

117



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

IMPLANTACION DE DATA
WAREHOUSE EN UNA
INSTITUCION DE CREDITO

T E S I S

QUE PARA OBTENER EL TITULO DE
" A C T U A R I A "P R E S E N T A :
MARIA CONCEPCIÓN URIBE ZUÑIGA

DIRECTOR DE TESIS:
M. EN C. ELISA VISO GUROVICH



FACULTAD DE CIENCIAS
UNAM



FACULTAD DE CIENCIAS
SECCION ESCOLAR



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Paginación

Discontinua



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

M. EN C. ELENA DE OTEYZA DE OTEYZA
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:

"Implantación de Data Warehouse en una institución de crédito"

realizado por **María Concepción Uribe Zúñiga**
con número de cuenta **073428163**, quién cubrió los créditos de la carrera de **Actuaría**
Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis
Propietario
Propietario
Propietario
Suplente
Suplente

M. en C. Elisa Viso Gurovich
M. en C. Virginia Abrín Batule
Act. Mauricio Aguilar González
Dr. Mario Magidin Matluk
M. en C. Agustín Román Aguilar

Elisa Viso Gurovich
Virginia Abrín Batule
Mauricio Aguilar González
Mario Magidin Matluk
Agustín Román Aguilar

Consejo Departamental de Matemáticas

José Antonio Flores
M. en C. José Antonio Flores



FACULTAD DE CIENCIAS
CONSEJO DEPARTAMENTAL
DE
MATEMÁTICAS

A MIS PADRES

Por haber me dado vida y sobre todo parte de la suya
Por su amor y porque a pesar de todo, siempre han
Estado ahí.

A JOSE MANUEL

Por creer en mí, gracias por tu cariño, paciencia y
apoyo.

A MIS HERMANOS Y HERMANAS

Por su fidelidad y compartir juntos los buenos y
malos momentos.

MI HIJA MAGALI

Fuente de inagotable
amor y felicidad

Indice

Capítulo 1 Introducción

Introducción	1
--------------	---

Capítulo 2 ¿Qué es el *Data Warehouse*?

2.1	Definición del <i>Data Warehouse</i>	8
2.2	Características	9
2.3	Arquitectura del <i>DW</i>	11
2.4	Componentes de la Arquitectura <i>DW</i>	12
2.5	Funciones del <i>DW</i>	21
2.6	Topologías para el <i>DW</i>	29
2.7	Minería de Datos (<i>Data Mining</i>) y <i>DW</i>	34
2.8	Pasos para la Elaboración de un <i>DW</i>	36

Capítulo 3 Implantación del concepto *Data Warehouse* en una Institución de Crédito

3.1	Descripción de el Banco (Institución de Crédito)	39
3.2	Necesidad del Sistema de Análisis	44
3.3	El <i>DW</i> del Banco	49
3.4	Plan del Proyecto	51
3.5	Plan Tecnológico del <i>DW</i> del Banco	58
3.6	Desarrollo	61
3.7	Implantación del <i>DW</i>	62
3.8	Resultados para el Negocio	67

Capítulo 4 Crítica y propuestas a futuro para el caso *DW* del Banco

4.1	Objetivo	70
4.2	Componentes	71
4.3	Funciones	75
4.4	Topología	81
4.5	Pasos para la elaboración de un <i>DW</i>	82
4.6	Recomendaciones y propuestas	84

Capítulo 5 Conclusiones del estudio

Conclusiones del estudio	107
Bibliografía	109
Anexos	111

Capítulo 1 Introducción

Los negocios actualmente se están desarrollando bajo un ambiente de alta competitividad, razón por la cual es prioritario contar con información que apoye a las organizaciones a tomar decisiones estratégicas y tácticas y que los lleve a lograr ventajas en el entorno actual.

Para lograr ventajas competitivas sostenibles, es necesario hacer un cambio en la visión de cómo la tecnología es utilizada. Debe de ser integrada a otros componentes de las organizaciones, como:

- La estrategia
- Las personas
- Sus operaciones

lo que significa la habilidad de ser lo suficientemente flexible a la transformación de la organización de una manera más rápida como respuesta a los cambios estratégicos y al reto del mercado.

Por otro lado, los negocios que mejor ventaja competitiva tienen son aquellos que además reconocen y manejan su información como su principal activo, creándose una necesidad constante de conocer a sus clientes/negocio a través del análisis de información en un periodo de tiempo definido.

La forma en la que la organización administre su información, es la forma en la que se verá beneficiada, es decir, la manera como la almacenan, la ordenan, la protegen, la distribuyen, y la forma en que permiten que se accese, podrá ocasionar la diferencia positiva de esta búsqueda.

Sin embargo, existen problemas que se enfrentan en el uso de la tecnología. Entre los más típicos podemos mencionar:

- Tratar de forzar nuevos sistemas para ajustarse a viejas e inflexibles metodologías y formas de trabajo, sin integrar la estrategia, operación y gente es ineficiente, produciendo sistemas que no satisfacen las necesidades de los negocios modernos e ignorando los muchos beneficios que puede dar la nueva tecnología.
- La información de los negocios está creciendo en extraordinarios porcentajes. Para quien administra la información, el manejar este volumen se vuelve cada vez un reto mayor. De esta forma la organización podría fallar en esta inmensidad de información a menos que se haga algo. También la información que un usuario necesita nunca podría ser encontrada una vez que ésta desaparezca en la enorme y creciente montaña de información.
- Otro problema es el uso que se le dé a la información. Frecuentemente vemos que entre más antiguo sea el negocio, más obsoletos pueden ser sus sistemas transaccionales u operacionales y si a esto le adicionamos que no identificamos adecuadamente el tipo de información que necesitamos ni tampoco de que tipo de sistema la obtenemos, creamos inconsistencias e insatisfacciones.

Se vuelve factor crítico de éxito el saber identificar ante un requerimiento, cuando es del tipo para análisis y cuando es sobre sistemas transaccionales u operacionales.

Bill Inmon y Kelley [Kelley 1994] hace referencia a que las aplicaciones transaccionales u operacionales soportan adecuadamente el proceso diario pero carecen de procesos de tipo analítico (sumarizada, conciliada, etc).

Entre las diferencias de un sistema transaccional u operacional y uno para análisis, se pueden mencionar:

Sistemas transaccionales	Sistemas de Análisis
Orientadas a solucionar la operación por aplicaciones	Integran información de varias aplicaciones
Operación por transacción	Operación por consultas de altos volúmenes de información
Sensibles a los tiempos de respuesta (segundos)	No preocupa los tiempos de respuesta (minutos, horas)
Los procesos corren en forma repetitiva	Corren en forma heurística
Actualizaciones en línea, al día	Datos que no se actualizan, sólo se consultan.

Algunos ejemplos de requerimientos de tipo analítico que se están formulando los negocios modernos son:

- Proporcionar los 10 mejores clientes con que se cuenta
- Qué tipo de servicios tienen los clientes que tienen un saldo promedio en sus cuentas de \$10,000.00
- En dónde acostumbran consumir los clientes que saldan sus Tarjetas de crédito mensualmente
- Que características en cuanto a edad, tipo de servicios que solicitan, etc. tienen los clientes que solicitan un crédito para vivienda
- Cuál es el promedio de venta de boletos a Cancún en Diciembre en los últimos 2 años.
- Qué clientes hacen más de 10 llamadas telefónicas nacionales a la semana.

Por otro lado, los problemas más fuertes a los que se enfrenta el área de Sistemas con respecto a estos nuevos requerimientos de información son:

- No hay integración de la información de las empresas. Los datos se encuentran repartidos en muchas aplicaciones. Los requerimientos necesitan, para ser satisfechos, de información que existe en más de un sistema. Además de que ésta se encuentra guardada en diversos medios de almacenamiento.
- Los sistemas que guardan esta información no están diseñados para este tipo de consultas. El enfoque con el que fueron creados implican un modelo rígido para resolver requerimientos precisos y con un mínimo tiempo de respuesta.
- No se tiene información histórica. Los sistemas transaccionales no la manejan y sólo tienen archivos de respaldo en cintas u otros dispositivos fuera de línea y sin forma alguna de accederlos (papel, microfichas, cintas, discos, etc.)
- No puede haber accesos para este tipo de requerimientos en horarios de operación. Soportan la operación del día a día y el hecho de que sean utilizados para consultas de análisis, impacta fuertemente a la operación, haciendo muy lenta la respuesta a las transacciones.
- Se debe integrar a varias personas para un requerimiento. Debido a que los sistemas pueden ser administrados por diversas personas, y cada quien es experto en lo suyo, es necesario coordinarlos para resolver un requerimiento de análisis.
- No necesariamente existe una terminología implementada en los sistemas por lo que se debe hacer un esfuerzo adicional para identificar los datos que se pueden llamar de distinta manera en diversos sistemas.
- Los sistemas no tienen el mismo punto de sincronía, ejemplo: fechas de cierre. En ocasiones los nuevos requerimientos implican información en cierto momento o estado y se debe coordinar dicho evento en los diversos sistemas que involucra la petición.

- Los tipos de datos no corresponden en su diseño físico entre sistemas, por ejemplo: la longitud del nombre, apellidos, etc.
- Los datos necesarios no existen o están en blanco. Es común encontrarlos con estructuras que sugieren algún dato, pero al explorarlo nos encontramos que no se captura o es utilizado para otro fin.
- Otro problema importante es la falta de un responsable de la información, generando confusión en su interpretación.

Para tratar de resolver este tipo de problemas y satisfacer las necesidades de información de análisis, existen dos formas que podemos resumir en:

- Obtener extractos para atacar los requerimientos
- Construir una arquitectura de información basada en crear Almacenes de Información, conocidos como *Data Warehouse (DW)*.

La creación de extractos de información es una forma de resolver este tipo de problemas. La idea es generar un archivo o conjunto de archivos de donde se accederá a la información para producir las soluciones a los requerimientos. Cuando un usuario hace una consulta, la gente de sistemas busca los archivos operacionales maestros donde se genera esta información y la replica en un dispositivo independiente. Luego mediante consultas satisface el requerimiento.

Este método se caracteriza por ser sencillo, ya que depende de hacer copias de información, pero tiene varias desventajas, entre ellas:

- Tiende a generar tantos extractos como requerimientos, con el riesgo de pérdida de control de las versiones y/o archivos.

- No son flexibles al cambio, implica desarrollo adicional y mantenimiento por nuevos requerimientos
- No garantiza la integridad de la información

La generación de un *DW* trata de construir un almacén de información de análisis, construyendo una base de datos con visión corporativa garantizando que los datos llegarán homogéneos; tendrán una definición y estructura única, y las aplicaciones operacionales sólo se preocupan por dejar los datos bajo un estándar establecido.

Esta propuesta beneficia a los responsables de los sistemas aplicativos, ya que elimina la generación de muchos extractos copiados de su sistema.

El presente trabajo, se realiza probando el concepto de *Data Warehouse* en una de las principales Instituciones de Crédito de nuestro país, e identificando los beneficios que trae su uso y consideraciones para futuras instalaciones.

En el siguiente capítulo se expone brevemente la importancia de contar con una arquitectura tecnológica claramente definida, que permita mayor flexibilidad en su evolución. Su importancia es tal que influye directamente en el éxito o fracaso de la acción de la tecnología sobre el negocio. El *Data Warehouse* tiene su fuente en esta arquitectura. Dentro de este concepto ésta contenida como un componente más la arquitectura de información.

Antes de iniciar el desarrollo del caso práctico, se expone una propuesta sobre *Data Warehouse*, sus componentes y funciones.

Posteriormente se presenta un caso práctico realizado para probar el concepto, describiendo los pasos efectuados y los resultados obtenidos, mostrando el efecto positivo que tiene el *Data Warehouse* sobre los sistemas de análisis y de toma de decisiones y presentándose como una opción eficaz para las organizaciones que necesitan este tipo de procesos.

Finalmente se describen algunas consideraciones para futuras implementaciones de *Data Warehouse*.

Barry Devlin [1997] dice " La información es una representación del entendimiento de los negocios y su uso por usuarios finales". Este trabajo está dirigido a las personas interesadas en cómo la Tecnología de Información apoya a los negocios y, más específicamente a los interesados en los sistemas de análisis, ya que éstos aportan ventajas competitivas a las organizaciones.

CAPITULO 2 ¿Qué es el *Data Warehouse*?

2.1 Definición del *Data Warehouse*

El concepto de *Data Warehouse* se traduce como Almacén de Datos Almacén de Información. Durante el presente trabajo, cuando nos refiramos a él, usaremos *DW*.

Iniciaremos con una definición de Bill Inmon [1994], considerado como padre de *DW*, quien propone:

"*DW* es una colección de datos que soporta la toma de decisiones y cuenta con cuatro características básicas:

- Orientada a Tópicos
- Integrada
- No editable
- Acumulable en el tiempo"

En otras palabras, el *DW* es una base de datos alimentada directamente de los sistemas operativos o transaccionales, organizada por tópicos, con los datos totalmente integrados, éstos no se actualizan, sólo se consultan y son fotografías en el tiempo.

La forma en la que se almacena la información, está determinada por las características arriba mencionadas y que describiremos a continuación.

