento, independiente de los sistemas operacionales.
4. Bajos costos de desarrollo y mantenimiento, debidos a la estandarización de las aplicaciones de consultas basadas en Internet, independientemente del sistema operativo que soporte el navegador y de la reducción de los costos de distribución de software en los puestos clientes.
5. Protección de los datos, debido al uso de tecnologías consolidadas de protección en entornos de red (cortafuegos).
6. Bajos costos de formación, debido al uso de interfaces tipo WEB.

La interactividad de las aplicaciones en este entorno puede tener varios niveles:

1. Publicación de datos: las páginas distribuyen información obtenida del Data Warehouse, volcada en las páginas intra o Internet.
2. Distribución de reportes: dando soporte a consultas simples elaboradas por los usuarios.
3. Aplicaciones dinámicas: sirviendo de soporte de decisión a servicios solicitados desde el puesto cliente, ejecutando la petición en el servidor y devolviéndolas al cliente, vía el navegador de Internet o haciendo uso de "*applets*" de Java.

Las arquitecturas base de una implantación de Data Warehouse en Internet, pueden tener las siguientes alternativas:

- ✓ Usar el Servidor Internet como enrutador y ejecutar la petición desde el cliente al servidor directamente.
- ✓ Hacer uso del navegador para visualizar una página Internet, residente en el servidor de Internet. Esta página contendría información que se actualizará en el servidor

Internet desde el servidor del Almacén de Datos, a petición del usuario haciendo uso de CGI.

- ✓ El cliente podría lanzar su consulta directamente al servidor de DW, con "*applets*" de Java, haciendo el servidor Internet únicamente de enrutador.
- ✓ Realizar una descarga masiva de datos con un protocolo de transferencia de ficheros (FTP), para su proceso local.

El alcance funcional de la implantación del Almacén de Datos, basado en tecnologías Internet, puede ser la misma que la realizada sin su uso. En este sentido las críticas que se le pueden achacar en la actualidad, provienen de la baja velocidad de las líneas actuales, que se solventa parcialmente mediante el uso de aplicaciones Java, en lugar de uso de páginas HTML o CGI. Solución parcial, mientras la velocidad de transferencia se incrementa día a día mediante nuevos algoritmos de compresión de datos o el uso de líneas de alta capacidad RDSI.

Terminamos este capítulo resumiendo la teoría presentada y los beneficios que aporta un DWH.

Un DWH es una herramienta para la ayuda en el proceso de toma de decisiones. Una base de información estratégica, integrada y granulada como fuente de soporte a la decisión. Un proceso de integración de informaciones a partir de varios orígenes a fin de ofrecer una visión integrada de los negocios, combinado las tecnologías optimizadas para consultas de análisis interactivo. En virtud de esto el DWH presenta una fuente única de datos integrados, accesibles, consistentes flexibles y adaptados, la tarea de toma de decisiones en este ambiente es inmensamente más fácil que en el ambiente clásico operacional

Un DWH está compuesto por varios Data Marts, cada uno de ellos es un subconjunto de datos del DWH, orientados a un proceso o parte del negocio, basado en un tema determinado, no volátil y variable con relación al tiempo, sirve de apoyo a las decisiones gerenciales. Los datos fluyen del ambiente operacional para el nivel del DWH sufriendo generalmente una cantidad significativa de transformaciones.

Los DWH son proyectados con el propósito de suplir las necesidades de los directivos con mejores informaciones y con informaciones actualizadas sobre el trabajo de la empresa, aportando los siguientes beneficios:

- Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.
- Facilita la aplicación de técnicas estadísticas de análisis y modelación para encontrar relaciones ocultas entre los datos del almacén; obteniendo un valor añadido para el negocio de dicha información.
- Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.
- Simplifica dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.
- Supone una optimización tecnológica y económica en entornos de Centro de Información, estadística o de generación de informes con retornos de la inversión espectaculares.

Problemas que se pueden encontrar en el desarrollo del Data Warehouse.

Existen diversos problemas que pueden ocurrir durante el desarrollo de un sistema DWH. Dentro de estos problemas mencionaremos los siguientes: [23]

a) **No darle participación a la alta dirección en el diseño del DWH:** En el diseño de un DWH deben participar desde el inicio todos los futuros usuarios que participaran en la explotación del mismo, pues esto facilita el cumplimiento de los objetivos reales del DWH para el negocio de la empresa en el momento de la implantación.
b) **Generar falsas expectativas con promesas que no serán cumplidas:** Citar frases del tipo "El DWH mostrará a los directivos las mejores decisiones", puede causar desconfianza en diseño", el DWH no mostrará las mejores decisiones, sino respuestas a las consultas efectuadas. Cabe a los usuarios elaborar consultas inteligentes y analizar las respuestas obtenidas.

c) **Cargar en el DWH informaciones que solo están disponibles en los sistemas transacionales:** No todos los datos disponibles en los sistemas operacionales de la empresa son necesariamente útiles para el DWH. Cabe al diseñador de los datos de analizar junto con los usuarios que datos son los realmente contienen informaciones necesarias y cuales son los que su información no forman parte de los objetivos del DWH.

d) Imaginar el diseño del banco de datos del DWH es el mismo que el diseño de un sistema transacional: En un proceso transacional se debe dimensionar los recursos para que se obtenga una alta velocidad de acceso y grandes facilidades en la actualización de los registros. EN los sistemas de apoyo a la toma decisión la realidad es totalmente otra. El objetivo de estos sistemas es utilizar accesos agregados, es decir, sumas, tendencias, etc. Otra diferencia entre los dos tipos de sistemas puede ser detectada en los usuarios, en los sistemas transacionales un programador desarrolla una consulta que puede ser utilizada millares de veces, en el DWH el usuario final desarrolla sus propias consultas, que puede ser utilizada solo una vez.

e) **En la selección del personal, escoger un administrador para el DWH con orientación esencialmente técnica**: Los sistemas de apoyo a la decisión son en verdad una prestación de servicios y no un servicio de almacenamiento de datos. Por eso, es fundamental que el administrador del DWH sea una persona que represente los intereses de los usuarios y principalmente que sepa de los términos utilizados diariamente por la alta dirección y por otro personal encargado de tomar decisiones.

f) **Dedicar solo al tratamiento de los datos de tipo numérico y de texto:** Muchos pueden imaginar que las informaciones que serán utilizadas en un DWH son originales específicamente de las bases de datos transacionales y que estas informaciones pueden ser palabras o números. Quien piensa de esta forma puede estar engañado, pues textos, imágenes, sonidos y videos pueden ser bastante útiles en el momento de analizar distintas situaciones de la empresa y del negocio.

g) **Diseñar un sistema en base a un Hardware determinado, que no podrá funcionar con el crecimiento del DWH:** La capacidad de los servidores está en constante crecimiento, porque pueden ser adquiridos uno o más equipos para que trabajen por uno o dos años, más las características de estos sistemas indican que la cantidad de las

informaciones almacenadas puede alcanzar algunos TeraBytes. Es importante que el servidor del banco de datos del DWH sea adquirido a una empresa confiable y que garantice la posibilidad de actualizaciones (upgrade) en dependencia de la situación del mercado.

h) **Imaginar que después de la implementación del DWH no existirán problemas:** Mucho esfuerzo se realiza para que en un sistema DWH sirva de plantilla, porque después de su implantación los usuarios comenzarán a crear más consultas, y estas consultas necesitarán de nuevos datos que resultarán en nuevas consultas. De esta forma, el diseño de un sistema de apoyo a la decisión, precisa ser revisado y actualizado constantemente, no solo con nuevos datos, también con nuevas tecnologías.

Después de tener bien claros todos los conceptos anteriores, válidos para el diseño de un Almacén de Datos, pasamos a las fases de implementación de un Data Warehouse.

Capítulo 2. Diseño del Almacén de Datos

2.1- Fases de Implantación.

Tal y como aparecía en un artículo en ComputerWorld: *"Un Data Warehouse no se puede comprar, se tiene que construir"* y como sabemos, la construcción e implantanción de un Almacén de Datos es un proceso evolutivo.

Como todo proceso, el de implementar un Almacén de datos se apoya en una metodología, que bien puede no ser la mejor, pero con un control riguroso, aseguramos el seguimiento de la misma, el seguir los pasos de la metodología y comenzar el Data Warehouse por un área especifica de la empresa, nos permitirá obtener resultados tangibles a corto plazo.

La metodología por nosotros adoptada es, la propuesta por el SAS Institute, la "*Rapid Warehousing Methodology*" [24], esta es interactiva y está basada en el desarrollo incremental del proyecto de un Almacén de Datos y consta de cinco fases:

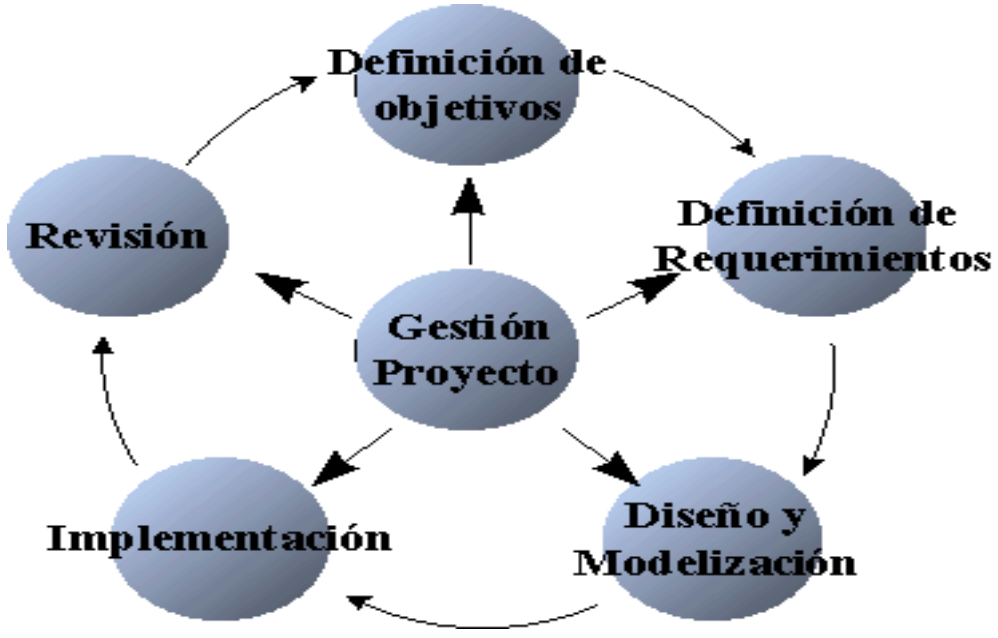

Figura 2.1. Metodología Rápida de Diseño de un DW según SAS Institute

- Definición de los Objetivos
- Definición de los requerimientos de la información
- Diseño y modelación
- Implementación
- Revisión

Definición de los Objetivos

Se definen los objetivos que deben cumplir la implementación del Almacén de Datos, estos deben ser precisos y desprovistos de ambigüedades, deben cumplir las expectativas de los usuarios de la información

Definición de los requerimientos de la información

Tal como sucede en todo tipo de proyectos, sobre todo si involucran técnicas novedosas como, las relativas a los Almacenes de Datos, aquí se debe analizar las necesidades y hacer comprender las ventajas que al sistema pueda aportar.