2.2 Características del *DW*

Bill Inmon [1994], define las características de la siguiente forma:

2.2.1 Orientada a Tópicos

Partiendo de que la información operacional es diferente de la analítica, la forma en la que podemos organizarla también es diferente. Pongamos como ejemplo a una institución bancaria donde existen muchos sistemas de información operacionales, como pueden ser cheques, tarjeta de crédito, préstamos hipotecarios, cartera comercial, inversiones, etc. Estos sistemas que soportan la operación del banco, pueden estar separados, de acuerdo a la conveniencia de la institución, incluso en diferentes plataformas (equipo central, plataforma *Risc*, *Unix*, *PC's*, etc.). Este tipo de organización podemos llamarla orientada a aplicaciones. Cuando hablamos de necesidades de información analítica nos hace pensar en una forma diferente de organizar la información, en base a tópicos.

Los tópicos son conceptos genéricos, que agrupan un conjunto de datos. En la operación pudieran estar separados. Ejemplos de tópicos son clientes, proveedores, actividades, contratos, tipos de transacciones, ubicaciones, etc.

2.2.2 Integrada

Una vez diseñado en *DW* el base a Tópicos, es necesario extraer de diversas fuentes de la operación para alimentar un tópico en especial. Esto implica la integración de diversas aplicaciones en torno a éste. Como las aplicaciones de la operación son independientes (uno de los problemas

mencionados en el capítulo primero), la integración de la información es un factor clave en el desarrollo del *DW*.

Dentro de las actividades que se deben llevar a cabo para integrar la información de varias aplicaciones al *DW* están:

- Estandarizar los tipos de datos y formatos que se almacenarán
- Conversión de tipos y formatos de datos originales al formato convenido
- Estandarizar la nomenclatura que se utilizará, etc.

Estas actividades tienden a ser claves y a regir la población de información, ya que por provenir de diferentes aplicaciones que deben ser integradas, nos podemos encontrar con muchas definiciones distintas y en el *DW* tenemos que establecer un acuerdo común para alimentarlo.

2.2.3 No editable

Una característica del *DW* es que esta información es cargada de la operación y no puede ser alterada después. Esto funciona como el tomar una fotografía de la operación de la organización, en un momento dado y pasarla a almacenar al *DW*.

Mientras que en la operación las aplicaciones están diseñadas para soportar en línea inserciones de información, borrados, alteraciones de una manera rápida, el *DW* sólo la extrae, la almacena y permite su consulta para análisis. De ninguna manera se puede alterar, ya que representa un estado de la información y por lo tanto del negocio en un tiempo determinado.

Las modificaciones a la información, siempre ocurren sobre la operación y posteriormente éstas se reflejan o se guardan en el *DW*.

2.2.4 Acumulable en el tiempo

El manejo de información histórica, es una característica muy importante para el *DW*, ya que como mencionamos en el punto anterior, es como si almacenáramos fotografías del negocio en el tiempo. Éstas fotos las podemos almacenar el tiempo que sea necesario.

En las aplicaciones operacionales, regularmente tenemos información actual, es decir, tenemos el estado más reciente del negocio. Si pudiéramos como ejemplo a una institución de crédito en una de sus múltiples aplicaciones como cuentas de cheques, en la aplicación operacional se maneja el estado de cuenta en donde se detallan los movimientos diarios de un período (por ejemplo: un mes) o cuando más hasta dos o tres períodos atrás, sin embargo en el *DW* nos puede dar capacidad para almacenar información incluso de años, todo dependerá de la frecuencia con la que se actualice el dato y el uso que se le dará. Con esto el *DW* nos da una perspectiva en algún tiempo específico.

El manejo y la implantación física de historia en el *DW* nos lleva a pensar en dispositivos no tan rápidos y caros como los sería en la operación.

2.3 Arquitectura del *DW*

Dentro del ambiente *DW* podemos identificar los componentes que juegan en este entorno y que se encuentran presentados en la figura 1 dentro de la arquitectura *DW*.

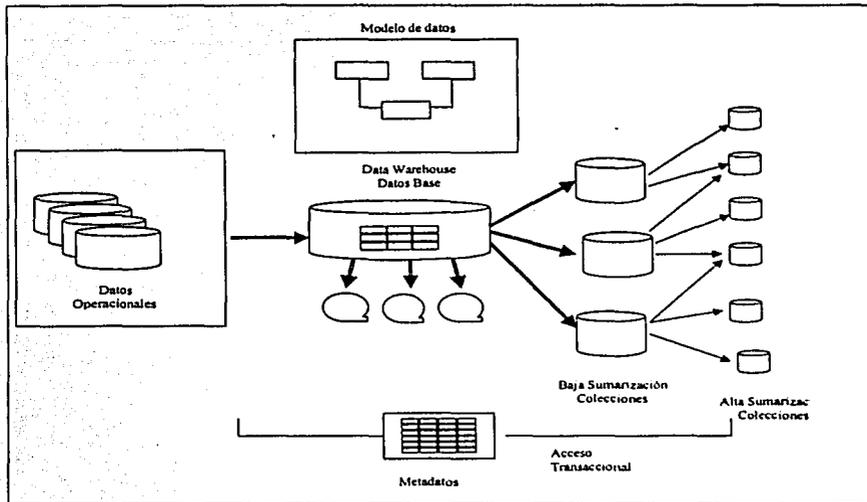


Figura 2.1 Arquitectura del DW

2.4 Componentes de la Arquitectura DW

2.4.1 Datos Operacionales

Se refieren a la información que se maneja en las aplicaciones operacionales del negocio (aplicaciones vitales). A partir de esta información se va a cargar el DW.

Las diversas aplicaciones que manejan los datos operacionales pueden estar en diversas plataformas, ubicaciones geográficas, programados en distintos lenguajes, utilizando diferentes manejadores de bases de datos o formas de almacenamiento, etc. El contar con los datos operacionales para poblar al DW,

muchas veces se convierte en un problema difícil, ya que al tenerlos dispersos requiere de conocer muy bien cada ambiente donde se encuentran los datos operacionales que necesitamos, cosa que para instituciones grandes y complejas no es nada trivial.

2.4.2 Base del *DW*

Es la colección de datos que nos apoya en las funciones de análisis en la organización. Esta base es de carga de los datos operacionales y guarda un alto grado de detalle que frecuentemente se puede hablar que existe el mismo nivel de granularidad de información que en la misma aplicación operacional. Se organiza de distinta forma, enfocándose al acceso solamente para el *DW*.

Otra característica es que permite guardar la información histórica. Para esto la información se almacena de forma similar que la que estará disponible para acceso. Es muy importante planear la capacidad de almacenamiento del *DW*, ya que como podemos analizar, tiende a crecer de forma muy rápida, por lo que debemos calcular y definir en qué tipo de dispositivos se guardará la historia. Estos dispositivos pueden ser no tan rápidos ni caros, como los que se utilizan para la operación, tales como cintas, discos ópticos, etc.

Es importante mencionar que el *DW* se integra de información operacional que incluye muchas veces información externa (de otras compañías, de dependencias oficiales y gubernamentales, etc.).

2.4.3 Información Histórica

Como vimos con anterioridad, cuando por conveniencia decidamos mover de la base del *DW* la información más antigua, esta puede ser movida a

dispositivos más baratos y que permitan el acceso cuando sea necesario. La información vieja tiende a ser menos solicitada que la actual. El ponerla fuera de los discos de acceso rápido nos puede ayudar con nuestra planeación de la capacidad a la vez que nos permite guardar más tiempo dicha información. Estos datos se utilizan, por ejemplo, al hacer pronósticos y analizar comportamientos en el tiempo. Al mover información de la base del DW a la información histórica se deberá hacer tomando en cuenta que ésta podrá ser necesitada y por lo tanto se debe construir la forma de consultarla.

2.4.4 Información de Baja Sumarización

La información de baja sumarización es aquella que es conciliada a partir de la información base que presenta un gran detalle. Se calcula de la dimensión jerárquica más baja representada por uno o dos pasos más arriba de la fuente original de la información y que sirve para generar el siguiente nivel

La información se sumariza en grandes rubros sin perder todo el detalle. Estas colecciones se generan pensando en las preguntas más usuales de un grupo o departamento de personas. Regularmente existen varios conjuntos de colecciones a este nivel y la frecuencia de acceso sobre los datos de sumarización baja es mayor a los datos de la base.

2.4.5 Información de Alta Sumarización

Es generada a partir de la información de baja sumarización y está en la parte más alta de la dimensión jerárquica de la información. Este nivel está orientado a grupos pequeños o individuos y por lo tanto su diseño está enfocado a necesidades muy específicas. Este bloque se conoce también como datos derivados.

Dentro del ambiente *DW*, este nivel tiende a ser muy dinámico ya que su enfoque es el de apoyar la toma de decisiones del momento. Responde a los cambios de análisis. Este trabajo se vuelve flexible, con lo que la formación de los datos derivados no debe implicar un gran esfuerzo.

2.4.6 Mercado de Datos (*Data Mart*)

Algunos autores utilizan el concepto de Mercado de Datos (*Data Mart*) para representar una colección específica de datos para un fin en particular. Se propone el siguiente concepto:

Un Mercado de Datos es un subconjunto de datos descentralizado que proviene del *DW*, diseñado para soportar los requerimientos de una sola unidad de negocio y para un sistema de toma de decisiones específico.

El concepto de Mercado de Datos se puede confundir fácilmente con el concepto de *DW*. El DM es una parte de la arquitectura *DW* que corresponde a las colecciones departamentales e individuales que surgen a partir de la base del *DW*. En este trabajo, a los DMs se les identificarán con las colecciones de baja y alta sumariazación .

2.4.7 Metadatos

Metadatos son los datos acerca de los datos [Inmon, 1994]. El componente de metadatos no es otra cosa que un directorio que nos permite conocer que es lo que tenemos en el *DW* y cuál es su estado. Mediante los metadatos podemos navegar de forma sencilla por la información contenida en el *DW*. Los metadatos apoyan para tener una mejor comprensión y credibilidad ya

que, dependiendo de su complejidad, almacenan información de qué hay en el *DW*, donde se originó, cuándo se originó y se cargó en el *DW*, quién es responsable de la información; nos aclara la transformación que sufrieron los datos desde su origen hasta el *DW*. Además, es una herramienta muy importante en el manejo de la información histórica ya que al paso del tiempo es normal que cambien las estructuras de los datos. Nos apoyan a conocerlas a través del tiempo por lo que, podremos acceder información de meses atrás aunque su estructura haya cambiado. Inmon [1994] propone que los componentes de los metadatos los podemos clasificar en:

- Básicos
- Mapeo
- Extracción histórica
- Misceláneos
- Sumarización
- Relaciones
- Propietario/Usuario
- Patrones de Acceso
- Tablas de referencia
- Modelo de datos

Básicos.- Los componentes básicos de los metadatos del *DW* incluyen información de las tablas que contiene el *DW*, las llaves de esas tablas, así como los atributos de las mismas.

Mapeo.- Este componente guarda información de la forma en que se llevaron los datos de la fuente al *DW*, incluyen la identificación de los campos fuente, relación del mapeo campo fuente a campo *DW*, la conversión del atributo, conversión de características físicas, cambio de nomenclatura, entre otros.

Extracción histórica.- La información referente a la extracción histórica de los metadatos simplemente le dice al analista usuario cuándo se cargaron los

datos al *DW* y por tanto permite conocer cuándo fue la última vez que la información se refrescó.

Misceláneos.- Este tipo de componente apoya la operación del *DW* almacenando información como los nombres alternos de los atributos o alias, los estados en que se encuentran las bases de datos, estadísticas de volumen de la información, criterios del ciclo de vida de la información en *DW*, etc.

Sumarización.- Como hemos visto, el *DW* se compone de información a diversos grados de sumarización; en este componente se documentan las fórmulas y algoritmos usados para estas sumarizaciones

Relaciones.- Este componente sirve para indicarnos las relaciones entre las tablas del *DW* y en base a qué atributos se puede ligar información de éstas, así como mantener la integridad de estas relaciones.

Propietario/Usuario.- Este componente tiene la información de las responsables de la información fuente (operacional) que ingresa al *DW*, lo cual es de gran ayuda para el analista usuario quien sabe con quién dirigirse en caso de alguna duda.

Patrones de acceso.- Aquí se guardan básicamente estadísticas del acceso a las tablas del *DW*, qué tipo de acceso, volumen de información, frecuencia, etc. Con esto se puede optimizar el diseño para tener mejores formas de obtener información del *DW*.

Tablas de referencia.- Como vimos en capítulos anteriores, la información de referencia tiene una estrecha relación con las tablas de base de *DW*. En este módulo se tiene información de estas tablas.

Modelo de datos.- Es una representación de los datos de un sistema. En los metadatos se guarda el diseño físico del *DW*.

Los metadatos en el *DW* juegan un papel importante ya que soportan la forma en que se accesan los datos. La forma de implementarlos puede ser tan sencilla o compleja como se requiera, dependiendo de cuántos componentes, de los que hemos visto, se le quieran incluir a los metadatos.

2.4.8 Modelo de Datos

El *DW* necesita de un modelo de datos que presente la información y la forma en que ésta se organiza.

Es pieza fundamental en el *DW*, al grado tal de que si no diseñamos adecuadamente el modelo, éste no logrará los objetivos que se han trazado.

La base de datos es la concretización del Modelo de datos.

Para el desarrollo o modificación de una base de datos se requiere de pasar por las siguientes tres etapas:

- Análisis de Datos
- Diseño del Modelo de Datos
- Construcción de la Base de Datos

2.4.8.1 Análisis de Datos

El objetivo del Análisis de Datos es el determinar de acuerdo a los requerimientos, determinar qué información se va a almacenar. Regularmente

esta fase contempla dos tipos de modelos: Modelo de Objetos y Modelo de Datos Lógico.

El Modelo de Objetos representa las necesidades de información desde un punto de vista de negocio puro. El Modelo de Objetos no es un Modelo de Datos, ya que en este modelo se analizan objetos existentes que los expertos del negocio conocen y manejan por ejemplo: gente, edificios, clientes, cuentas de bancos, etc., y se definen las relaciones de dichos objetos de manera muy general. De esta manera, los objetos de negocio que aquí se analizan deben ser capaces de contemplar las necesidades de la organización utilizando términos del negocio. A partir de este modelo, se genera el Modelo de Datos Lógico.

A partir del Modelo de Objetos transformándolo en tablas lógicas (o registros lógicos), se obtiene el Modelo de Datos Lógico que llamaremos simplemente modelo de datos. A este nivel agregan aspectos del proceso, tales como, seguridad de datos o auditoría. Este modelo presenta las tablas lógicas de datos necesarios para satisfacer requerimientos e incluye información de tablas lógicas, columnas o elementos y sus características. Las tablas lógicas pueden no coincidir con las tablas físicas. En la figura 2.2 se muestra un ejemplo de un modelo de objetos y un modelo de datos lógico.

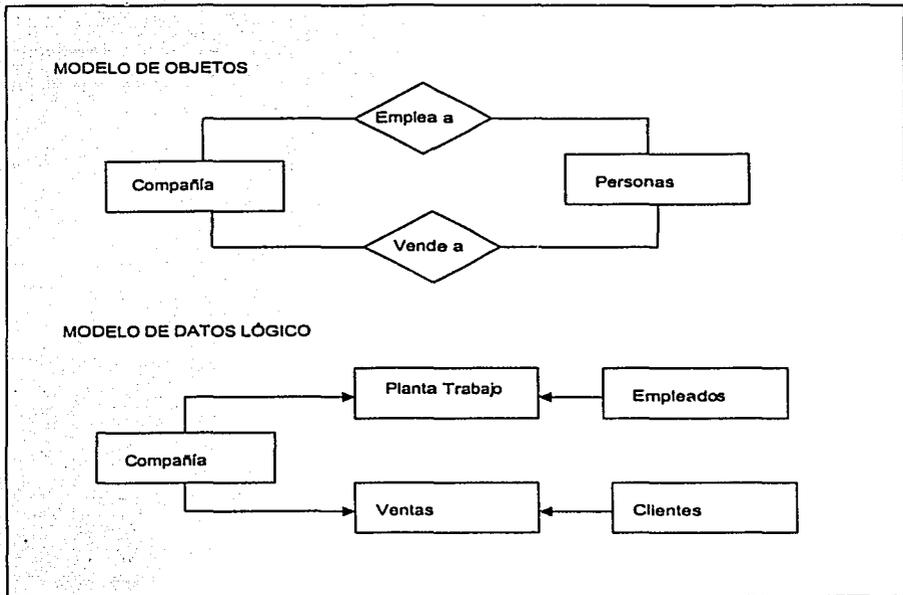


Figura 2.2 Modelo de Objetos vs Modelo de Datos Lógico

2.4.8.2 Diseño de la Base de Datos

Esta fase es la correspondiente a la elaboración del Modelo de Datos Físico. El objetivo de esta fase es describir la estructura que tendrá nuestra Base de datos. En este paso se transforma el Modelo Lógico en el Modelo Físico y se adicionan factores referentes a la tecnología y que se relacionan con el manejador de bases de datos específico en el cual se va a construir la base de datos. Con esto se busca que el diseño cubra los requerimientos de información, además de optimizarlo en rendimiento dentro de la tecnología que se va a utilizar.

Usualmente el DBA (Administrador de Base de Datos) es el responsable de esta fase.

2.4.8.3 Construcción de la Base de Datos

El objetivo de esta fase es tomar el diseño físico y generar la base sobre el manejador seleccionado. El gran principio de esta fase es en sí que se debe usar el modelo físico para desarrollar la base de datos. Si se encontrara un problema en la construcción de la base de datos, se para esta fase, se revisa y modifica el diseño físico para después regresar a esta fase de construcción. Si se tiene el problema en la parte de requerimientos, se analiza el problema desde el diseño lógico para continuar con las fases del modelo.

2.5 Funciones del *DW*

Para que la arquitectura del *DW* pueda funcionar correctamente se han detectado seis funciones básicas: Adquisición, Administración, Operación, Navegación, Acceso y Distribución. La figura 2.3 muestra dichas funciones.

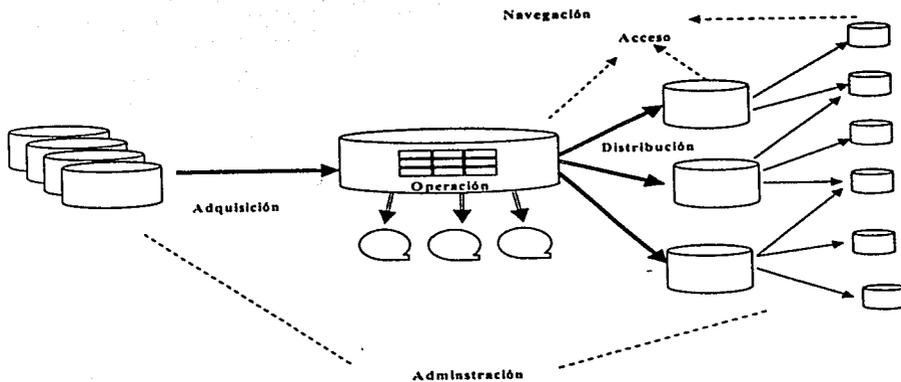


Figura 2. 3 Funciones del *DW*

2.5.1 Adquisición

La adquisición de datos, se refiere a la función de llevar los datos del nivel operacional al *DW*.

Dentro de las actividades que se realizan en esta función tenemos:

- Extracción

- **Transportación**
- **Depuración**
- **Mapeo**
- **Transformación**
- **Carga**

Extracción.- Es la actividad de obtener una imagen de la información de uno o más sistemas operacionales en un momento dado. Es como si tomáramos una fotografía de la información operacional que necesitamos para el *DW*.

Transportación.- Una vez que hemos realizado la extracción debemos enviar esta información desde la ubicación de los sistemas operacionales, hasta la ubicación del *DW*. Esto puede hacerse con procedimientos que transfieren archivos automáticamente o por medio de transporte de cintas.

Depuración.- Esta actividad inspecciona los datos en cuanto a calidad e integridad. Como resultado, puede descartar datos, marcarlos, modificarlos o enriquecerlos con el fin de mejorar el contenido de la información. Este es un proceso muy complejo y debe manejarse muy cuidadosamente ya que se trata de mejorar la información, sin embargo, no debemos cambiar los resultados que vienen de la operación, es decir, no debemos reportar cosas diferentes en la operación y el *DW*.

Mapeo.- Consiste en relacionar la entrada (sistema operacional) y la salida (*DW*). Esto lo hace al mantener la correspondencia entre los campos y/o columnas de la operación con columnas del *DW*.

Transformación.- Se encarga de convertir valores a estándares predefinidos. Se basa en cálculos y agregaciones para alimentar al *DW*, tomando como origen los campos de la operación.

Carga.- Consiste en poblar el *DW* desde uno o más sistemas operacionales. Una vez realizadas las actividades previas, se podrán obtener los datos preparados para cargar las tablas del *DW*.

2.5.2 Administración

Los servicios de administración están van más relacionados con cómo vamos a usar la información del *DW*. Contempla la creación de los niveles de sumarización así como el manejo de la información histórica. Dentro de estas actividades podemos mencionar:

- Construcción de Colecciones de Baja Sumarización
- Construcción de Colecciones de Alta Sumarización
- Inventario del *DW*
- Administración de datos históricos

Construcción de Colecciones de Baja Sumarización.- Corresponden a las bases de datos de tipo departamental.

Dentro de esta actividad, debemos conocer los datos que existen en la base del *DW* para poder encontrar la mejor manera (sencilla y rápida) de generar las colecciones departamentales.

Construcción de Colecciones de Alta Sumarización.- Son las colecciones de datos individuales que implican un alto grado de sumarización. Se alimentan de las colecciones departamentales; el punto está en conocer bien las fuentes para generar las nuevas colecciones de una manera óptima.

Inventario del DW.- Esta actividad sugiere el tener un catálogo de la información que se tiene disponible en todos los niveles de la arquitectura (base, departamental, individual e histórico) con lo que podemos detectar formas de optimizar el almacenamiento de dichos datos.

Administración de Datos Históricos.- Permite que la información histórica fuera de línea sea consistente para poder explotarla y que le dé sentido a nuestro análisis. Se deben manejar versiones de la información y almacenar los cambios para después poder interpretarla. No debe confundirse con los archivos de respaldo y recuperación. Esta actividad es compleja por el volumen de datos que maneja.

2.5.3 Operación

En esta función debemos asegurar que se realicen actividades que garanticen la continua operación del ambiente *DW*. Las actividades generales que esta función propone son:

- Uso de Monitores
- Administrador de Eventos
- Respaldo y Recuperación

Uso de Monitores.- Es importante hacer mediciones periódicas respecto a los patrones para el ambiente *DW*. Esto incluye situaciones como el verificar la capacidad del espacio, verificar tiempos de respuesta, tablas con mayor contención. Al establecer estas mediciones podrá analizarse qué acciones se tendrán que realizar para evitar la generación de cuellos de botella que perjudiquen el desempeño del *DW*.

Administrador de Eventos.- Permite especificar, invocar y dar seguimiento a eventos dentro del ambiente *DW*. Estos eventos pueden ser el extraer información, cargar información, reorganizar el *DW*, realizar respaldos, generar colecciones, etc y pueden llevar un orden específico y precedencias. El administrador de eventos permite manejar estas situaciones, además de controlar mensajes y alertas para apoyar la función de escalamiento.

Respaldo y Recuperación.- Otra actividad importante es el contar con procesos de respaldo y recuperación en casos de contingencia. Dentro del ambiente *DW*, esta actividad no es muy compleja en cuanto a que es un sistema de consulta, pero es complicado al manejar enormes cantidades de información.

2.5.4 Acceso

Esta función representa la forma en que los usuarios y consumidores obtendrán la información del *DW*. Este acceso puede llevarse con fines de llegar al usuario final o bien para generar otras colecciones a otros niveles de sumarización. Para realizar el acceso existen básicamente dos tipos de herramientas:

- Generales
- Propósito particular

Actualmente existen en el mercado diversas herramientas sencillas de usar con el fin de acceder bases de datos. Dichas herramientas se basan en el uso de interfaces gráficas, diseño cliente servidor, uso de metadatos para navegación y un sencillo control y mantenimiento.

Sistemas de Propósito General.- Estas herramientas por lo general accesan bases de datos y presentan información, permitiendo desarrollar consultas abiertas como son los sistemas reportadores o los sistemas de consultas (*Query Tools*). También se encuentran herramientas a las que se les agrega algún tipo de análisis general y presentación compleja, dando paso a los sistemas para ejecutivos (*EIS*) o sistemas de soporte a las decisiones (*DSS*).

Otro tipo de herramientas usadas sobre todo para generar colecciones, son utilerías que en Lenguaje estructurado de consultas (*SQL – Structure Query Lenguaje*), pueden generar archivos que después se cargan en tablas de otro nivel de sumarización.

Sistemas de Propósito Específico.- Son otra forma de acceder el *DW*, tales como sistemas de análisis de mercado (*Marketing*), de análisis de comportamiento de clientes, venta cruzada, análisis financiero, hojas de cálculo, etc. Estas herramientas están diseñadas para un tipo específico de análisis y tienen funciones complejas especializadas para dicho trabajo.

2.5.5 Navegación

Los servicios de navegación permiten a los usuarios saber que la información está disponible, dónde se ubica. Facilita encontrar más rápido la información dentro del *DW*. Dentro de sus actividades tenemos:

- Navegador de Datos
- Lista de Productores

Navegador de Datos.- Representa el punto central para entrar al *DW*. Tiene una relación muy estrecha con los metadatos, ya que funciona como el asistente para conocer la información disponible, formatos, versiones, fechas de

carga y de esta forma apoya al usuario a obtener la información correcta; y si no existiera ésta, apoya al análisis para detectar si puede ser generada con la ya existente.

Lista de Productores.- Es la lista de los datos de las personas responsables de las sistemas que alimentan el *DW*, ya que en caso de alguna duda o aclaración respecto a los resultados mostrados, facilita conocer a quién dirigirse para la aclaración de cualquier situación.

2.5.6 Distribución

La función de distribución permite el envío de información del *DW* a otro sitio de procesamiento, esto para creación de colecciones en un lugar más apropiado para el usuario final. Dicho de otra forma, muchas veces y como lo veremos en el siguiente punto, nos conviene tener alguna configuración donde ciertas colecciones se ubiquen en otro sitio físico, de esta forma con la función de distribución podremos enviar la información, de acuerdo a un requerimiento o sobre un programa de envíos. Este servicio puede funcionar como una suscripción a una revista donde periódicamente se recibe el producto; en este caso periódicamente recibirán información de la base de datos del *DW*.