Es en este punto, donde se detallan los pasos a seguir, es donde el usuario juega un papel destacado.

Diseño y modelación

En esta fase, partiendo de los requerimientos de la información identificados en la anterior, se forman las bases para realizar el diseño y modelación del Data Warehouse, aquí se identifican las fuentes de los datos (sistemas operacionales, fuentes externas), también se identifican las transformaciones necesarias que deben sufrir las fuentes, para obtener el modelo lógico de los datos del Almacén de Datos dicho modelo está formado por las entidades y relaciones que permiten resolver las necesidades de información del negocio en estudio, por parte de la organización.

Este modelo lógico se traduce posteriormente en el modelo físico de datos del Data Warehouse y que define la arquitectura de almacenamiento, adaptándose al tipo de explotación que se realice del mismo.

La mayor parte las definiciones de los datos del Data Warehouse están almacenadas en los metadatos y por tanto forman parte del Almacén de Datos.

Implementación

La implementación de un Almacén de Datos lleva implícitos los pasos siguientes:

- Extracción de los datos del sistema operacional y transformación de los mismos.
- Carga de los datos validados en el Data Warehouse. Este proceso debe ser planificado con una periodicidad que se adaptará a las necesidades de refrescamiento detectadas durante las fases de diseño del nuevo sistema.
- Explotación del Almacén de Datos mediante diversas técnicas, dependiendo del tipo de aplicación que tengan los datos, como las siguientes:

    - ❖ Consultas e Informes
    - ❖ Procesamiento Analítico en Línea (OLAP)
    - ❖ Sistemas de Información de Gestión (EIS)
    - ❖ Sistemas para la Toma de Decisión (DSS)
    - ❖ Visualización de la información
    - ❖ Minería de Datos

Con la finalización de esta importante fase, ya estamos en condiciones de la Explotación del Data Warehouse.

Revisión

La construcción del Almacén de Datos, no finaliza con la implementación del mismo, esta es una tarea iterativa, donde se trata de aumentar su alcance, aprendiendo de las experiencias anteriores.

Luego de implantarse el Data Warehouse, se debe realizar una revisión del mismo, planteando preguntas que permitan, después de un tiempo prudencial, en dependencia de la frecuencia de su uso, definir los aspectos a mejorar o potenciar.

Diseño de cursos de adiestramiento

Es vital para una exitosa puesta en funcionamiento de los Sistemas de Data Warehousing, que los usuarios y administradores del sistema tengan conocimiento de las herramientas y su aplicación al negocio en particular, por lo que se debe diseñar cursos de adiestramientos a todos los posibles usuarios del DWH. Como se puede ver, esta fase no está incluida en la metodología propuesta, pero es muy importante para la instalación generalizada del Almacén de Datos.

2.2. Justificación del Proyecto

En las entrevistas realizadas y en el análisis del sistema "informativo" actual, vimos que la información mostrada sigue una estructura fija, ya que la los reportes fueron programados partiendo de un patrón definido en el Manual de Procedimientos de Emprestur S.A. [25], por este sistema, los directivos encargados de tomar decisiones o los especialistas que recomiendan soluciones apenas pueden consultar los datos en este formato, o sea con filas y columnas inmutables.

En un sistema DWH los usuarios pueden decidir en cualquier momento cuales son las columnas y cuales son la filas de para satisfacer su necesidad de información, pueden totalizar una u otra dimensión, todo esto sin tener que escribir una línea de código de programa.

La herramienta utilizada para realizar estos análisis, permite que con simple movimiento de arrastrar y soltar con el ratón, se cambie totalmente el panorama del reporte, analizando la información desde diferentes puntos de vistas, con considerable ahorro en esfuerzo y gasto de tiempo.

Con los sistemas de apoyo a la toma de decisión se puede realizar proyecciones con los datos, además de acceder a informaciones que fueron colectadas hace mucho tiempo, pues la base de datos transaccional solo almacena información de dos años como máximo.

2.3. Diseño del DWH

2.3.1 Definición de los Objetivos

Con el diseño del DW perseguimos los siguientes objetivos:

1. Reunir en un mismo depósito los datos necesarios para la alta gerencia de la empresa.
2. Construir una historia de la empresa, en cuanto a los datos de la misma.

2.3.2. Definición de los requerimientos de la información.

La información necesaria para la alta gerencia es la siguiente.

La dirección de la empresa hoy adolece de información que le permita una dirección acertada de los negocios, la información que brindan los sistemas existentes es muy abrumadora, es decir, un volumen muy grande tablas, donde la dirección de la empresa debe hacer unas búsquedas agotadoras si desea conocer un dato en específico que tenga relación con un área X de la empresa.

Como la construcción de un DWH, no es una tarea fácil y menos para un proyecto de este tipo y teniendo cuenta que un DWH esta formado por varios DMs, es que nos proponemos diseñar el modelo estrella del DWH e implementar un DM para un área seleccionada.

2.3.3. Construcción del Data Mart.

2.3.3.1 Selección del Proceso

En el contexto analizado anteriormente se ve la necesidad diseñar e implementar un Data Mart en el área de Comercial, debido a la poca flexibilidad y a la falta de autonomía de los usuarios que pueden utilizar la información almacenada por el Sistema Submayor.

Sin embargo, la implementación del prototipo debe ocurrir en esta área, sin la intención de sustituir el Sistema Submayor, sino la de ofrecer nuevas posibilidades de consultas de los datos, disponibles solo ahora para el área económica.

Partiendo del sistema SUBMAYOR y particularmente del fichero DEMASxx.DBF, que es donde se guarda las transacciones contables de cada Centro de Costo (UPS), debemos conformar una tabla de hechos donde se reflejen las medidas referidas al importe de las transacciones.

Con el diseño del DWH le presentamos a la Dirección de la Empresa los datos que ella necesita, de una forma amena y rápida.

2.3.3.2 Análisis de las Necesidades de los Usuarios.

Los directivos en sentido general, manejan las informaciones relativas a:

- ✓ Los Ingresos por cada centro de costo.
- ✓ Los Gastos por centro de costo.
- ✓ Las Utilidades por centro de costo

De las informaciones antes mencionadas, en el área de Comercial los datos necesarios son aquellos relativos a los ingresos, es decir, el monto de dinero que se ingresa tanto en Moneda Nacional como Divisas, por el concepto de ventas de producciones o servicios de Emprestur S.A.

Una vez definido el área donde se diseñará el DM, y teniendo en cuenta que esta información se necesita también al nivel de filial, total de la sucursal y en un tiempo determinado, ofrecemos a la dirección de la empresa una herramienta que le permite hacer análisis de ventas, análisis de los principales clientes y hacer pronósticos entre otros.

Para la obtención de estos datos debemos hacer un pase desde el fichero DEMASxx.DBF, este fichero como hemos detallado anteriormente, es el que guarda las transacciones, podemos realizar este pase de datos con la siguiente frecuencia:

1. Cada 10 días que es cuando los centro de costos hacen los cierres decenales de su actividad.
2. Después del cierre mensual.
3. Diario.

Con la frecuencia del pase diario presentamos un problema de disciplina en cuanto a la captación de los datos en el sistema submayor, ya que no es frecuente que se capten los datos diarios, aunque debemos reconocer que esta frecuencia es muy buena para mantener una información actualizada.

El pase mensual de la información, es una idea correcta desde la teoría de los DW, ya que se hace una sola carga de las transacciones, pero dejamos a Dirección de la Empresa con información desactualizada de sus producciones y servicios.

El pase cada 10 días es la media correcta, para la carga de información, por las siguientes razones:

- ✓ Los cierres decenales están normados, por la dirección de contabilidad de la empresa, por lo que debe cumplirse en la mayoría de los casos.
- ✓ Puede accionarse después de un análisis de los datos, con vista a mejorar cualquiera de los indicadores antes expuestos.
- ✓ La carga se hace pequeña, la cantidad de datos a transformar es mucho menor, que con la frecuencia mensual.

En el fichero DEMASxx.DBF, existe información banal para la dirección, después de un análisis conjunto con el área de contabilidad y la dirección de la empresa, llegamos a la conclusión de eliminar campos en dicha tabla, para conseguir mayor claridad en los datos almacenados.

Analizando cada uno de los sistemas descritos en acápite 1.1.1 y la información almacenada, se llegó a la conclusión de que un Data Marts(DM) en el área de Comercial sería interesante para la fase inicial del proyecto para toda la empresa, ya que esta área no tiene hoy información para recomendar líneas de comercialización de los servicios ofrecidos por Emprestur S.A., además, todos los datos necesitados por los especialistas se encuentran en las tablas vistas anteriormente. Existe una cultura de búsqueda de información, pero esta es agotadora, debido a la gran cantidad de tablas a analizar para poder llegar a conclusiones.

Los pasos a seguir para la transformación de los datos son los siguientes:

- Carga decenal de todos los datos del fichero DEMASxx.DBF.
- Carga de los nuevos clientes y nuevos centros de costos.
- Carga de las nuevas cuentas contables definidas por el área de contabilidad.
- Actualización de la tabla fecha, con la fecha de esta carga.

Después que tenemos los datos cargados entonces pasamos a la etapa de limpieza de los mismos, esta etapa consiste en la eliminación de los campos no necesarios.
En el fichero DEMASxx.DBF existen datos que se pueden transformar en virtud de las necesidades de información, estos son los siguientes:

Los campos BALMN y BALDIV contienen el valor del importe de cada transacción en Moneda Nacional y Divisa respectivamente, este valor puede transformarse en el campo importe, y que lo diferencie la cuenta contable que tiene asignada la transacción, los campos PAGMN y PAGDIV no son interesantes, ya que ellos solo reflejan el valor del

pago, SALDOMN y SALDODIV no brindan información necesaria según los objetivos definidos, ya que ellos guardan el saldo en Moneda Nacional y Divisa a pagar, como podemos apreciar estas son las medidas que contiene el fichero DEMASxx.DBF, conjuntamente con los usuarios, definimos que el campo importe de la nueva tabla es la medida que más necesitan ellos.