2.5.6.1 Administración de Ordenes de Envío

Dentro de esta actividad, los usuarios podrán concertar un servicio de distribución. Se hace mediante una orden donde se especifica qué información y en qué periodos se necesita. En otros casos el envío de los datos no dependerá necesariamente de algún periodo constante, sino que la distribución se puede hacer de acuerdo a otro tipo de evento, como por ejemplo podemos solicitar que

se envíe la información sólo cuando el total de las ventas sea menor a una cantidad.

2.5.6.2 Monitoreo de Agentes

Una vez que se han puesto las órdenes, la ejecución de éstas se relacionan con un agente. Este agente se encargará de ejecutar los procedimientos necesarios para distribuir la información. En este punto se sugiere tener un mecanismo para dar seguimiento a estos agentes.

En general, la función de distribución tiende a ser muy compleja, ya que no existen herramientas maduras que apoyen esta función. Para poner un ejemplo de lo difícil de esta función, podemos imaginarnos una implantación de un *DW* donde la base se ubique en un equipo mayor y las colecciones puedan estar en minicomputadoras o en PC's. La función de distribución debe ser capaz de enviar automáticamente la información de las bases a las colecciones, sin importar la plataforma en que residan.

2.6 Topologías para el *DW*

La forma en que estructuramos el *DW* depende en gran medida de la estructura de la empresa.

Si tenemos una institución fuertemente distribuida, es posible pensar en una topología de *DW* distribuido, pero si nuestra empresa tiene una administración centralizada, nos convendrá centralizar nuestro *DW*. A continuación se muestran las topologías generales en las que un *DW* puede ser constituido. Experiencias en *DW*, indican que en un principio es mejor empezar con una topología centralizada y de acuerdo a las necesidades de la institución, buscar la distribución.

2.6.1 DW Central-Único

La topología centralizada propone el desarrollo del *DW* en una plataforma central. La información de las aplicaciones operacionales que van a alimentar el *DW* pueden estar en diferentes plataformas, pero el almacén de información estará en una sola plataforma. Este esquema incluye la información base como las colecciones sumariadas, lo que hace que el acceso se haga en la misma plataforma, es decir, las herramientas o sistemas de acceso deberán conectarse a esta plataforma única.

Esta opción presenta facilidades en el desarrollo por su sencillez, pero en ocasiones puede mostrar dificultades cuando queremos hacer uso de herramientas de acceso que no fueron diseñadas para la plataforma donde está el *DW*. La figura 2.4 ilustra esta topología.

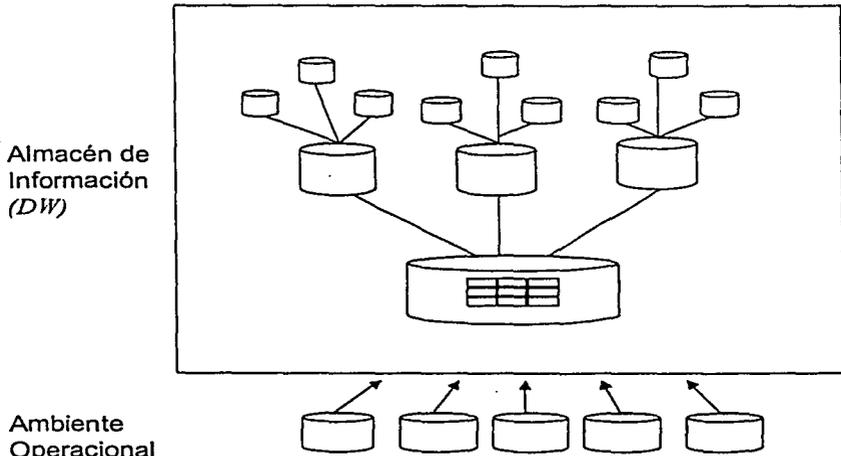


Figura 2. 4 Topología DW Central-Único

2.6.2 DW Central con Fuentes Distribuidas (separadas geográficamente)

Esta topología sugiere que la información que alimenta al *DW* sea organizada previamente en almacenes (similares a un *DW*). Este esquema es recomendable cuando los sistemas operacionales de una compañía están organizados en forma distribuida; de esta forma se agrupa la información por tópicos o regiones antes de cargarla al *DW*. La información del *DW* permanece centralizada en una misma plataforma. La figura 2.5 muestra un diagrama de esta topología.

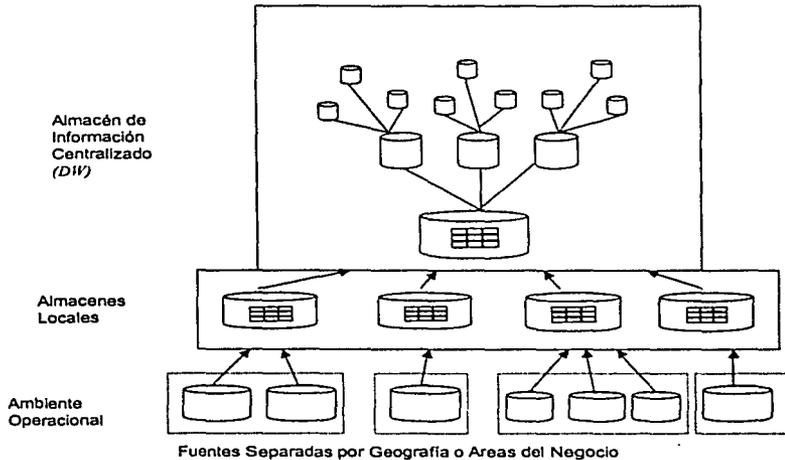


Figura 2.5 Topología de DW Central con fuentes distribuidas

2.6.3 DW Central con Colecciones Finales Distribuidas

Dentro de las variaciones de las topologías del *DW*, una de las opciones más comunes es la separación de la información correspondiente a la

base del *DW* y las colecciones sumariadas o niveles departamentales e individuales (colecciones finales). La razón de esta separación es que en muchas ocasiones conviene que la información departamental o individual resida en otra plataforma e incluso en otra ubicación geográfica. Esta topología propone que la información base del *DW* permanezca centralizada, pero las colecciones que de ahí se generen se ubiquen de manera distribuida. La parte de las fuentes operacionales puede ser desde una misma plataforma o de diversas, pero las colecciones se almacenarán en diversos puntos. La ventaja de este esquema es que es muy flexible al usuario final y facilita la utilización de diversos medios de acceso a la información del *DW*, sin embargo es más complejo su desarrollo y administración. La figura 2.6 muestra un esquema de colecciones finales distribuidas. En este esquema las fuentes están distribuidas también pero es válida la combinación de fuentes centralizadas.

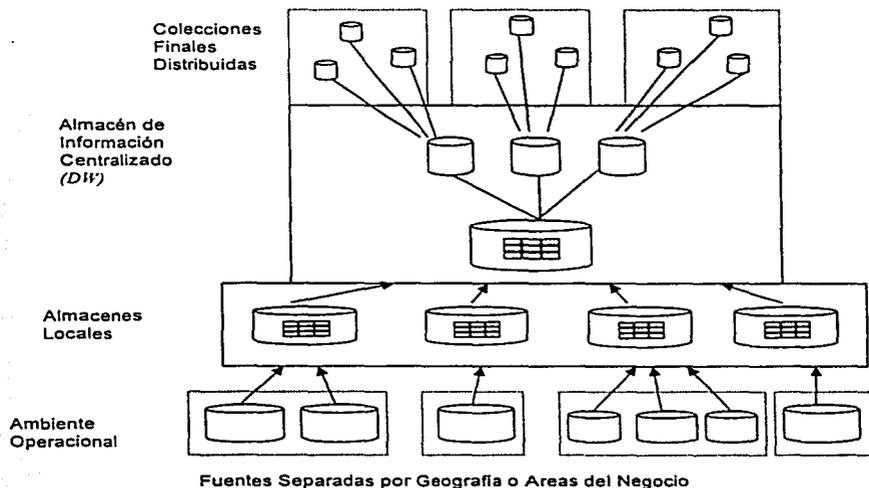


Figura 2.6 *DW* Central con fuentes y colecciones finales distribuidas

2.6.4 DW Distribuido sin Base Central

La siguiente topología corresponde al de un *DW* completamente distribuido. Esta opción es recomendable cuando la naturaleza del negocio sea distribuida al cien por ciento. Aquí el detalle está en construir un *DW* por cada localidad dentro de la organización distribuida, para después interconectar dichos *DW*'s. La conexión de los almacenes de cada localidad, hace complejos el acceso y administración de este esquema. La figura 2.7 ilustra esta topología.

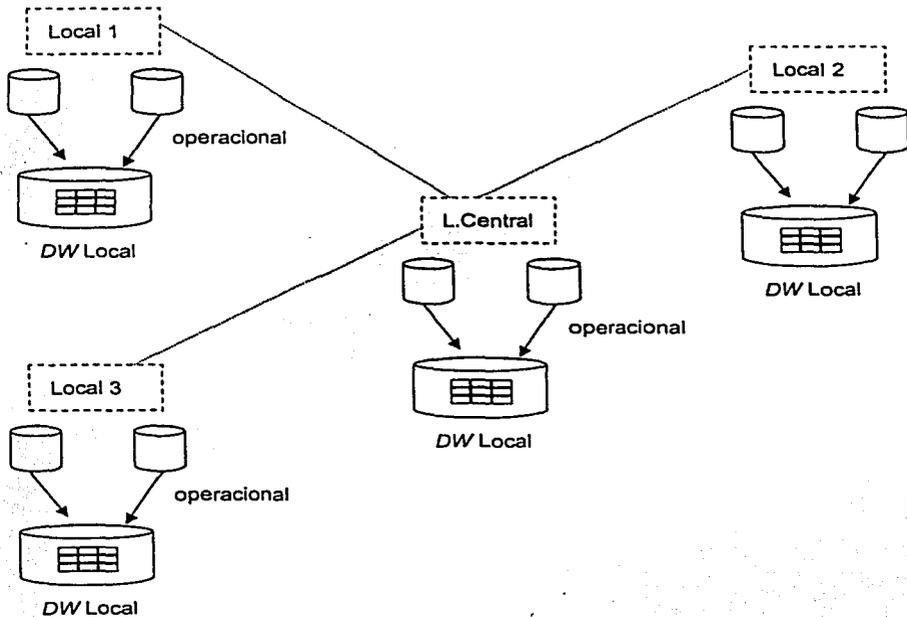


Figura 2.7 DW Distribuido

2.7 Minería de Datos (*Data Mining*) y DW

La Minería de datos es un concepto muy reciente, que está tomando mucha fuerza y que propone técnicas para el análisis de información para encontrar ventajas competitivas en la organización.

Una definición de este concepto es:

" La Minería de Datos es el proceso de descubrir correlaciones, patrones y tendencias, a través del análisis de grandes cantidades de datos, almacenados en repositorios, usando tecnologías de reconocimiento de patrones, así como de técnicas matemáticas y estadísticas"

Dentro de las técnicas que pueden ser aplicadas durante la Minería de datos se encuentran las de visualización, inducción, estadísticas y redes neuronales.

La Minería de datos propone cinco pasos principales como metodología:

- Selección y preparación de Bases de Datos
- Ordenamiento y clasificación
- Selección de herramientas
- Hipótesis, pruebas y descubrimiento de conocimiento
- Aplicación del conocimiento

La selección y preparación de los datos consiste en identificar la información necesaria para nuestro análisis, así como de buscar la forma de accederla. Esto implica obtener información, muchas veces, de diversas fuentes y diferentes dispositivos, lo que hace compleja la actividad. La preparación significa el limpiar los datos y verificar que los datos estén completos y que sean coherentes.

El ordenamiento y clasificación crea vistas de esta información para facilitar nuestro análisis por las bases de datos.

Se requiere de herramientas específicas para la explotación del *DW*. La selección de éstas debe hacerse con cuidado, ya que aún las herramientas de análisis no son maduras y presentan problemas de escalabilidad, es decir, funcionan bien en ambientes pequeños pero presentan dificultades para crecer. Este tipo de herramientas son básicamente para la toma de decisiones que contempla la generación de hipótesis, pruebas y conocimiento. Es en este sentido en que la Minería de Datos ayuda, permitiendo conformar nuevas relaciones e ideas. Con los resultados obtenidos con la Minería de datos, se puede aplicar el análisis de tipo arriba-abajo (*top-down*), abajo-arriba (*Bottom-up*) y el de "que pasa si ..." (*what-if*).

El resultado del análisis frecuentemente lleva a la obtención del conocimiento, mismo que se aplica en reglas o recomendaciones para el analista de este proceso.

La relación entre el *DW* y la Minería de Datos es complementaria y se hace muy notable en los dos primeros pasos de la metodología, que corresponden a la selección y preparación de bases de datos, así como de su ordenamiento y clasificación.

El *DW* proporciona una base de datos integrada con información de la organización, misma que puede ser explotada con técnicas de Minería de Datos. Dentro de los factores críticos tenemos conseguir la información y asegurarse de que ésta es confiable. Por su parte, el *DW* presenta información de donde se puede generar una colección de datos para que la Minería de Datos logre su objetivo.

Para el DW, la Minería de Datos es una aplicación más de análisis y no existe impacto en el diseño del DW.