2.4. Diseño y Modelación

Luego de seleccionar el área donde se aplicará el DM, vistos los requerimientos del DWH y del diseño del sistema legado, coincidimos con la dirección de la empresa, que para el diseño del modelo estrella necesitamos los siguientes datos:

- ✓ Clasificador de Cuentas.
- ✓ Clasificador de Clientes.
- ✓ Clasificador de Centro de Costos.
- ✓ Clasificador de Filial.
- ✓ Fechas.
- ✓ Clasificador de Sucursal.

Estos datos son necesarios en el diseño general del DWH, coincidiendo que en diseño del DM del área comercial de la filial los datos necesarios pueden ser:

- Clasificador de Cuentas.
- Clasificador de Clientes.
- Clasificador de Centro de Costos.
- Fechas.

Modelo Estrella del Datawarehouse de Emprestur S.A. Sucursal Cienfuegos

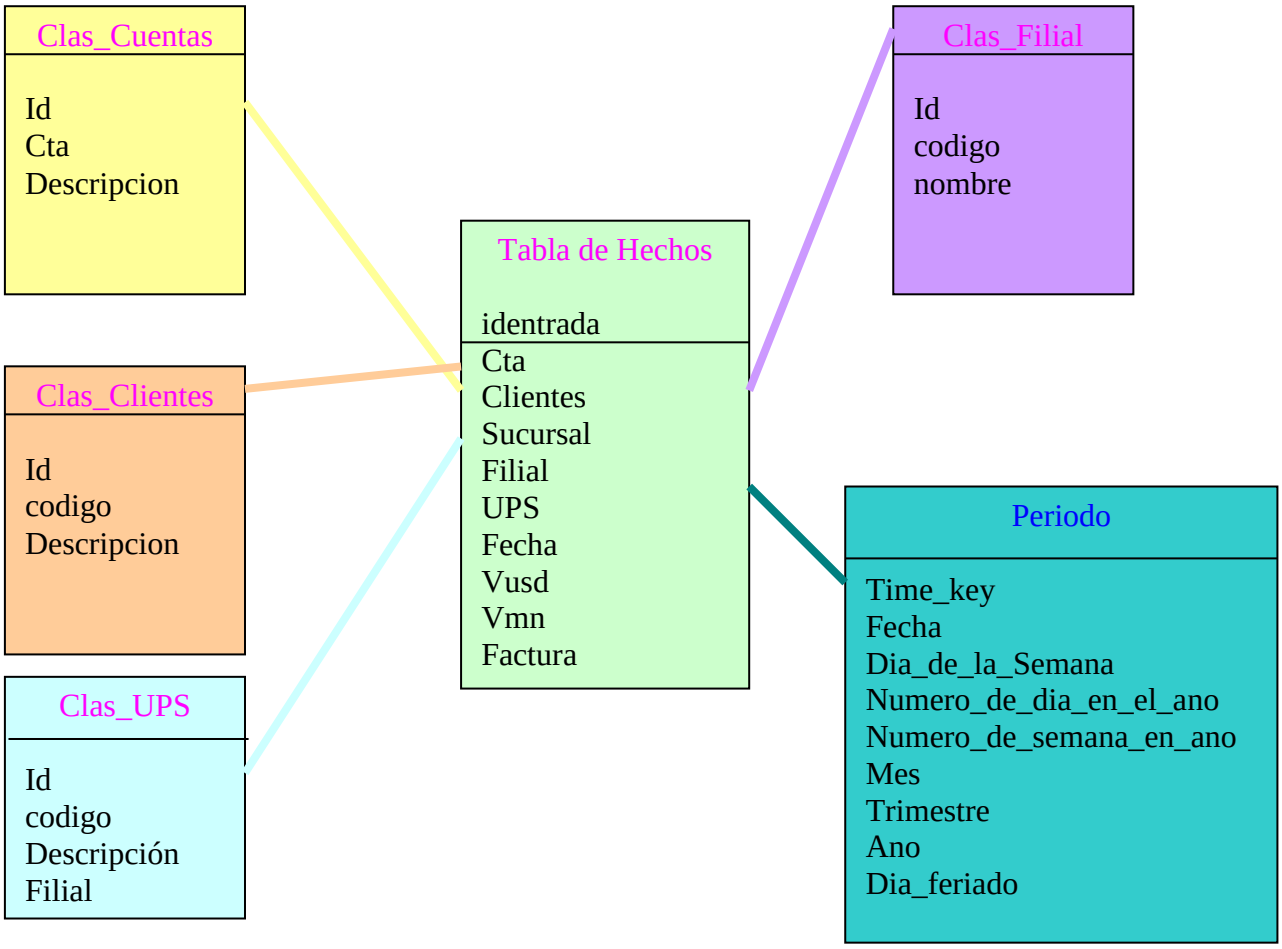

Gráfico 2.1. Modelo estrella del DWH

A continuación se explica con más detalle cada dimensión seleccionada.

2.4.1. Definición de las Dimensiones del Modelo Estrella.

Cuando se proyectan modelos estrellas departamentales se debe tener presente los proyectos futuros dentro de varios sectores de la empresa, pues no deben existir datos redundantes y que no se correspondan los unos a los otros, una vez que tratan del mismo tipo de información.

Un ejemplo de dimensión que se debe tener mucho cuidado de formular es la de Cliente, cuando se haga una carga de la tabla de clientes de la base transacional el responsable por la carga debe estar atento para que otras bases de datos de otros sistemas de apoyo a la decisión utilicen esta misma actualización.

En el modelo estrella propuesto, las tablas de dimensiones y sus atributos, están relacionadas con la tabla de hechos de forma uno a muchos.

Dimensión Clas_Cta

Esta dimensión es el clasificador de cuentas de la organización y está formada por los siguientes atributos,

Nombre del Atributo	Significado
id	Es un consecutivo de cada cuenta, pudieramos utilizar el código de la cta como id, pero es más fácil la manipulación por la explicación del campo siguiente.
Cta	Una cuenta contable, está formada por cuenta, subcuenta, elemento, entre otros parametros, el número de cuenta agrupa a un conjunto de subcuentas, la combinación Cta + scta es única.
Scta	Forma parte de la cuenta, como explicamos anteriormente.
Descripcion	Nombre de la cuenta, explica el significado de sus operaciones.

Dimensión Clas_Clientes

Esta dimensión es el clasificador de clientes que tienen relación con Emprestur S.A. Sucursal Cienfuegos.

Nombre del Atributo	Significado
Id	Id del Codificador de Clientes, utilizamos el id, debido a que el código del cliente tiene 9 caracteres.
Codigo	Código del Cliente según el Manual de Procedimientos de la Casa Matriz
Descripcion	Nombre del Cliente (solo empresas)

Dimensión Clas_UPS

Nombre del Atributo	Significado
Id	Id del los Centros de Costos
Codigo	Código del centro de costo u unidad de producción y servicios
Descripcion	Nombre del centro de costo u unidad de producción y servicios
Filial	Código de la Filial al que pertenece el Centro de Costo

Dimensión Clas_Filial

Nombre del Atributo	Significado
Id	Id de cada Filial
Codigo	Código de las Filiales, tiene 2 caracteres
Descripcion	Nombre de la Filial

Dimensión Periodo

Nombre del Atributo	Significado
Time_key	Id de las Fechas
Fecha	Fecha en que ocurre la operación
Dia_de_la_semana	Dia de la semana de la fecha
Numero_de_dia_en_el_mes	Número del dia de la Fecha en el año
Numero_de_semana_en_ano	Numero de la semana de la Fecha en el año
Mes	Mes de la fecha
Trimestre	Trimestre de la fecha
Ano	Año de la Fecha

En la tabla anterior se puede apreciar la forma en se estructuró la dimensión tiempo, de esta manera existe la posibilidad de realizar operaciones "*drill down*", es decir, detallar los datos hasta que sean mostrados los días, también se puede realizarla operación inversa "*roll up*", estas operaciones se pueden realizar gracias a la herramienta OLAP del SQL server 2000, el **Analysis Service**, utilizando las bondades de esta herramienta la dimensión periodo la dividimos en Año, Trimestre, Mes, pero no se excluye la posibilidad de realizar análisis más detallado, como en que semana del año, en que semana del mes, depende de las facilidades que brinde la herramienta OLAP a utilizar.

2.4.2. Tabla de Hechos del DWH y que se utiliza en el prototipo.

Nombre del Atributo	Significado
identrada	Id de los articulos de la tabla de Hechos
idCta	Id de la Cuenta según su Clasificador
idClientes	Id de los Clientes según su Clasificador
idFilial	Id de las Filiales según su Clasificador
Idsucursal	Id de las Sucursales según su Clasificador
idUPS	Código de los Centros de Costos según su Clasificador
idFecha	Id de la Fecha en que se realiza la operación según su Clasificador
VUSD	Importe de la Operación en USD (Dólares Estadounidenses)
VMN	Importe de la Operación en Moneda Nacional
Factura	Número de la Factura

Como se puede apreciar en el diseño, existen los id en cada una de las tablas a parte de los códigos naturales de cada articulo, esto se debe a dos razones:

- Evitamos la ambigüedad de los códigos predefinidos por el Manual de Procedimientos de la Casa Matriz.
- Hacemos más rápido el trabajo de la aplicación asociada.

En cuanto al Id de la tabla de hechos se debe, a que se controla de forma automatica la cantidad de transacciones efectuadas hasta la fecha, esto pudiera traer un problema en cuanto al tamaño de la tabla, pero debido a la cantidad fisica de espacio disponible en el servidor de la empresa y la disponibilidad en el mercado de soportes de alta capacidad, no es preocupante para nosotros, además como la clave es sencilla la Aplicación asociada mejora su rendimiento, en los Store Procedure se garantiza que esta llave cumpla con la unicidad de la multiple.

2.4.3.Modelo estrella para el prototipo

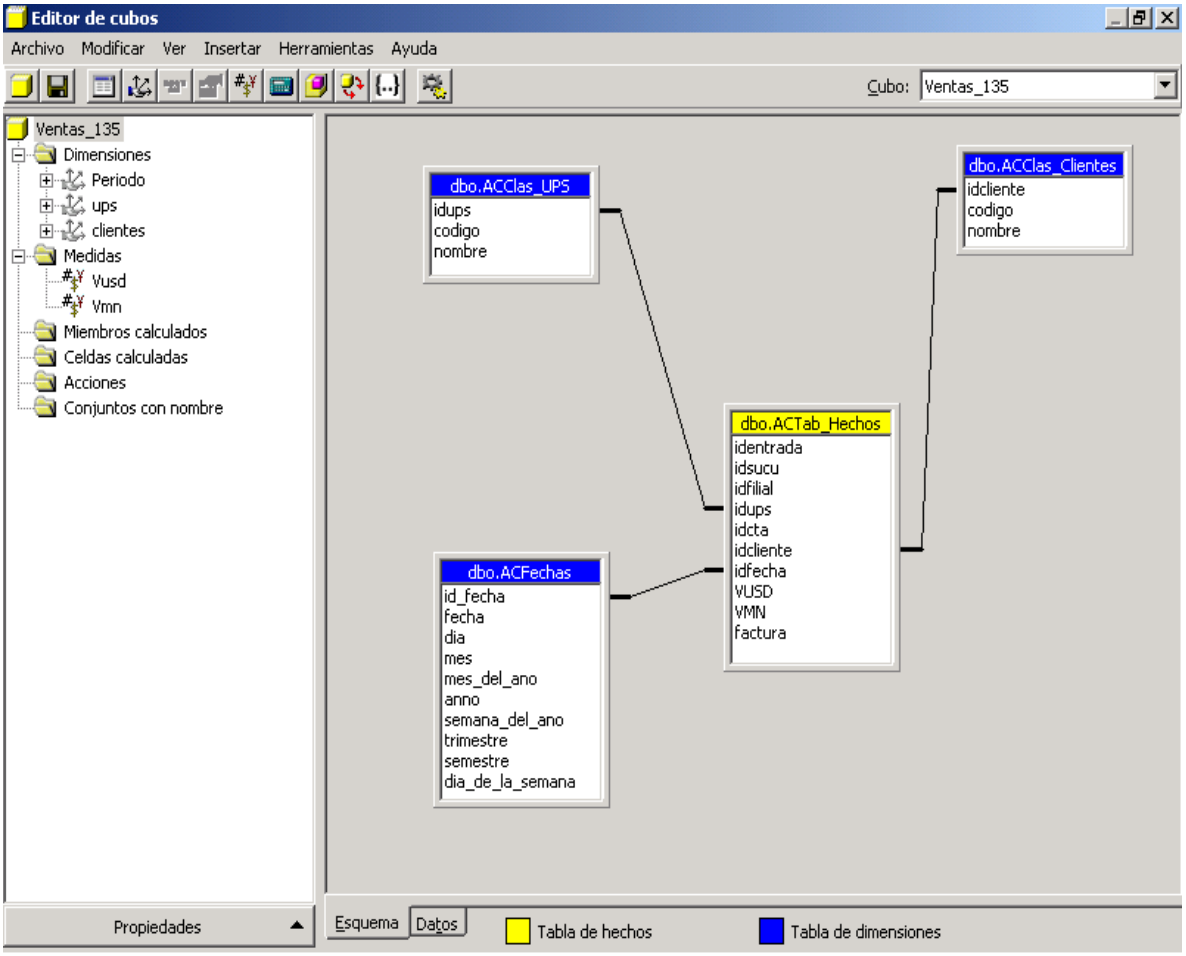

Figura 2.2. Modelo Estrella para el DM, según el Analysis Service

Si se compara la figura 2.2. con el modelo estrella del DWH, se puede apreciar que no se formará un conjunto de DMs, sino que con la información contenida en el DWH Global de la empresa, se pueden formar distintos DMs.

2.5 Nivel de agregación de las dimensiones.

Para el prototipo, se define el nivel de agregación siguiente para las dimensiones declaradas.

Agregaciones del modelo estrella del prototipo.

Dimensión	**Campo**
Clientes	**Nombre del Cliente**
UPS	**Nombre de la Unidad de Producción y Servicio (centro de costo)**
Periodo	**Año (con 4 dígitos)** **Trismestre del año (1 a 4)** **Mes del Año (1 a 12)** **Nombrados por sus nombres y no por sus números.**

El nivel de agregación de las dimensiones anteriores está definida de la siguiente forma: conocer el total de la venta por cada UPS en un periodo determinado y con clientes predefinidos, se puede obtener la suma general por cada dimensión, como explicamos anteriormente en la dimensión Periodo podemos hacer *drill down* y *roll up,* con cada una de las restantes dimensiones.

2.6. Tamaño de la Base de Datos.

Teniendo en cuenta que la base de datos a diseñar, debe ser capaz de guardar información de 3 años, que la cantidad de transacciones que hacen por día en la sucursal son 65 como promedio, detallaremos a continuación el cálculo del tamaño posible de la base de datos, de manera similar a como lo brinda Kimball en los ejemplos de su libro [26].

Dimensión tiempo: 3 años x 354 días= 1062 días.
Dimensión Clas_Cuentas: 20 cuentas
Dimensión Clas_Filial: 3
Dimensión Clas_UPS: 20
Dimensión Clas_Clientes: 40
Cantidad de Artículos de la Tabla de Hechos = 3 x 354 x 20 x 20 x 40=5 371 200 artículos.
Cantidad de campos llaves en la tabla de hechos = 6
Cantidad de campos hechos de la tabla de hechos = 2
Total de campos de la Tabla de Hechos= 10
Tamaño de la tabla de hechos = 5 371 200 articulos x 10 campos x 4 bytes = 2148480000 bytes.
Aproximadamente 2.09 GB.

2.6.1. Tamaño del Prototipo:

Dimensión periodo: 3 años x 354 días = 1062 días
Dimensión Cliente: 40
Dimensión UPS: 20
Cantidad de articulos de la Tabla de Hechos: 3 x 354 x 40 x 20 = 849 600 artículos
Cantidad de campos llaves en la tabla de hechos = 6
Cantidad de campos hechos de la tabla de hechos = 2
Total de campos de la Tabla de Hechos= 10(*)
Tamaño de la tabla de hechos = 849 600 artículos x 10 campos x 4 bytes = 33 984 000 bytes.
Aproximadamente 33.1 MB

2.7. Definiendo las medidas numéricas del DMs.

Las medidas numéricas son definidas en función de los valores que se desea mostrar en la herramienta OLAP seleccionada, formando normalmente la parte interna de la tabla que es divida en filas y columnas. Se puede definir como los valores que componen los gráficos

analíticos, siendo las dimensiones las que forman los ejes de estos gráficos. En el modelo en cuestión se definen dos medidas, que son las siguientes.

a) **Vusd** es el importe en USD de cada producción o servicio ejecutado por cada UPS a un cliente en una fecha determinada.
b) **Vmn** es el importe en Moneda Nacional de cada producción o servicio ejecutado por cada UPS a un cliente en una fecha determinada.

Estos valores pueden ser visualizados con la herramienta de consulta Analysis Service del SQL 2000.

2.8 Definiendo los Miembros Calculados

Los miembros calculados poseen carateristicas semejantes a la de las medidas, una vez que son derivados de estas medidas utilizando la herramienta del *Editor de Cubos* como sigue:
Para el prototipo definimos como Miembro calculado a la suma de Vusd y Vmn, este valor es necesario para tener una visión general de la producción de la empresa, los balances de contabilidad y la confección de los planes de producción contemplan este valor.

Vtotal = Vusd + Vmn.

Esta operación es efectuada por la propia herramienta OLAP.

2.9. Definiendo los atributos de las dimensiones.

En las tablas anteriores se define los atributos de cada dimensión, estos atributos se tomaron tal y como lo almacena el sistema submayor, debido a que hoy no existe información detallada de cada uno de ellas, solo en el caso de la dimensión periodo es donde se definen más atributos.

2.10. Definiendo la edad de los datos.

La edad de los datos se refiere al tiempo en que serán mantenidos en el DWH. Existen sistemas analíticos que guardan informaciones de décadas, no es necesario guardar todos los datos de distintos años.

Una práctica bastante usada es la de guardar los últimos 5 años en la granularidad original de los datos, tal como fueron cargados originalmente y posteriormente pasarlos para datos más resumidos, con una granularidad menor.

En caso del prototipo desarrollado existen datos almacenados a partir del segundo trimestre del año 2000.

2.11. Implementación

- Extracción de los datos del sistema operacional y transformación de los mismos.

Para la extracción de los datos del sistema operacional seguimos los siguientes pasos:

1. Se crearon en SQL Server 2000 dos bases de datos, una llamada el AreaDeCarga y la otra ESSA, la primera como su nombre indica, es la encargada de recibir todos los datos del sistema operacional y la segunda es el Almacén de Datos.
2. Crear un fichero .bat, que realice la copia desde el sistema operacional, situado en una máquina de la red, hacia una carpeta del servidor, con este paso logramos la actualización diaria de la carga del Almacén de Datos.
3. Ejecutar un fichero creado en Visual Basic 6.0 con un fichero .MDB donde se especifica la ubicación de los ficheros que se van a convertir a SQL, este paso es necesario ya que los ficheros DBF´s están protegidos, por lo que este programa se encarga de la desprotección de los mismos.
4. Ya desde SQL, se creó un procedimiento almacenado el cual se encarga de borrar los datos de la última carga, para que no exista duplicación en los datos de la carga.

5. Un paquete del Servicio DTS del SQL Server 2000, ejecuta la conversión de los ficheros .dbf a tablas de SQL en la base AreaDeCarga.
6. Se cambia en la tabla DEMAS25 de la base de datos AreaDeCarga, los códigos de la UPS, ya que pueden existir, por deficiencias del sistema heredado, datos que no contenga el clasificador de las UPS, este error se produjo cuando se cambiaron por orientación del nivel superior los códigos de las UPS, por lo que se cambió el clasificador, pero no el contenido de las transacciones efectuadas anterior a este cambio. El procedimiento es temporal y no afecta la carga del DW, cuando los códigos de la empresa sean totalmente confiables se puede excluir de la carga.

- Carga de los datos validados en el Data Warehouse.
    1. Desde SQL varios procedimientos almacenados se encargan de la carga y validación de los datos hacia el Data Warehouse. Cada uno de estos procedimientos almacenados, validan la existencia de los códigos correspondientes, en caso de no existir alguno de ellos, se reporta el artículo en cuestión en una tabla llamada AcErrores de la base de datos ESSA, esta tabla garantiza la calidad de los datos.

A continuación mencionamos los procedimientos creados para la carga del DWH.

Nombre	**Función**
CARGA_DE_LAS_CUENTAS	Carga el clasificador de cuentas
CARGA_DE_LAS_UPS	Carga el clasificador de UPS
CARGA_DE_LOS_CLIENTES	Carga el clasificador de clientes
CARGA_LA_FECHA	Carga el clasificador de las fechas
CARGA_DEL_DEMAS	Carga la tabla de hechos partiendo del fichero DEMASxx.DBF, valida que todos los datos sean correctos.

El código de cada uno de estos procedimientos almacenados puede consultarlos en el Anexo No.1

Para la extracción y carga de los datos, se utilizó el DTS del SQL Server 2000, donde se ejecutan en orden lógico todos los pasos descritos anteriormente, como se puede apreciar en la siguiente figura.