2.8 Pasos para la Elaboración de un DW

En este punto daremos una serie de puntos generales por los que se debe pasar en la construcción de un DW. Estos pasos, se proponen de forma muy global y pueden cambiar de acuerdo a las necesidades o tamaño de una organización así como de la calidad en la que queremos armar el DW.

- Identificar las necesidades del usuario final. Definitivamente nuestro almacén debe iniciar por medio de una necesidad del usuario final. Esta actividad es muy importante considerando que el éxito de este proyecto depende de qué tan bien entendemos la problemática. Para este punto se sugiere el uso de técnicas como la de sesiones de usuarios en reuniones de diseño de aplicaciones (*JAD Joint Application Design*), las cuales son muy dinámicas y aseguran el entendimiento de las necesidades del usuario. Con estos requerimientos identificados, podremos elaborar el modelo de datos del DW, el cual representa la información que nuestro almacén contendrá.
- El segundo paso, corresponde a la búsqueda de las funciones necesarias para alimentar el DW de acuerdo al modelo de datos desarrollado. Debemos encontrar la mejor fuente existente, lo que implica que sea la más exacta y oportuna, ya que debemos recordar que dependiendo de la calidad de la entrada, tendremos nuestra salida (si entra basura, basura sale). Este punto está relacionado con la función de adquisición.
- Identificar las reglas de negocio necesarias para producir información de análisis. Una vez que encontramos las fuentes, debemos asegurarnos que entendemos lo que el usuario espera, es decir, típicamente, la información que

va de la operación al *DW* debe sufrir transformaciones para que sea mejor aprovechada por los usuarios. Este tipo de transformaciones debe ser muy claro por todos los participantes, de modo que la información contenida ofrezca confianza al usuario.

- Generar las bases de datos físicas que soportarían el *DW*. De acuerdo al modelo de datos y después de haber revisado las fuentes y reglas del negocio (estas actividades pueden cambiar el modelo) se inicia el proceso de generación de bases de datos. La construcción de tablas, índices y otros objetos, sobre algún manejador (típicamente relacional) debe contemplar el hecho que el *DW* es una base de datos para análisis y toma de decisiones y la forma en que se realiza su dueño es diferente a la de una base de datos para una aplicación operacional.
- Generar los procedimientos para llevar la información de la operación al *DW*. Esta actividad puede convertirse en un fuerte problema si se descuida. Es recomendable utilizar herramientas automatizadas. Inmon [Inmon, 1994], comenta que el 80% de esfuerzo en la construcción del *DW* corresponde al desarrollo de los procedimientos para alimentar el *DW* de la operación. En el mercado existen herramientas para el copiado de archivos o herramientas que generan código para traslado y transformación de datos. Se debe considerar que algunas veces se deben construir procedimientos para alimentar el *DW* por primera vez y otros para la actualización del almacén. Al término del desarrollo de esta fase tendremos todo para poblar el *DW*.
- El acceso al *DW* constituye otro paso muy importante dentro del entorno *DW*. Actualmente existen diversas herramientas de acceso para usuario final en ambiente gráfico, para aplicaciones de análisis y toma de decisiones. El *DW* puede consultarse tanto por aplicaciones de tipo genérico como *Query Tool* (herramienta de consultas abiertas), *DSS's* (Sistemas de toma de decisiones), *EISS's* (Sistema para ejecutivos), accesos por Internet, etc., así como

herramientas de propósito específico, como rentabilidad, de estudio de mercado, estudio de clientes, etc. La selección o desarrollo del mejor sistema de acceso para usuario final repercute fuertemente en el éxito del *DW*.

- Implantación de los mecanismos de administración, operación, soporte y mantenimiento del *DW* son indispensables para darle continuidad en el servicio y lograr los objetivos de apoyar a los usuarios analistas de la organización.

Algunos expertos comentan que iniciar un proyecto de *DW* llevaría de 8 a 10 meses de trabajo con un equipo de 10 personas aproximadamente, pero definitivamente depende del tamaño del proyecto y la calidad que se tenga en cada función que se desarrolle. Lo importante al respecto es elegir un proyecto inicial que aporte beneficios y no sea muy grande. El *DW* implica un proceso continuo, a largo plazo y caro, por lo que se debe planear el primer esfuerzo de modo que muestre las ventajas e inspire confianza a la organización para continuar con este proyecto.

Capítulo 3 Implantación del *DW* en una Institución de Crédito

3.1 Descripción de el Banco (Institución de Crédito).

En la Institución de Banca Múltiple donde se realizó el caso, a partir de este momento la llamaremos el Banco.

El Banco fue fundado en 1924, contaba con tan solo 28 empleados y un capital de \$500,000.00. A partir de ese momento inició su crecimiento y evolución, llegando a ser una de las principales empresas en su género.

Uno de los factores que han apoyado este crecimiento es el uso de la tecnología para el control y procesamiento de su información y esto se puede notar en el alto volumen de operaciones que actualmente controla en forma automatizada.

En 1982 el presidente de la república, el Lic. José López Portillo, decretó la nacionalización de la Banca, en un momento en que el país pasaba por un reajuste económico muy serio. Durante casi 10 años la administración de la Banca estuvo en manos del Gobierno. En el año de 1991, el Gobierno decidió desincorporar a la Banca de su administración, generando así un proceso de reprivatización. A partir de esta nueva administración, se forma un Grupo Financiero que incluye compañías de Arrendamiento, Almacenadora, Factor de Capitales y Casa de Bolsa.

Actualmente el banco se define como una institución de Banca Múltiple, ya que maneja una gran variedad de servicios financieros, haciendo de esta empresa una de las más completas e importantes de México. Dentro de los productos que ofrece se tiene: cuentas de cheques, tarjetas de crédito, préstamos

hipotecarios, créditos comerciales, Mercado de Dinero y Capitales, etc. y ofrece otros servicios como el de Banca por Teléfono, Servicio de transacciones por Cajero automático, Servicio electrónico de transacciones desde la casa o empresa, etc.

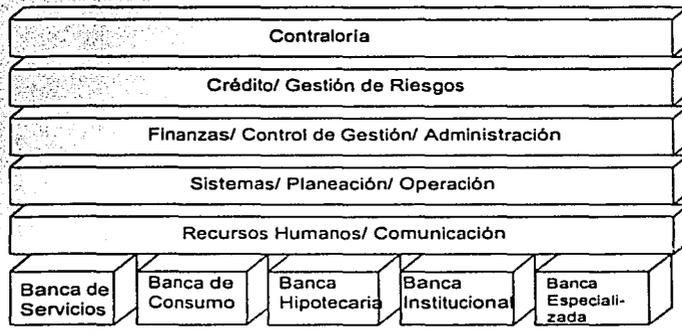
A pesar de los fuertes problemas económicos que sufre en general el país, desde diciembre de 1994 a la fecha, el Banco se ha mantenido como uno de los bancos más fuertes en América Latina. Actualmente este banco cuenta con más de 1000 sucursales ubicadas en todo el país. Su red propia de cajeros automáticos se extiende por toda la república, logrando tener a más de 1,700. Tiene oficinas en diversas ciudades del mundo como: Los Ángeles, Nueva York, Londres y Gran Caimán. Sus clientes son del orden de más de 4 millones en cuentas de captación (cheques, mercado de dinero y capitales, inversiones, etc.), más de 2.3 millones de clientes en tarjetas de crédito y aproximadamente 8.5 millones de afiliados al SAR (Sistema de Ahorro para el Retiro).

3.1.1 Organización del Negocio

La nueva administración ha estructurado a la empresa en cinco bancas, las cuales trabajan como sub-bancos enfocados a diversas secciones del mercado, es decir cada sector del mercado será atendido por una Banca, aunque los servicios pueden ser generales en todas las bancas o particulares para alguna en especial. A la forma de estructurarse se le llama "Modelo Estratégico" Existen otras áreas que funcionan como soporte en las funciones de la empresa. En la figura 3.1 muestra el diagrama del modelo estratégico organizacional implantado.

Este modelo define a cinco Bancas que a su vez agrupa unidades de negocio y unidades de producto y también define cinco estructuras funcionales que apoyan a toda la corporación.

Estructuras
Funcionales



Unidades de Negocio y de Producto

Figura 3.1 Modelo Estratégico

La siguiente tabla (tabla 3.2) , muestra las relaciones de cada Banca con sus unidades de negocio y productos.

Banca de Servicios	Banca de Consumo	Banca Hipotecaria	Banca Institucional	Banca Especializada
Red de Sucursales	Tarjeta de Crédito	Crédito individual	Banca Corporativa	Tesorería
Créditos personales y Microempresas	Tarjeta de Débito	Promotores	Banca de Gobierno	Cambios
Crédito Agropecuario	Crédito al Consumo	Financiamiento Inmobiliario comercial	Banca Empresarial	Fiduciario
	S.A.R.		Productos Internacionales	Banca Patrimonial
	Banca Electrónica		Cash Management	Arrendadora
	Crédito Automotriz			Banca Electrónica
				Factoraje
				Casa de Bolsa
				Red Internacional

Tabla 3.2 Bancas y sus Unidades de Negocio y Producto

3.1.2 Organización del área de Sistemas

El área de Sistemas, buscando satisfacer al negocio de la mejor manera, se ha estructurado tomando como modelo la organización del negocio, y de la misma forma se han creado las funciones de ejecutivos de cuenta, quienes son responsables de atender las necesidades de los funcionarios del negocio que tienen a su cargo alguna Banca. La figura 3.3 muestra la estructura del área de Sistemas (el nombre completo es: Sistemas, Planeación y Operación).

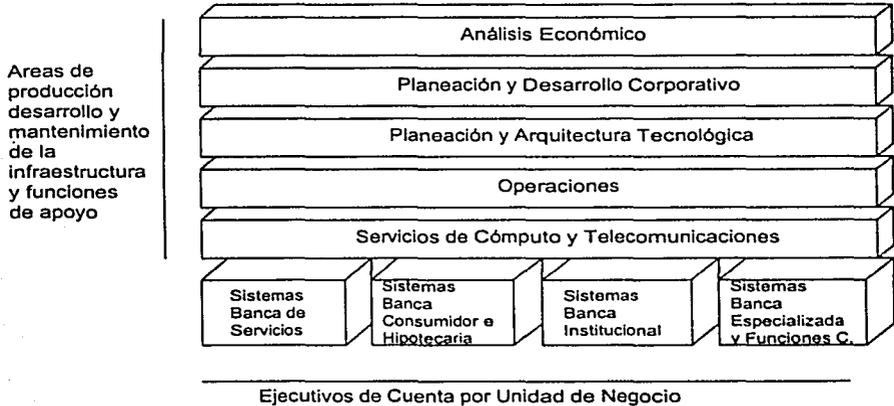


Figura 3.3 Estructura del área de Sistemas, Planeación y Operación

Como vemos en la figura anterior, existen también áreas de apoyo en las funciones de Sistemas, las cuales apoyan a las funciones de planeación e infraestructura a lo largo de la institución.

3.1.3 Infraestructura Tecnológica

La empresa cuenta con una de las infraestructuras más complejas del país, tiene una red de comunicaciones privada que abarca toda la república mexicana, cuenta con un centro de procesamiento centralizado para aplicaciones de cobertura nacional localizado en el Distrito Federal y el cual se compone de tres equipos centrales. Adicionalmente cuenta con varios equipos medianos de la familia de los *Risc* y *AS/400* para aplicaciones departamentales y con más de 20,000 computadoras personales conectados a la infraestructura en una red interna con tecnología *Ethernet*.

En las sucursales existen redes locales que a su vez se interconectan a la red nacional de la empresa, permitiendo el acceso a los equipos centralizados.

La infraestructura de cómputo soporta en promedio 70,000,000 de transacciones al mes dentro del país y en horas hábiles se tienen hasta 300 transacciones por segundo. Para darnos una idea de la magnitud de la infraestructura, la tabla 3.4 muestra parte del equipo instalado.

Procesadores	3 equipos IBM serie 9000
Capacidad de Proceso	734 MIPS por equipo
Terminales conectadas al equipo central	16,500
Capacidad de almacenamiento	Más de 1,500 Gigabytes (> 1.5 Terabytes)
Computadoras personales (No incluidas las sucursales)	8,500

Tabla 3.4 Algunos datos sobre la infraestructura

3.2 Necesidad de Sistemas de Análisis

En los últimos meses y debido a la evolución de necesidades de la empresa, se ha requerido contar con sistemas de análisis para apoyar la toma de decisiones, para Ejecutivos. Estos sistemas son con fines específicos como Estudio de Mercado (*Marketing*) o de costos.

Como podemos intuir, estos sistemas requieren de información que se genera en diversas aplicaciones operacionales.

Estos sistemas son requeridos básicamente por los niveles tácticos y estratégicos de la institución, y los ejecutivos de cuenta de la parte de Sistemas se han avocado a satisfacer estos requerimientos.

La forma en que comenzaron a resolver estos sistemas fue con la creación de aplicaciones intermedias que funcionan con extractos, es decir, se copian los archivos maestros de las aplicaciones operacionales como cheques, bursátil, préstamos, etc. y se almacenan para dar información de los ejecutivos del negocio.

Este método sirve muy bien cuando se resuelven pocos requerimientos de un departamento, pero si hablamos de escalabilidad, o dicho de otra forma, si pensamos en sistemas corporativos, ésta solución no es flexible y puede traer otros problemas.