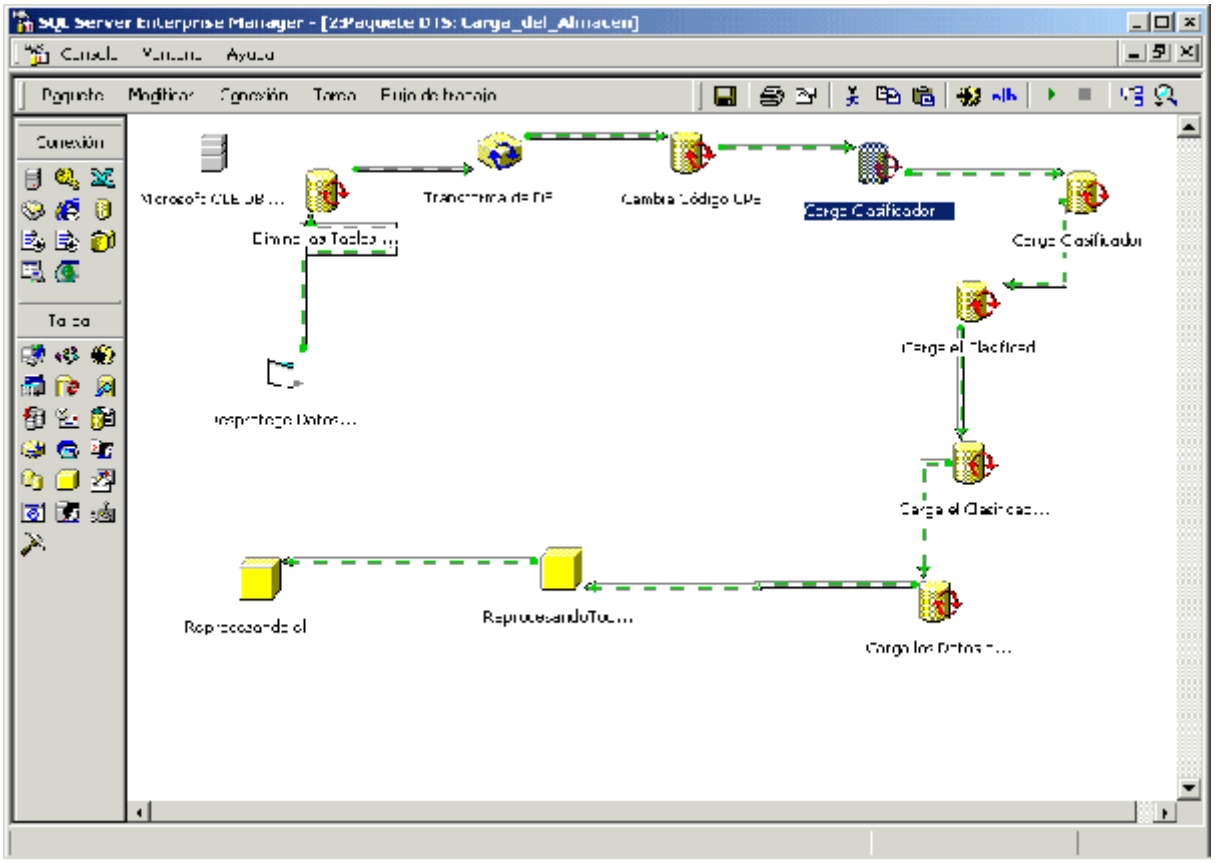

Figura 2.3. Paquete creado con el DTS del SQL 2000

Con los pasos anteriores, comenzamos el uso del Data Warehouse de Emprestur Sucursal Cienfuegos, con la explotación del DM del área Comercial, hoy solo tenemos los datos de una Filial (Cienfuegos) por encontrarse el Departamento Económico en el mismo edificio de la Gerencia de la Sucursal, lo que hace que la carga sea fácil.

Este DM puede ser implementado en todas las filiales de la SUCURSAL CIENFUEGOS, con solo instalar el SQL Server 2000 y los procedimientos descritos anteriormente.

2.12. Revisión

La construcción del Almacén de Datos, no finaliza con la implementación del mismo, esta es una tarea iterativa, donde se trata de aumentar su alcance, aprendiendo de las experiencias anteriores.

Luego de implantarse el Data Warehouse, se debe realizar una revisión del mismo, planteando preguntas que permitan, después de un tiempo prudencial, en dependencia de la frecuencia de su uso, definir los aspectos a mejorar o potenciar.

Algunas de las preguntas que se pueden plantear, para la validación del DWH.

1. Los Datos que se muestran ¿se corresponden con la realidad?

   En caso de que los datos mostrados no se correspondan con la realidad, se debe conjuntamente con los especialistas y la dirección, revisar cuales son datos a mostrar.
2. La Frecuencia de Carga de los Datos, ¿es la correcta?

   Los Datos que muestran información no actualizada, se debe revisar la frecuencia de carga de los datos, puede que este proceso, lleve al Almacén de Datos, información atrasada, con la cual es imposible tomar alguna decisión.
3. ¿Por qué no se explota el DWH?

   Se debe analizar en conjunto con el consejo de dirección de la empresa, las causas que motivan la no explotación del DWH, una de ellas, puede ser, la falta de capacitación al personal que debe explotarlo, para la cual se debe diseñar un plan de capacitación intensivo.

Al comenzar el análisis y diseño del DW, se detectaron algunos problemas que vale la pena mencionarlos, con el fin de evitar estos, en posteriores diseños.

- No existe un Clasificador único de clientes, esto trae consigo que se procese un mismo cliente como dos o tres cuando se consolida en el ámbito de SUCURSAL CIENFUEGOS. Por ejemplo, La Filial de Trinidad y la Filial de Cienfuegos ha desarrollado trabajos en la Base de Campismo Guajimico, contablemente cada filial registra a este cliente con un código distinto, por tanto en el consolidado al nivel de sucursal se analiza como si fueran dos clientes distintos.
- Algo parecido ocurre con el Clasificador de Centros de Costos.

Al cabo de un mes de implantado y en consulta constante con los especialistas del área de contabilidad y los jefes de centros de costos, nos dimos cuenta que los valores tomados distorsionaban la información, es decir considerábamos los importes de la cuenta **135** y no los valores de la cuenta contrapartida **905**, que es en definitiva la cuenta que controla los ingresos por cada centro de costo.

Se tiene planificado que a los seis meses de explotación del DM, se realice una nueva revisión, aunque no se descarta la revisión puntual a pedido de los usuarios.

Capítulo 3. Implementación del Almacén de Datos.

Este capitulo detalla todo lo relacionado con la implementación del Data Warehouse, se explica la forma de explotación, las herramientas utilizadas así como el plan de capacitación para la explotación del Almacén de Datos.

La implementación del DM fue desarrollada en una plataforma IBM-PC, con sistemas operativos MSDOS, Windows 2000 Server y Windows 2000 Professional, según la figura 3.1 y se explicó en el capítulo 2, se desarrolló una herramienta para la desprotección de los datos en los ficheros DBF, que es donde el Sistema Submayor almacena los datos primarios, utilizando las opciones que brinda SQL 2000 Server, en este caso el DTS (Servicio de Transformación de Datos) se diseñó un paquete en el DTS como se puede apreciar en la figura 2.2.

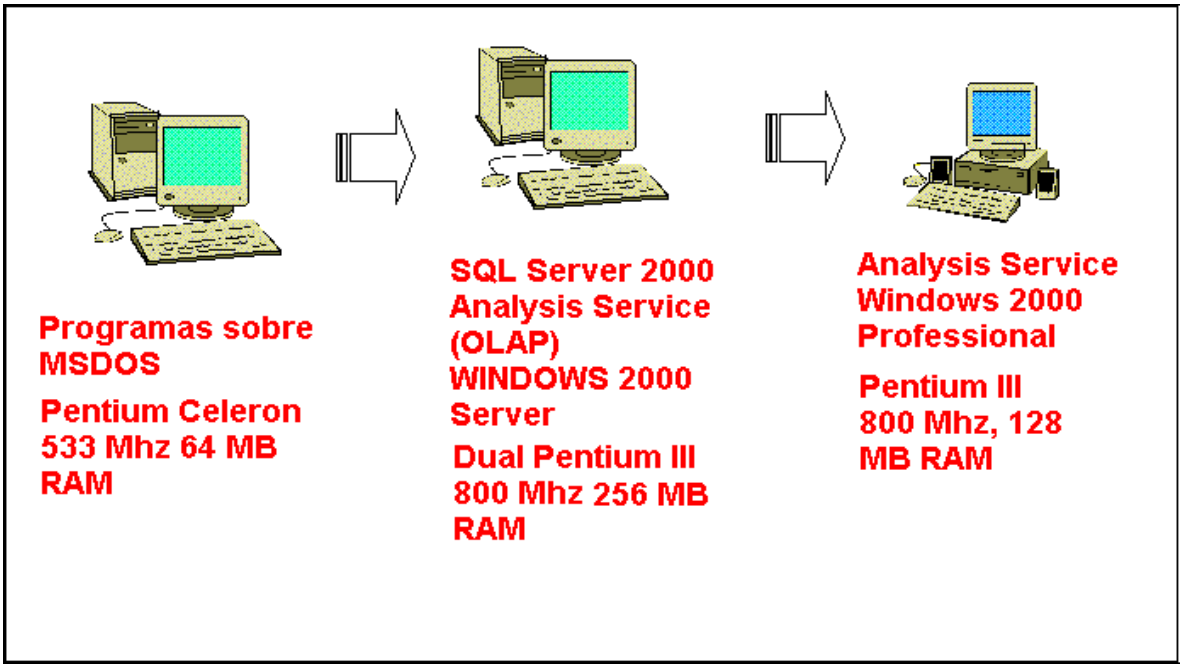

Figura 3.1. Ambiente en que se desarrolla el DM

Explotación del Almacén de Datos mediante diversas técnicas, dependiendo del tipo de aplicación que tengan los datos.

La aplicación que tienen los datos está orientada, a la consulta por parte del director de la empresa y al área Comercial, a los datos relativos a la venta, para poder dar un seguimiento a las fluctuaciones de la producción en las distintas UPS (Unidades de Producción y Servicio) y al comportamiento de los Clientes dentro del radio de acción de la Empresa.

El DW se actualiza de forma automática, por lo que los usuarios siempre tendrán a disponible la información más actual, esto depende de la captación de los datos del área de contabilidad

3.1. Herramientas utilizadas.

3.1.1 Herramientas *Back End*

Estas herramientas son las responsables de los pasos para la carga de un DWH, es decir, la extracción, limpieza, carga y restaura de los datos utilizados en el DWH.

Para el DMs definido, solo se hizo un tipo de carga de datos, está fue de datos internos a través de las herramientas creadas en el cuarto trasero, basadas en los procedimientos mencionados en el capítulo 2, como Sistema de Gestión de Base de Datos, que sirve de soporte al DM, se utilizó el SQL SERVER 2000, la extracción de los datos se efectuó utilizando el DTS del SQL SERVER 2000, estas herramientas son las encargadas de cargar los datos desde los sistemas operacionales hacia el DWH. Todas estas operaciones se detallan como se ha dicho en el capítulo 2.