Para entender mejor lo que sucede en el Banco, veamos algunos ejemplos.

3.2.1 Sistemas de análisis del Banco

En el Banco existen varios sistemas intermedios (ya sean de toma de decisiones, ejecutivos o de propósito específico) como el BCU (Sistema de Centros de Distribución), BOS (Sistema de Gestión Operativa en Sucursales), BGO (Sistema de Gestión Operativa de Cartera), BNI (Sistema de Medición de Incentivos), etc. Cada uno de estos sistemas, como lo veremos a continuación, está dirigido a un segmento diferente de la organización, y por lo tanto, es administrado por diferentes ejecutivos de cuenta del área de Sistemas. Estos sistemas son desarrollados sin tomar en cuenta los existentes.

En la figura 3.5 vemos un diagrama del sistema de BCU

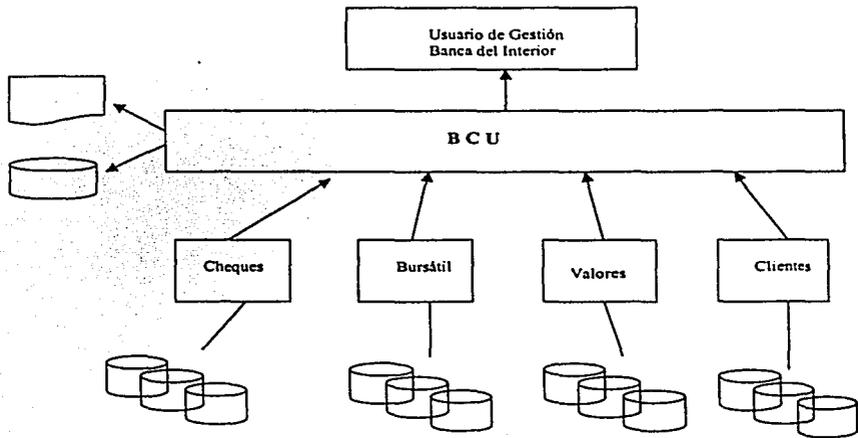


Figura 3.5 Esquema de Sistema Intermedio BCU

En este sistema vemos que las fuentes de información son los sistemas operacionales de cheques, bursátil y clientes y está enfocado a un departamento en especial.

La figura 3.6 presenta el diagrama para el sistema BGO .

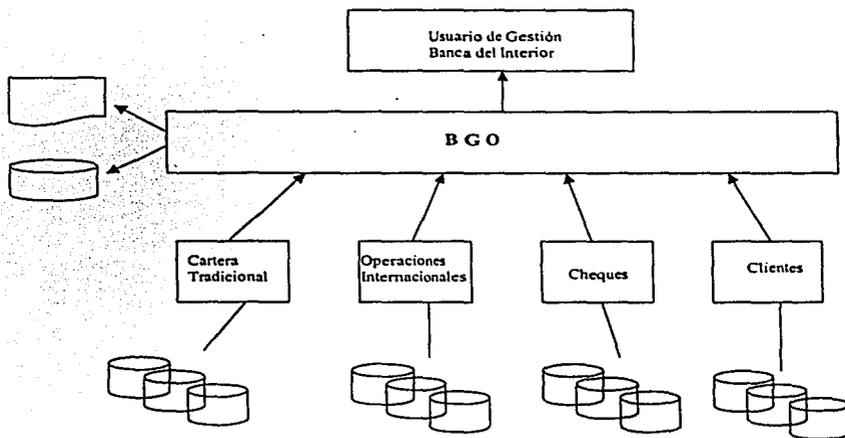


Figura 3.6 Esquema del Sistema Intermedio BGO

Como vemos, este sistema, que está enfocado a otro departamento gerencial, tiene de entrada a cheques, entre otros. Esto provoca que la aplicación de cheques tenga que generar varias interfases o extractos.

La figura 3.7 muestra el diagrama de algunos de los sistemas intermedios que existen en el banco en relación con los sistemas operacionales.

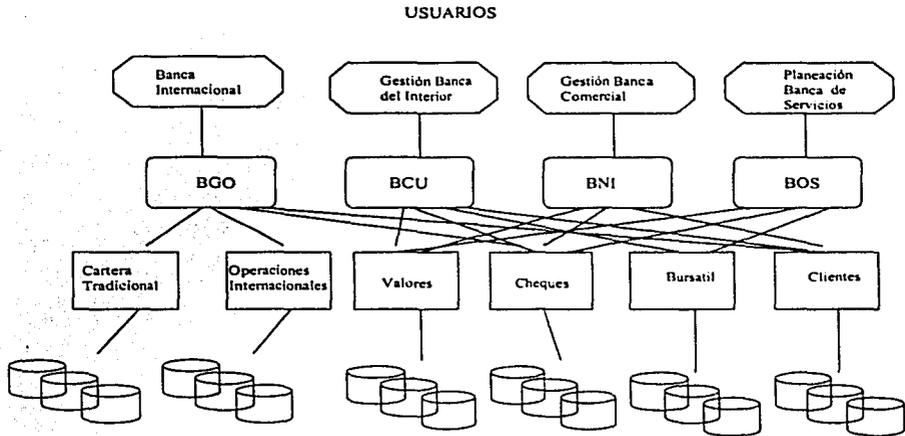


Figura 3.7 Diagrama de 4 Sistemas intermedios y su relación

El diagrama de la figura 3.7 es una pequeña porción de la duplicidad y redundancia de datos que se veían obligados a generar las distintas aplicaciones.

3.2.2 Problemática del Banco en sistemas de análisis

Los problemas que se presentan en el Banco bajo este entorno son los siguientes:

Algunas aplicaciones operacionales se encuentran saturadas de trabajo, y la generación de interfaces hace que se reduzcan sus ventanas de tiempo

en procesos. Dichas ventanas se usan para dar mantenimiento a los sistemas.

La generación de diferentes extractos no garantizan resultados confiables, ya que depende en mucho del manejo y control que se le dé a la información en los sistemas intermedios.

Para aplicar alguna modificación a un sistema operacional, se debe coordinar con diversos ejecutivos de cuenta para no impactar a los sistemas intermedios.

El tiempo de respuesta a los requerimientos de modificaciones es muy alto, ya que se involucran varias áreas para hacer el ajuste. Por ejemplo, si en el esquema de BCU, figura 3.5, se modificara el sistema para añadir más información, el ejecutivo de cuenta tendría que coordinar esfuerzos de los grupos responsables de los sistemas de Cheques, Bursátil, etc.

Estos sistemas son cerrados: si el ejecutivo de negocio necesita información de otro sistema intermedio, no hay forma de conectarlos y tendrá que esperar a que incorporen dicha información a su sistema.

Existe información duplicada sin control y de la misma forma esfuerzos innecesarios, además que los resultados de los sistemas intermedios rara vez cuadran con los otros sistemas, ya que los criterios para los procesos de obtención de información son diferentes.

3.3 El DW del Banco

El DW se planea en el Banco como parte de la arquitectura de información del negocio. De hecho, la primera fase (caso de nuestro estudio) se desarrolla a cargo del área de arquitectura tecnológica.

Dentro de la arquitectura de información del Banco, nace la necesidad del almacén de información para toma de decisiones. Esta arquitectura, desarrollada por el área de Planeación y Arquitectura, se presenta en la figura 3.8.

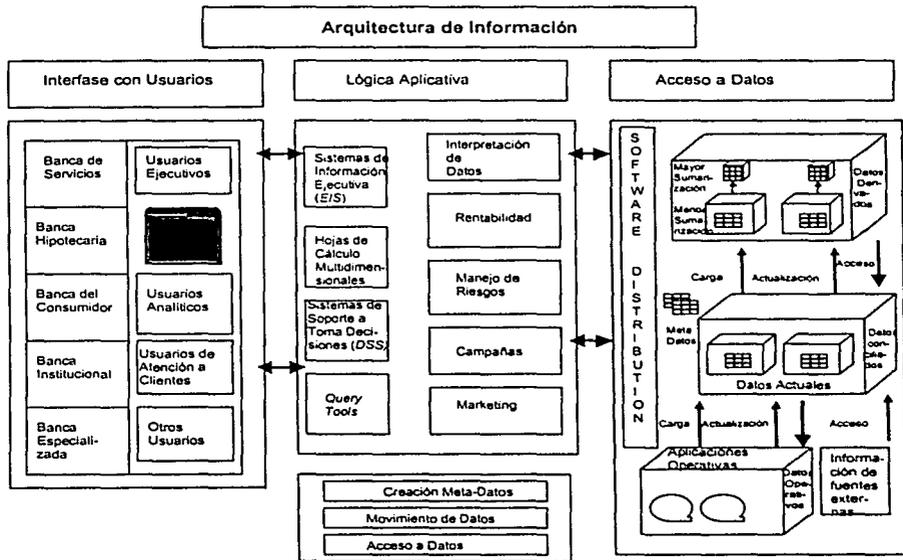


Figura 3.8 Arquitectura de Información del Banco

La arquitectura de Información queda dividida en tres grandes partes: Usuarios, Lógica Aplicativa y Datos. En la parte de usuarios están representadas las cinco Bancas de la organización mostrando el carácter corporativo de esta

arquitectura. Dentro de la Lógica Aplicativa se presentan las aplicaciones para análisis, por una parte los sistemas de carácter general como son los sistemas de consulta (*Query Tools*), los de toma de decisiones (*DSS*), etc. y por la otra parte los sistemas que cubren un caso particular como los de análisis del mercado (*marketing*), o de riesgos. Por último, se define la sección de los datos que presenta tres tipos de información correspondiente a la operacional, la departamental (conciliada) y la individual (derivada). Es conveniente mencionar que dicha arquitectura, con pequeños cambios, es la que actualmente rige la información del Banco.

El *DW* entra para apoyar la estrategia de los datos, y para este caso de estudio, se completaron también las otras secciones de la arquitectura de Información del Banco.

El alcance que se le da al *DW*, comprende información de toda la institución que apoye los sistemas de análisis. Esto como se puede apreciar, es un proyecto muy grande y a largo plazo que se estima tomará 3 años para terminar el 80% de la construcción y explotación del *DW*.

Dada las dimensiones del proyecto, se empezó con una fase relativamente sencilla, pero que da grandes ventajas por ser un tema que el 100% de los clientes tienen, cuentas de cheques.

Cheques son cuentas de depósito a la vista. Algunas de ellas pagan intereses y otras no, algunas manejan chequeras y otras no, y todas están ligadas a servicios alternos como Cajeros Automáticos, además de que el número de la cuenta es el número único de servicios y productos del Banco.

Para esta primera fase, se definió como alcance el incorporar la información de todas las regiones a nivel nacional de cheques (BCH) y algunos sistemas relacionados como la base de datos de clientes (BKL).

El sistema de cheques (BCH) es el principal sistema operacional de cheques y maneja información de diferentes tipos de cuentas de cheques. Así mismo es el sistema con mayor transaccionalidad (70 millones mensuales de transacciones aprox.)

El sistema BKL es la base de datos de clientes que maneja aproximadamente 8 millones con todos sus datos demográficos y el detalle de sus cuentas/servicios que tiene con la empresa.

Todos estos sistemas operacionales que intervienen, residen en el equipo central y el *DW* residirá inicialmente también en dicho equipo.

3.4 Plan del Proyecto

La planeación de la primera fase del *DW* tomó dos meses y consistió en capacitarse en el concepto, buscar el proyecto inicial para el *DW*, desarrollar el plan tecnológico y armar el equipo de trabajo.

Una de las políticas más importantes para el inicio del *DW*, es la de no impactar resultados actuales o necesidades inmediatas, es decir, el *DW* en esta fase no tiene la intención de sustituir a algún otro sistema ni de parar algún esfuerzo, la estrategia es la de formar el almacén de información y después hacer re-ingeniería con los sistemas intermedios actuales.

Para seleccionar el proyecto inicial, se tomo en cuenta la necesidad del negocio, el entusiasmo del ejecutivo del negocio y por lo tanto su apoyo (siendo este punto importante, ya que el ejecutivo del negocio es el patrocinador de esta fase). De esta forma se seleccionó iniciar con información de Cheques.

Con esta información se ha logrado que los usuarios finales tengan a su disposición información sobre el cómo se comporta Cheques en el Banco, permitiéndoles dar seguimiento y comparando entre diversos puntos del tiempo, de esta forma se presenta un panorama global, y es factible llegar al detalle en caso de ser necesario.

El *DW* fase inicial, almacena detalle de cuentas de cheques y después genera sumariación a nivel organizacional (por funcionario, sucursal, centro regional, etc.), o también por productos (tipos de cuentas de cheques), por tipos de intereses, etc. En esta organización de los datos permite que el usuario obtenga diversas vistas de la información, y al mantener el detalle de los datos, hace posible el elaborar nuevas vistas que no se habían contemplado.

La información Histórica que permanece en el *DW*, comprende de 24 meses de información detallada diaria (según lo pactado con el usuario final).

La planeación del proyecto se basó en secciones de seguimiento (Tracks). Estas son:

- Administración del proyecto
- Modelo de Datos
- Usuarios Final
- Sistemas Fuentes
- Plan Tecnológico
- Herramientas del Usuario
- Operaciones
- Soporte

Las actividades dentro de cada sección que se realizaron en el desarrollo de la fase uno del *DW* se presentan a continuación. La forma en que se presentan las actividades no tiene un orden cronológico en el cual se

desarrollaron, están ordenadas por secciones. El desarrollo de estas actividades se realizó en paralelo en ciertos puntos y con diferente responsable por sección.