3.1.2 Herramientas *Front End.*

En el epígrafe anterior se especifica el uso de las herramientas *back end* utilizadas, se muestra en el capítulo 2 las funciones de cada *store procedure*. Después la fase de carga de los datos en el DWH es posible realizar diversas consultas que finalmente podrán ser mostradas utilizando una herramienta del tipo *Front End* (Herramientas del Cliente), la herramienta utilizada, como se ha mencionado es el Analysis Service del SQL 2000, la cual mostrará la potencialidad de los sistemas OLAP.[27]

A partir de ahora podemos distinguir con claridad los tipos de herramientas utilizadas en los sistemas DWH: las herramientas *back end* y las herramientas *front end*. Estas diferencias se tornan evidentes cuando analizamos sus funciones dentro del sistema como un todo, que son de, una visión bastante superficial, cargar los datos en el DWH y consultar los datos cargados en el DWH.

El proceso de consulta de los datos debe ser optimizado de forma tal que permita a los usuarios finales del sistema, es decir, a los directivos y otro personal encargado de tomar decisiones, transformar sus preguntas en consultas adecuadas a la base de datos, una vez que no posean y no tengan la necesidad de poseer conocimientos profundos sobre la teoría de las Bases de Datos.

Estas herramientas poseen una interfaz que ayuda a los análisis comparativos y presenta los resultados más claramente, tanto para los usuarios especializados como para otros colaboradores de la empresa que estén vinculados al asunto en cuestión.

Además de permitir la consulta de los datos, las herramientas *front end* deben poder realizar operaciones sobre los datos consultados, transformando los datos en informaciones verdaderamente útiles a los usuarios. Estas herramientas constituyen las llamadas "aplicaciones sobre el negocio", las cuales permiten que un usuario pueda realizar consultas también especializadas en una BD. De esta manera, las cuestiones que surjan en reuniones o hasta en el mismo trabajo diario, pueden obtener respuestas rápidamente y con índice de certidumbre muy elevado.

A continuación mostramos algunas de las operaciones realizadas por las "aplicaciones sobre el negocio".

a) Selección del conjunto de datos necesario para la consulta la DWH.
b) Presentación de los resultados de las consultas, a través de gráficos, tablas, informes, indicación de puntos críticos, etc.

c) Publicación de los resultados a otros usuarios, bien sea a través de impresos, correo electrónico o en la misma intranet empresarial o corporativa en caso de existir, y más aún, si se desea por la Internet.

Para la explotación del Almacén de Datos utilizamos el Servicio de Análisis del SQL Server 2000 (Análisis Service), este servicio se realiza mediante un ambiente de consulta completo y flexible, el cual permite que se realicen diversas operaciones del universo de las herramientas analíticas. En la figura 3.2 está representada la herramienta OLAP con los datos del prototipo.

En el Análisis Manager, creamos un cubo llamado Ventas_135, con las dimensiones UPS, CLIENTES y PERIODO, con las medidas VUSD, VMN y la medida calculada Vtotal, con la dimensión PERIODO (Año, Trimestre y Mes). A continuación se describen cada una de estas operaciones que permiten al usuario final navegar por los datos del cubo.

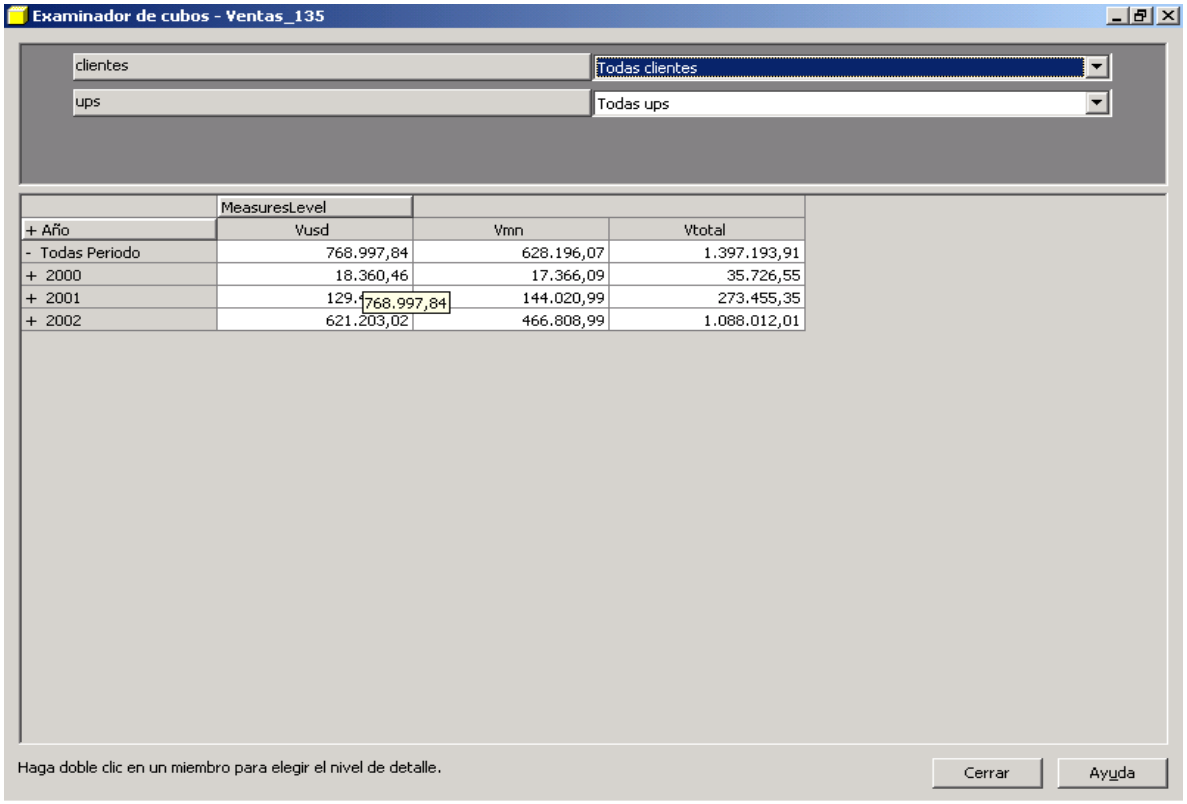

Figura 3.2 Interfaz del Analysis Service

a. *Drill down:* (Detallar los datos) Esta operación se realiza cuando el usuario haga doble click sobre un miembro para elegir el nivel de detalle, en caso del periodo se presentará el trimestre del año seleccionado.
b. *Roll up:* Es la operación inversa al *drill down*, cuando el usuario haga click en una de las dimensiones expandidas, estos datos serán agregados de manera tal, que todos los periodos serán totalizados en el año.
c. *Pivoting:* (alternar filas y columnas). Si el usuario arrastra la dimensión UPS para dentro de la columna Periodo que compone las filas de la tabla, todos los valores por UPS serán recalculados dentro de cada intervalo de la dimensión Periodo.
d. *Slice dice:* Fijar partes de los datos. En la parte superior de la interfaz, se puede seleccionar un Cliente específico y de esta forma se presentarán los datos relativos a las ventas efectuadas a ese cliente seleccionado.

En la parte superior de la interfaz están disponibles las dimensiones del cubo al usuario, para que sean arrastradas a la porción inferior, donde los datos son representados de acuerdo a las características seleccionadas.

Se puede notar que las dimensiones están configuradas para listas *todas*, o sea, todas las tuplas de las medidas, en la parte de fondo blanco del formulario.

En este ejemplo están totalizadas las medidas Vusd, Vmn y Vtotal, separadas por UPS y todos los clientes en un periodo determinado.

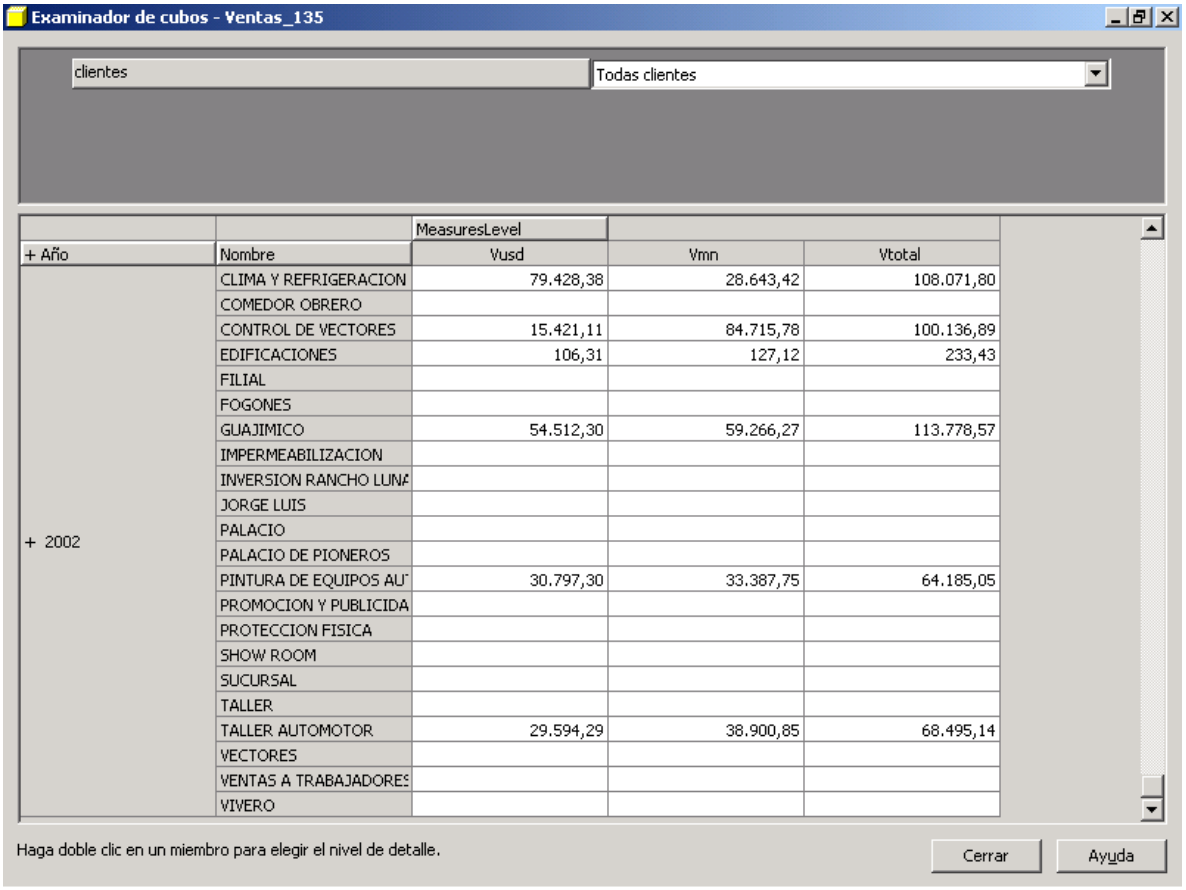

Figura 3.3. Otra visión del Análisis Service

Con la técnica de arrastrar y soltar, es fácil para el usuario hacer consultas de referencias cruzadas, con cada una de las dimensiones.

El análisis de esta forma se mantiene lineal, aunque se deja ver su utilidad comparativa, a continuación mostramos consulta de referencias cruzadas, ejemplo, cuanto vendió cada UPS en un periodo determinado y su comparación con igual periodo del año anterior, con todos los clientes ejecutados.

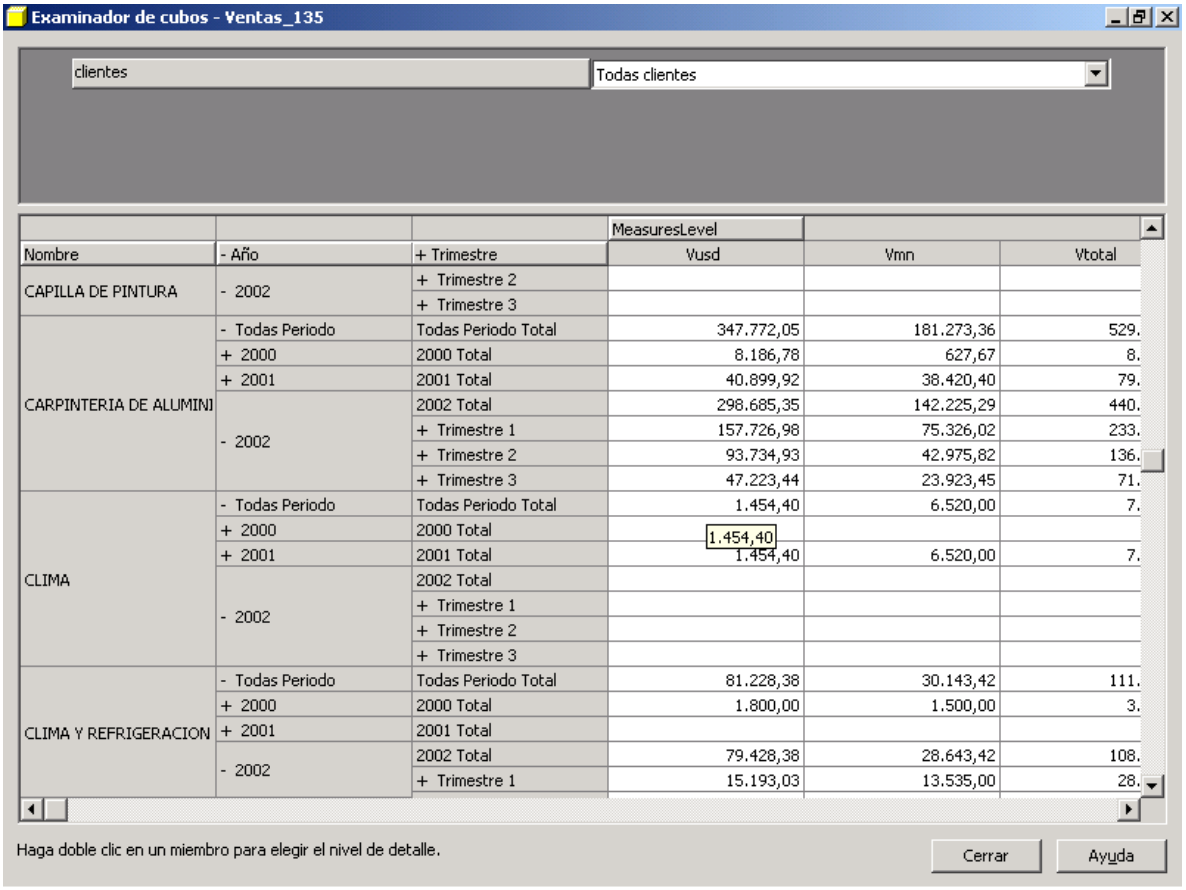

Figura 3.4. Referencia cruzada

Si se desea, se puede seleccionar un cliente en específico y analizar como se han comportando las ventas, esto permite tener seleccionado los clientes que más compren los servicios de la empresa.

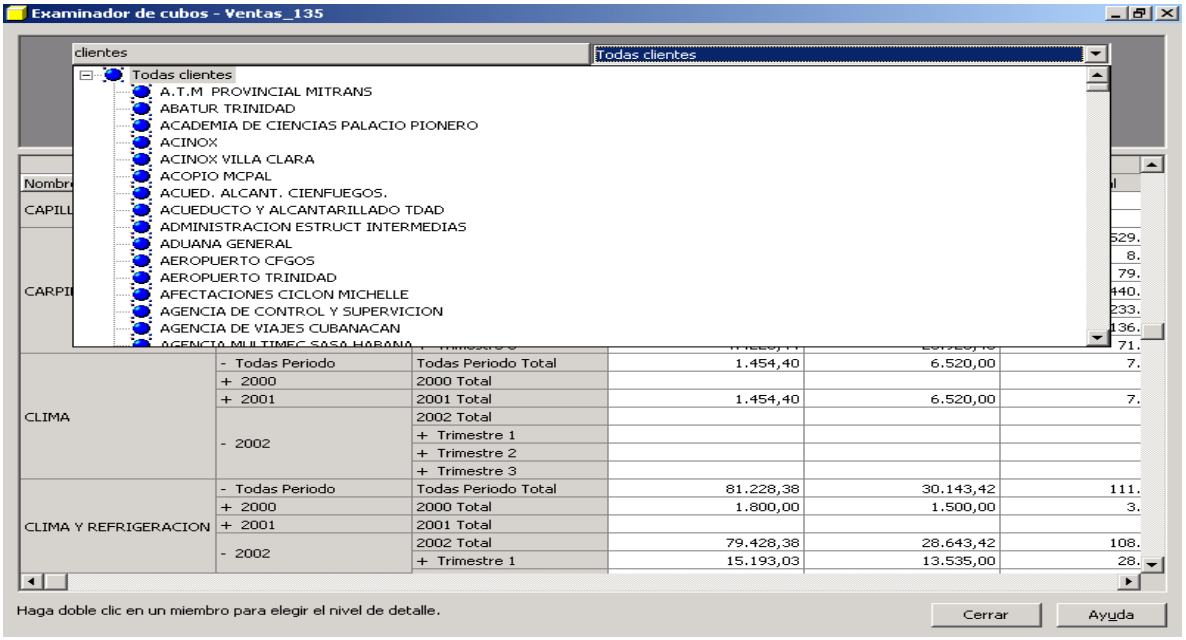

Figura 3.5. Haciendo *Slice* a un cliente

Como resultado del proceso anterior obtenemos la información mostrada en la figura 3.6.

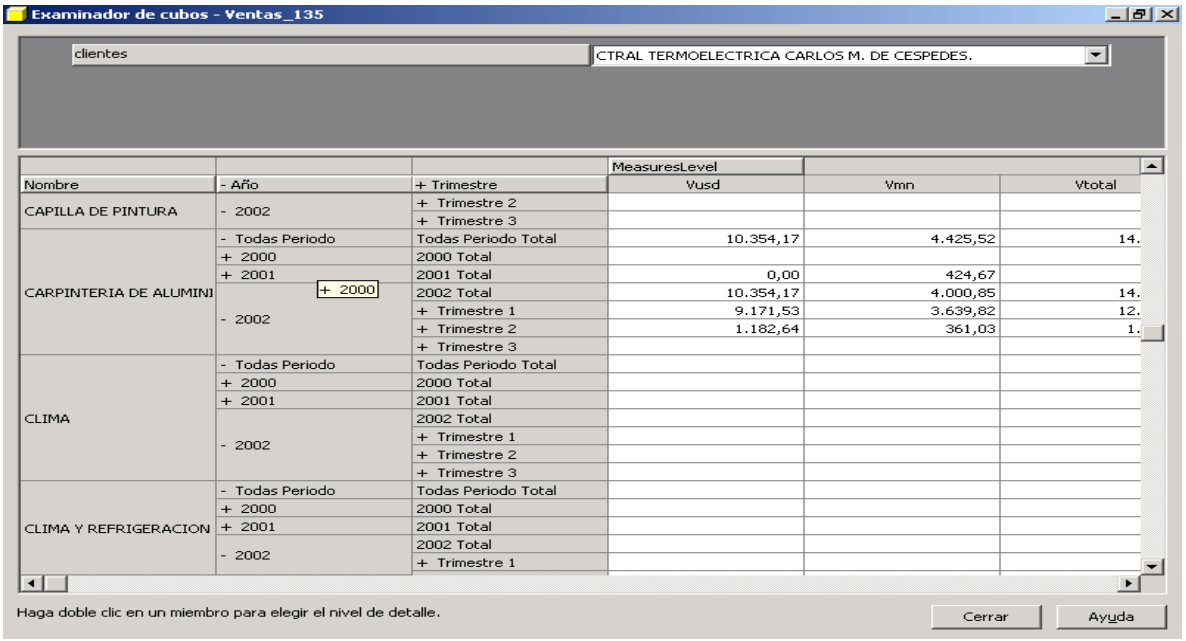

Figura 3.6 Resultado del Slice de la figura 3.6.

En fase de construcción se encuentra un conjunto de páginas WEB que hacen más asequible el acceso de toda la información contenida en el DW, para esto se utilizan páginas .asp con instrucciones MDX.

3.2. Definición de los Metadatos.

Por las características de prototipo que tiene el sistema desarrollado, fueron definidos solo los metadatos del negocio que auxilian al usuario final a utilizarlas herramientas de forma que pueda entender el papel de cada uno de los artículos que le son mostrados.

Para el depósito de los metadatos se utilizó una página WEB, donde se definen los objetivos de cada dimensión y el significado de cada medida. Ver Anexo No. 2.

La estructura definida en este depósito es la siguiente:

a) Fecha de la Última actualización.
b) Descripción Resumen del DWH.
c) Nombre y significado de cada dimensión.
d) Significado de las medidas en la tabla de hechos.

Con estos datos se puede entender mejor la función de cada tabla de dimensión utilizada en el sistema, así como los valores de interés que pueden ser explorados con más atención.