Sección de Administración del Proyecto

Definición del Plan Global del Proyecto. El ejecutivo de cuenta, líder del proyecto inicia la fase, coordina a los grupos de trabajo y elabora un plan macro del proyecto.

Presentación ejecutiva a usuarios. El proyecto inicia con la presentación a las direcciones usuarias con lo que se da arranque al proyecto.

Definición del alcance. En base al análisis del modelo, requerimientos y fuentes, se define el alcance del proyecto.

Revisar el plan basado en el alcance. El ejecutivo de cuenta revisa el plan y le da mayor detalle de acuerdo al alcance.

Revisar el plan del *DW* para implantación. El ejecutivo revisa que todo este listo para el uso en producción de la fase del *DW*.

Monitorear y revisar actividades. Esta actividad es continua durante el proyecto y trata de asegurar que el plan se este cumpliendo adecuadamente.

Liberación completa de fase del *DW*. En este punto es cuando se libera el uso de la fase del *DW* a los usuarios en ambiente de producción.

Sección de Usuario Final

Obtener requerimientos de Usuarios Finales. El usuario define sus requerimientos generales. Esta actividad se realiza con entrevistas para conocer sus necesidades.

Clasificar las necesidades de usuarios. El Ejecutivo de cuenta da forma a los requerimientos para definir el objetivo del sistema

Analizar los *DSS/EIS* existentes. Se revisa si existe alguna solución previa que pueda generar consideraciones al desarrollo.

Definir colecciones de datos del *DW*. Se implementan las colecciones de datos en la plataforma seleccionada.

Generar el código de las colecciones y probar. En este punto, con herramientas de consulta se generan las consultas para formar las colecciones del *DW*.

Sección del Modelo de Datos

Seleccionar los tópicos para el diseño del modelo *DW*. Inicia con el análisis de datos para el modelo de datos.

Definir el Modelo Lógico del *DW*. Se elaboran los diagramas de entidad relación.

Definir el Modelo Físico del *DW*. Se construyen las entidades físicas del modelo.

Construir Bases de Datos. Se hace la construcción física del modelo de datos y de los objetos de la base de datos.

Construcción de modelo de datos de colecciones. En este punto se modelan las colecciones departamentales y de usuario, esta actividad no involucra

a grupos centrales ya que el ejecutivo de sistemas decide la plataforma y forma de dar los datos.

Construir Base de Datos de colecciones. Se construyen las colecciones que presentará al usuario.

Desarrollar proceso iterativo de colecciones y vistas. Buscar mecanismos para ser dinámicos en la formación de las colecciones.

Sección de Operaciones

Solicitar requerimientos de las áreas técnicas. De acuerdo con lo avanzado del proyecto, el Ejecutivo convoca reuniones con las áreas involucradas para detectar sus requerimientos y para prepararlas para la entrada de este nuevo sistema.

Análisis de Actividades de apoyo. De las reuniones con las áreas técnicas, se catalogan los requerimientos y responsabilidades para darle seguimiento, por ejemplo alta al sistema de los catálogos.

Desarrollar plan de operación. En ese momento se realiza el plan que apoye las actividades respecto a la parte de operaciones como el planear el cambio estratégico, etc.

Adecuar el plan de la capacidad. Se solicita el espacio en disco necesario y se revisa capacidad de equipo de cómputo.

Inicia etapa de pruebas. En este punto los usuarios deberán estar probando su información en las colecciones, a partir del DW que se ha estado cargando en pruebas.

Iniciar uso en producción del *DW*. Empezar el *DW* en producción normal.

Sección de Sistemas Fuentes

Investigar las fuentes candidatas. Con el resultado del análisis de datos, se revisan que sistemas (operacionales) sirven para captar la información requerida.

Analizar alimentadores existentes (fuentes). De los sistemas seleccionados se revisan los archivos o tablas con la información requerida.

Investigar las condiciones de los datos. Se revisan las fuentes propuestas junto con el Ejecutivo para tomarlas como fuentes válidas (esto retroalimenta al diseño físico).

Documentar mapeos y transformaciones *DW*. El equipo de *DW* trabaja en la parte de llevar los datos y cargarlos al *DW*.

Generar código de adquisición y prueba. Trabajo bajo alguna herramienta o no. En este caso se uso la herramienta llamada *PRISM* y que es un generador de código entre otras funciones.

Preparar fuentes candidatas. Se solicitan las fuentes que utilizará el *DW* a las diversas aplicaciones operacionales.

Utilizar fuentes para carga del *DW*. Se hacen las pruebas del proceso de adquisición de la información.

Poblar el *DW*. A partir de este momento se comienza a cargar el *DW* (antes de la liberación).

Sección de Soporte

Definir plan de soporte. Es importante contar con un plan que refleje el apoyo al desarrollo y buscar responsables en éstas actividades, planear reuniones, capacitación, documentación, etc.

Definir plan de Capacitación de herramientas. Definir cómo capacitar al usuario en el uso de las herramientas de acceso y explotación.

Definir plan de Documentación. Realizar la documentación del sistema y del área de operación de éste.

Implantar mecanismo de soporte. Se debe planear la forma en que el usuario recibirá el soporte en el uso del sistema.

Sección del Plan Tecnológico

Elaborar Premisas Tecnológicas. Este plan se inicia con una lista de lo que se busca en cuanto a infraestructura tecnológica, qué se tiene y posibles productos a evaluar.

Revisar Arquitectura Tecnológica. Se verifica que exista la Arquitectura Tecnológica y tenga sentido con las premisas.

Definir plataforma del *DW*. Se busca la plataforma del *DW*, en base a la Arquitectura Tecnológica.

Diseñar Infraestructura. Se seleccionan y adquieren los productos que integran la infraestructura que embone dentro de la Arquitectura.

Implantar la Infraestructura. Se instala la infraestructura del *DW* y comienza un proceso continuo de mantenimiento a ésta.

Sección de Herramientas del Usuario

Definir requerimientos. Identificar requerimientos de usuario respecto a herramientas de acceso.

Investigar herramientas. Se evalúan herramientas de acceso en el mercado.

Seleccionar herramienta. Se selecciona la o las herramientas de acceso.

Elaboración de prototipo. Construcción del prototipo para mostrar al usuario final y obtener su aprobación.

3.5 Plan Tecnológico del *DW* del Banco

En el plan tecnológico se define cómo se desarrolla la infraestructura de cómputo en base a la arquitectura.

En este punto se definen actividades como dónde va a residir el *DW*, con qué herramientas se va a administrar, cómo se le dará mantenimiento, cómo se va a acceder, con qué herramientas se va a consultar, etc.

Dentro del plan tecnológico del *DW* del Banco, se definió lo que a continuación se describe.

3.5.1 Topología

La topología que se eligió para esta primera fase es la Central-Única. En esta fase inicial es conveniente por sencilla, además de que las fuentes que alimentan al *DW* están en la misma plataforma del *DW*. El acceso se hace utilizando una herramienta puente (*gateway*) para acceder los datos del equipo central.

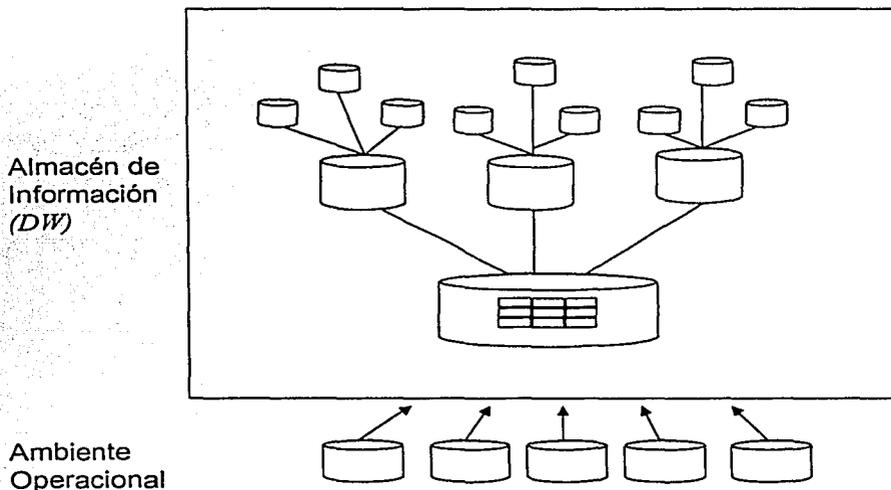


Figura 3.9 Topología *DW* Central-Único

3.5.2 Infraestructura para el DW

La tabla 3.10 muestra la infraestructura que se va a utilizar, la forma en que se contruyó la infraestructura fue analizando cada componente de la arquitectura, después se investigó sobre herramientas o productos existentes en el banco que apoyaran este componente, cuando no existía dicha herramienta, se evaluaban y seleccionaban herramientas del mercado.

Plataformas	
CPU Hardware	Equipo Central IBM Serie 9000
Sistema Operativo	MVS
Redes	SNA, TCP/IP
Administrador de Archivos y Bases de Datos	
Administrador de Archivos	VSAM
Administrador de Bases de Datos	DB2
<i>Middleware Gateways</i>	<i>Net Gateway Sybase</i>
Ambiente de Desarrollo	
Lenguajes de Programación	Cobol (generado con la herramienta <i>Prism</i>)
Manejo de bibliotecas	PDS
Software de Operación	
Manejo de calendario de procesos <i>Batch</i>	UCC11, UCC7
Respaldo y Recuperación	DB2 (Procesos manuales)
Acceso	
Herramientas de Acceso	<i>Business Object – Windows</i>

Tabla 3.10 Infraestructura del DW del Banco – Primera fase

De los productos en la infraestructura, sólo se adquirieron dos que fueron el *Prism Warehouse Manager* y el *Query Tool Business Objects*. El primero auxiliando a la parte de copia de los datos operacionales al DW mediante la generación de programas Cobol, y el segundo se seleccionó para acceder el DW desde un ambiente gráfico.

3.6 Desarrollo

Para el desarrollo de esta fase, se contó con un equipo de desarrollo formado por siete personas de sistemas y cinco personas de los negocios con las siguientes funciones:

- Líder del Proyecto
- Líder Técnico
- Administrador de Datos (Modelo de Datos)
- Consultor de herramientas de acceso
- Analista de sistemas (3)
- Especialistas en datos del cliente (2)
- Especialista en datos de cheques (3)

El proyecto fue coordinado por un grupo con reporte directo a la Dirección General de Sistemas, Planeación y Operación y en el desarrollo participaron personas de las Direcciones de Arquitectura y de Desarrollo de Sistemas para Banca Comercial.

Las actividades que vimos en la sección de planeación fueron desarrolladas por este equipo y tomó 8 meses aproximadamente, teniendo la supervisión y apoyo del grupo de evaluación que se conformó por altos ejecutivos. Estos fueron:

Líder del Grupo especial designado para coordinar el proyecto

Director de Planeación y Arquitectura Tecnológica
Director de Sistemas de Banca Comercial (Ejecutivo de Cuenta de
Sistemas)
Director de la Banca Comercial (Usuario líder)
Director del área de Control de Gestión de Finanzas

Se tomó la premisa de usar herramientas automatizadas para el desarrollo, evitando así la elaboración del código. Se utilizó un CASE llamado *ADW* para el diseño de la Base de Datos DB2; el producto de *Prism Warehouse Manager* para la generación automática de programas *Cobol*, así como de los procedimientos de control. Para el acceso a los datos del *DW*, se utilizó *Business Objects*, una herramienta de consulta (*Query Tool*) que permite al usuario final hacer consultas al *DW* manejando gráficos y términos del negocio.

3.7 Implantación del *DW*

A continuación veremos algunos puntos concretos que describen el resultado de implantación de la fase I del *DW* en el Banco. La intención es visualizar cómo se desarrolló y terminó éste.

Se comentarán los puntos que, a nuestro juicio, fueron los más relevantes y que permitieron contar con el *DW* inicial.

Se inició con la definición clara de la necesidad a cubrir en esta primera fase del proyecto, siendo ésta:

Descripción de la necesidad:

Conocer mensualmente cómo se estaban comportando los saldos de captación para poder planear las estrategias de negocio a aplicar en la red de sucursales y poder medir las metas que se fijarían anualmente a cada funcionario.

Con la descripción anterior, se procedió a definir el objetivo, alcance del proyecto, y principales usuarios, puntos de partida para iniciar el modelo de datos que después evolucionó en el *DW*. Estos son:

Objetivo del proyecto:

Lograr darle seguimiento cuantitativo no contable, como fuente única de información para el control de los recursos, productos y mercado objetivo de la institución.

Entendiendo como: a:

Seguimiento:	Registro histórico de sucesos
Recursos:	Unidades de Negocio, funcionarios
Control:	Puntos de evaluación, tomar acciones para modificar el comportamiento.
Mercado objetivo:	Clientes, segmentos, competidores, áreas geográficas.
Productos:	Captación, Colocación, Fiduciario, etc.

Principales Usuarios:
Dirección de Control de

Uso
Analizar información.

Gestión de Finanzas

No toman decisiones.

Fuente única de información para la alta dirección.