3.3. Plan de Capacitación

A modo de capacitación primaria, se instaló en la estación de trabajo del director las herramientas de clientes del SQL 2000 y el Análisis Manager con el que podemos realizar las tareas antes descritas, esta instalación nos permitió conocer algunas causas por la que no se explota un DWH, una de ellas es, por la ausencia de preparación del personal en las técnicas de explotación, por tanto, para que los usuarios exploten eficientemente el DWH, se necesita un conocimiento mínimo de la teoría relacionada a los Almacenes de Datos, así como el manejo de la aplicación utilizada como interfaz, teniendo en cuenta estos

problemas y en coordinación con el área de Recursos Humanos de la SUCURSAL CIENFUEGOS se diseña el siguiente Plan de Capacitación.

Este plan de capacitación está dirigido a los técnicos del Area Comercial, a los Directores de las Filiales y al Director General de la Sucursal, esta superación se debe a que son ellos los encargados de estudiar el mercado para la subsistencia productiva de la empresa.

3.2.1 Objetivos de la Capacitación.

Que los usuarios de los datos:

1. Conozcan los aspectos generales acerca de los Almacenes de Datos.
2. Conozcan los aspectos generales de las Bases de Datos y por las partes que las componen.
3. Aprendan el uso de la interfaz para la explotación del DW.
4. Conozcan el uso de las páginas WEB, como alternativa para la explotación del Almacén de datos.

3.2.2. Programa de Capacitación para la Explotación del DW

No. de Tema	Contenido	Horas
1	Introducción a los Almacenes de Datos Utilizamos de este material "Breve Introducción a los Almacenes de Datos"	2
2	Introducción al SQL 2000 • Concepto de Tabla (Campo, Fila, llave) • Concepto de relaciones. • Concepto de Consultas • Concepto de Base de Datos en SQL	2
3	Introducción al Analysis Service	2

	• Concepto de Dimensión • Concepto de Medida • Concepto de Cubo	
4	Uso de la Aplicación con los Datos de la Sucursal. • Drill Down • Roll up Recuperar Datos	6
5	Introducción a la Navegación (WEB)	2
	TOTAL	**14**

Conclusiones

El presente trabajo siguió un basamento teórico sobre las Bases de Datos Analíticas. Fueron abordados diversos factores que sirven de base al desarrollo de sistemas de Data Warehouse.

Como se puede notar en el trabajo, se presentaron las características de los sistemas de soporte a la toma de decisión, también se presentaron distintas arquitecturas que pueden ser utilizadas en los proyectos de diseño de un DWH, además de hacer una comparación entre los sistemas transacionales y los sistemas OLAP, quedando claro que es bastante interesante separar las aplicaciones dedicadas al control de las operaciones día a día de la empresa, de las aplicaciones destinadas a proveer el análisis de los datos obtenidos de las operaciones anteriores.

Se demostró con este trabajo, la viabilidad de diseñar e implementar un sistema de apoyo a la toma de decisiones, se citaron los pasos a seguir durante el diseño, para cumplir los objetivos propuestos.

En la práctica este trabajo demostró como se implementa un sistema DWH, utilizando una de las herramientas OLAP disponibles en el mercado, la cual analiza efectivamente los datos cargados en el DWH. Esta implementación se demostró a través del diseño y construcción de un prototipo en el área de Comercial.

Se llega a la conclusión de que los ambientes transacionales y analíticos son muy diferentes y que esta tecnología aunque haya surgido hace algunos años, todavía constituye un campo para la investigación y búsqueda de mejores métodos, vinculando conceptos del área informática, como Bancos de Datos, Sistemas Distribuidos, Internet, Intranet, Análisis y Diseño de Sistemas, entre otros.

Este trabajo sirvió también para despertar el interés para las áreas de apoyo a la toma de decisiones, un área carente de profesionales que diseñen e implementen sistemas y

consecuentemente pongan a disposición de las altas direcciones de las empresas, informaciones con gran calidad, estas altas direcciones necesitan con más frecuencia, desvincularse de la relación directa de dependencia entre los subordinados, para obtener las informaciones que apoyaran las decisiones por ellos tomadas.

Recomendaciones

- Al comienzo del diseño del DWH uno sus objetivos era “Reunir en un mismo depósito los datos necesarios para la alta gerencia de la empresa.”, este objetivo por las causas explicadas en la Fase de Revisión, solo se aplica a la Filial Cienfuegos, por tanto teniendo en cuenta esas causas, es que recomendamos:

    1. Crear el Clasificador de Clientes único para toda la empresa.
    2. Crear el Clasificador de UPS único para toda la empresa.

- Nuestra empresa hoy se encuentra en el Perfeccionamiento Empresarial, por lo que nuestras Filiales, son Unidades Básicas Económicas las cuales tienen bastante independencia en su gestión, por esta razón recomendamos de forma transitoria, hasta tanto no se creen las condiciones anteriores, implantar el DWH al nivel de cada una de ellas, para los análisis correspondientes.

- Continuar desarrollando el Sitio de WEB de la empresa, donde existan vínculos con el DWH.

- Recomendar a la Casa Matriz utilizar este diseño para lograr la base de datos informativa de Emprestur S.A. en Cuba.

Referencias Bibliográficas.

[1]. Inmon W. "Building the Data Warehouse". Wiley Computer Publishing, 1996, pág 14.

[2]. Kimball, Ralph “The Data Market Splita”. Data Warehouse Architect, pág 1 september 1995.

[3]. Kimball, Ralph “Is ER Modeling Hazardous to DSS”. Data Warehouse Architect, pág 1 october 1995.

[4]. Chaudhuri, Surajit e Dayal, Umeshwar. An Overview of Data Warehousing and OLAP Technology. ACM Sigmod Record, Mar/1997, pág. 21.

[5]. Campos, Maria Luiza e Rocha F, Arnaldo V. Data Warehouse. XVII Congresso da Sociedade Brasileira de Computação. 1997, pág.124.

[6]. Feltrin, Christian, Projeto e Desenvolvimento de Data Warehouse Hospitalar**.** Universidade Federal de Santa Maria, Centro de Tecnología, Departamento de Electrónica y Computación, Curso de Informática, Trabajo de Graduación. http://www.hcaa.com.br/antiga/dw/index.htm

[7]. Idem

[8]. Idem

[9]. Kimball, Ralph “The Data Warehouse Toolkit”. Wiley & Sons 1996, pág. 29.

[10]. Inmon W. "Building the Data Warehouse". Wiley Computer Publishing, 1996, pág. 37.

[11]. Kimball, Ralph “The Data Warehouse Toolkit”. Wiley & Sons 1996, pág.31.

[12]. Kimball, Ralph “The Data Warehouse Toolkit”. Wiley & Sons 1996, pág. 40.

[13]. Inmon W. "Building the Data Warehouse". Wiley Computer Publishing, 1996, pág 45.

[14]. Kimball, Ralph y Otros. “The Data Warehouse Lifecycle ToolKit”. Wiley & Sons 1998, pág. 187.

[15]. Silberschatz, Abraham. Database System Concepts. 3rd ed. McGraw Hill, 1996, pág 28.

[16]. Feltrin, Christian. Projeto e Desenvolvimento de Data Warehouse Hospitalar**.** Universidade Federal de Santa Maria, Centro de Tecnología, Departamento de

Electrónica y Computación, Curso de Informática, Trabajo de Graduación. http://www.hcaa.com.br/antiga/dw/index.htm

[17]. Idem.

[18]. Inmon, W.H. - Como construir um Data Warehouse - Editora Campus – 1996, pág 31.

[19]. Kimball, Ralph "The Data Market Splita". Data Warehouse Architect, pág 2 september 1995.

[20]. Kimball, Ralph "Mastering Data Extraction". Data Warehouse Architect, pág 1. June 1996.

[21]. Thomsen E. "OLAP Solutions". Wiley Computer Publishing, 1997, pág. 16.

[22]. Llacer, Enrique y otros. **"**DATA WAREHOUSING, Un paso más hacia la gestión del conocimiento en las empresas". Universidad de Sevilla, 1998, pág 2.

[23]. Barquini, Ramon. Planning and designing the warehouse. New Jersey: Prentice-Hall, 1996, pág. 311.

[24]. SAS Institute Inc. "SAS Rapid Warehosing Methodology". SAS e-Intelligence 2001, pág 17.

[25]. Dirección Económica, Manual de Procedimientos Emprestur S.A. Casa Matriz, 2002, Capitulo 5, pág 5.12.

[26]. Kimball, Ralph "The Data Warehouse Toolkit". Pág. 46. Wiley & Sons 1996.

[27]. Gunderloy, Mike. Jorden, Joseph L. Mastering SQL Server 2000. Sybex 2000.

Bibliografía.

Barquini, Ramon. Planning and designing the warehouse. New Jersey: Prentice-Hall, 1996.

Campos, Maria Luiza e Rocha F, Arnaldo V. Data Warehouse. XVII Congresso da Sociedade Brasileira de Computação. 1997.

Chaudhuri, Surajit e Dayal, Umeshwar. An Overview of Data Warehousing and OLAP Technology. ACM Sigmod Record, Mar/1997.

Dirección Económica, Manual de Procedimientos Emprestur S.A. Casa Matriz, 2002.

Feltrin, Christian, Projeto e Desenvolvimento de Data Warehouse Hospitalar**.** Universidade Federal de Santa Maria, Centro de Tecnología, Departamento de Electrónica y Computación, Curso de Informática, Trabajo de Graduación. http://www.hcaa.com.br/antiga/dw/index.htm.

Gunderloy, Mike. Jorden, Joseph L. Mastering SQL Server 2000. Sybex 2000.

Inmon W. "Building the Data Warehouse". Wiley Computer Publishing, 1996.

Kimball, Ralph "The Data Market Splita". Data Warehouse Architect, pág 1 september 1995.

Kimball, Ralph "The Data Warehouse Toolkit". Wiley & Sons 1996.

Kimball, Ralph "Mastering Data Extraction". Data Warehouse Architect, pág 1. June 1996.

Kimball, Ralph y Otros. "The Data Warehouse Lifecycle ToolKit". Wiley & Sons 1998.

Kimball, Ralph y Otros. "White Paper". Septiembre/1995 hasta Abril/2001

Kimbal, Ralph . "Dimensional Data Warehouse, The Business Perpective". Ralph Kimball Associates, 2000.

Llacer, Enrique y otros. **"**DATA WAREHOUSING, Un paso más hacia la gestión del conocimiento en las empresas". Universidad de Sevilla, 1998.

Ross, Margy. "Lifecycle Project Management". Ralph Kimball Associates, 2000.

Silberschatz, Abraham. "Database System Concepts". 3rd ed. McGraw Hill, 1996.

SAS Institute Inc. "SAS Rapid Warehosing Methodology". SAS e-Intelligence 2001.

Thomsen E. "OLAP Solutions". Wiley Computer Publishing, 1997.