Equipo de la Dirección de Banca Comercial

Toman decisiones
Analizan cambios en el comportamiento de los elementos de negocio y características del cliente.

Alcance de la fase 1:

Para fines de la primera fase del proyecto, se delimitó por aplicativo, definiéndose la incorporación de los datos demográficos del cliente y su relación con el banco, y la aplicación de Cheques; esto es, no se consideraron las aplicaciones de otros productos de captación como: valores a plazo, valores mismo día y día siguiente, mercado de dinero, Casa de Bolsa.

Para la construcción del Modelo de Datos del *DW*, se hizo a partir de los modelos de datos de los sistemas operacionales, en otras palabras, se analizaron los modelos de las aplicaciones de "Base de Datos de Clientes" y "Cheques", en los cuales se basó el modelo del *DW* (anexo 1).

El modelo de datos del *DW* generó 103 tablas físicas en DB2, de éstas, 58 pertenecen a la base del *DW* donde se incluyen datos detallados y tablas de referencia (número de sucursal, funcionario, calendarios, etc.). Las 45 restantes son colecciones que fueron definidas también en DB2, aunque en diferente base

de datos. La figura xx muestra un diagrama de los niveles desarrollados en esta fase del *DW* del banco.

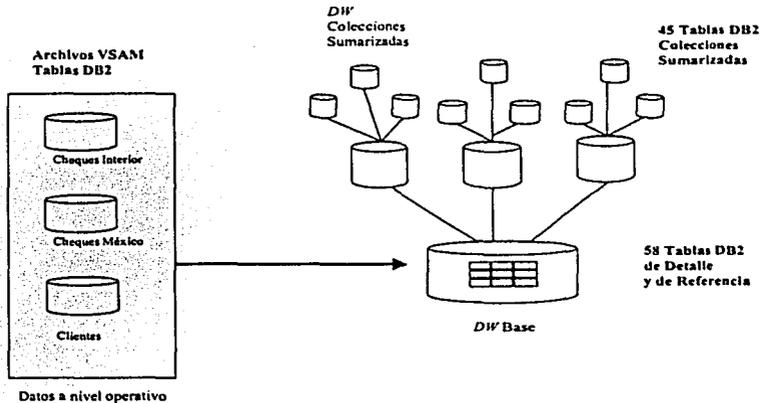


Figura 3.11 esquema general del *DW*

El contenido y manejo de la Base de Datos son actualmente los mismos que presentaba el sistema cuando se elaboró este trabajo, por lo que se deprecia sigue siendo vigente.

Dentro de los Datos Básicos, tenemos información a nivel del nombre de cada cliente, su dirección desglosada, condiciones y características del tipo de persona, origen del cliente, actividad, fecha de nacimiento, nivel de estudios, sucursal y funcionario que lo atiende, números de servicios y productos que tiene con el banco. En el caso del producto de cheques se tiene tipo de producto con características y condiciones, saldo diario por cuenta y moneda, detalle de transacciones en cuanto a monto, tipo de transacción y lugar de operación. El periodo de actualización es semanal para datos demográficos y diario en el caso

de cheques. En cuanto al manejo de la historia de este nivel se manejan 13 meses.

En los datos contenidos en las colecciones sumariadas se encuentra información organizada por 4 grandes tópicos y se generan con una periodicidad semana. Estos tópicos son:

- Estructura organizacional
- Tipo de producto
- Cliente
- Cuenta

Cada tópico representa una vista de información facilitando al usuario sus consultas y su análisis. Estas vistas o colecciones semanales son generadas de la información base, de modo que resulta relativamente sencillo modificar las colecciones o generar nuevas según se requiera. La historia que se prepara para este nivel es de 13 meses en las colecciones sumariadas.

El acceso al *DW* se hace de dos formas:

- Consultas y pantallas previamente definidas vía terminal asincrónica por todos los funcionarios de la red de la Dirección de Banca Comercial (1300 sucursales aproximadamente).
- Utilizando la herramienta *Business Objects*, un producto que corre bajo la ambiente *Windows* y permite hacer consultas en un ambiente gráfico. Para ello, se usó un portal de modo que se conectara a la base de datos del DB2 en el equipo central. Este acceso es utilizado por las áreas de apoyo de la Dirección de Banca Comercial y por la Dirección de Control de Gestión de Finanzas para análisis de información haciendo consultas no programadas.

La siguiente figura muestra la conexión de los productos utilizados en el DW del Banco.

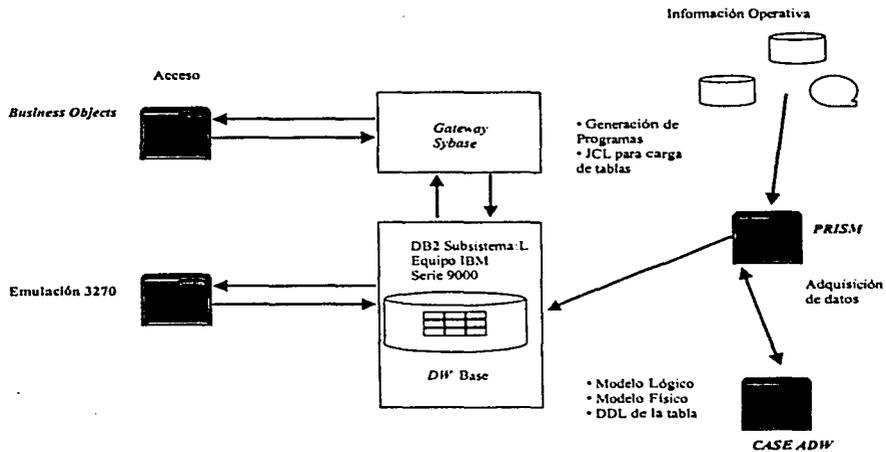


Figura 3.12 Herramientas y Productos usados en el DW del Banco

3.8 Resultados para el Negocio

Los resultados de esta fase del proyecto pueden ser divididos en dos puntos de vista:

- El del área de sistemas
- El del Negocio

Para Sistemas, los resultados se pueden listar de la siguiente forma:

- Se pudo probar el concepto del *DW* y ha convencido en su funcionamiento
- Se probaron y utilizaron herramientas automatizadas para el desarrollo del proyecto.
- Se constató una buena comunicación con el área usuaria
- Se tienen los elementos necesarios para comenzar un plan para la incorporación de más aplicaciones.

Los usuarios que actualmente accesan el *DW* son: directores de áreas de apoyo de la Banca Comercial y de Finanzas y ejecutivos de línea.

El punto de vista del negocio consiste en:

- Se ha establecido una fuente única de información para Cheques (uno de los productos más importantes porque va dirigido a todos los segmentos de clientes y a todas las bancas)
- Los ejecutivos de línea semanalmente pueden revisar como se está comportando su paquete de clientes y compararlo con el de la semana anterior y/o mes anterior, permitiéndoles reaccionar ante cualquier variación importante dentro del mes.
- Las áreas de apoyo, cuentan con una herramienta que les permite obtener información anual para apoyar su análisis en el comportamiento de los saldos por cliente, ejecutivo de línea, sucursal, banca, Banco.
- Brinda apoyo para el otorgamiento de créditos al menudeo (personas físicas).
- La información se puede comparar contra el mercado y otros bancos.

Capítulo 4 **Crítica y propuestas a futuro para el caso DW del Banco**

En el presente capítulo se revisa el caso del *DW* del Banco y se compara con el modelo propuesto del *DW* general.

Para llevar a cabo este estudio, se han tomado como base los conceptos, componentes y funciones vistos en el capítulo tres, para después compararlo con el caso práctico del banco y así generar algunas propuestas para el caso.

A continuación se detallan los conceptos, funciones y componentes que se usarán como base para esta crítica:

- Objetivo del Caso
- Componentes:
 - Información
 - Metadatos
 - Modelo de Datos
- Funciones:
 - Adquisición
 - Administración
 - Operación
 - Acceso
 - Navegación
 - Distribución
- Topología
- Pasos para la elaboración de un *DW*

4.1 Objetivo

Dentro de la problemática planteada, pudimos observar que la falta de un almacén de información único es el principal problema para los sistemas de análisis y de toma de decisiones ya que:

- Las aplicaciones operacionales se encuentran saturadas en cuanto a trabajo.
- La generación de extractos de información no es confiable.
- Es muy tardado la incorporación de nuevas funciones y/o modificaciones y,
- se tiene información duplicada que no permite dar una sola versión del negocio.

Por ello, el objetivo manejado en el caso del Banco respecto a la problemática está bien enfocado al uso de un almacén de información único (*DW*) en la organización.

En este sentido, se logró identificar la problemática y se ha planteado una solución de largo plazo, que contiene la creación de infraestructura de información al seleccionar el *DW* como solución. Dada la complejidad de ésta, se plantea su desarrollo en fases perfectamente alcanzables y acotadas para los sistemas de análisis.

Se parte de una arquitectura de información definida por el banco, y que consiste a grandes rasgos en tres grandes partes o capas: Interfase con Usuarios, Lógica aplicativa y Acceso a datos. La interfase con Usuarios es la capa en donde quedan representados todos los diferentes tipos de usuarios de información, dicho en otras palabras, la organización misma. La capa de Lógica aplicativa presenta los tipos de aplicaciones para análisis, ya sea de tipo general o de tipo específico. Por último la parte de Acceso a datos presenta tres tipos de información: operacional, departamental e individual.

Dicha arquitectura permite visualizar el objetivo y fijar los lineamientos por donde se definirá el proyecto completo, así como el armado de la infraestructura. Contar con una arquitectura permite tener una visión de largo plazo, dándole continuidad a los proyecto del Banco.

4.2 Componentes

Como vimos anteriormente, existen diversos componentes en esta propuesta. A continuación revisaremos la forma en la que fueron aplicados en el caso del Banco.

4.2.1 Información

Dentro de la información organizada en el *DW* y sus niveles operacionales, el caso presentó lo siguiente:

- La mayoría de los sistemas de tipo operacional estaban (y están) ubicados dentro del equipo central, lo que apoyó, de alguna manera, que el *DW* se creara dentro de esta misma plataforma. De igual forma las colecciones se ubicaron en el equipo central lo que permite acceso a nivel nacional en forma rápida y con un mínimo de inversión.
- Es importante reevaluar si el equipo central sigue siendo la mejor opción en cuanto a la plataforma para el *DW* corporativo y sus colecciones, es decir, es importante diseñar cuál es la plataforma ideal para que los almacene. Dentro de las topologías de *DW* existen diversos estados que convendría analizar si la plataforma de un equipo central, que es administrado para aplicaciones operacionales, es la plataforma para un *DW* que fundamentalmente

demandará espacio en disco y ventanas de servicio mayores a lo que requiere una aplicación transaccional.

- En esta fase no se definieron datos de tipo individual, sino que se presentaron solo las capas de la información base y después las colecciones de nivel departamental. Algo interesante es que son casi el mismo número de tablas físicas para los datos sumarizados (45) que para la base (58), y de acuerdo a lo visto anteriormente, las tablas base presentan un modelo que tiende a ser normalizado, mientras que las colecciones son propensas a desnormalizarse. Partiendo de este punto, no suena lógico el tener casi el mismo número de tablas en las colecciones que en la base. Una causa fue el que al analizar la información de colecciones, se decidió formar todas las vistas que el usuario pidió y a partir de esto se formaron las tablas físicas. Es conveniente revisar nuevamente el diseño para las colecciones, tratando de generar las tablas físicas que el usuario realmente utilice.

4.2.2 Metadatos

En la primera fase sólo se contó con un inventario de los objetos de la base de datos como tablas, columnas, índices, etc. y algo de información sobre las fuentes en los sistemas operacionales, pero falta detallar la infraestructura de metadatos y automatizarla, así como incluirle más información como: frecuencia de actualización de los datos, responsable de cada dato, procesos, etc.

Conforme transcurra el tiempo, los metadatos conforman un componente muy importante para el éxito del *DW*. En el caso del Banco, no se detalló este componente lo suficiente. Quizás en el momento en que se terminó la primera fase, no fue crítico el no contar con una definición de metadatos completa, pero para las siguientes fases es necesario no avanzar si no se considera la

creación de los metadatos en forma consistente y formal, para lo cual se pueden explorar herramientas del mercado orientadas a este fin.

Éste es un aspecto de gran oportunidad en el caso del Banco y debe ser atendido lo más rápido posible, no sólo por las grandes ventajas potenciales, sino porque sin el puede haber muchos problemas de administración y de credibilidad para el *DW* del Banco.

Dentro de las complicaciones que se podrían evitar en caso de tener una estrategia y desarrollo claro para los metadatos, podemos mencionar:

- No contar con un inventario de la información que contiene el *DW*
- Un acceso complicado al usuario final al no tener un directorio de datos, lo que provocaría problemas para navegar sobre la información
- Mezcla de conceptos y definiciones de los datos contenidos en el *DW*
- Desconocimiento del nivel de actualización de los datos accedidos
- Desconocimiento del responsable de la información

4.2.3 Modelo de Datos

La forma en la que el Banco decidió comenzar su *DW* tiene algunos puntos a considerar. Por un lado tiene ventajas el hecho de que desarrolló un proyecto dando resultado tangible a corto plazo y probando diferentes conceptos. Esto tiene grandes ventajas pensando en la venta del concepto al negocio y a la Alta Dirección del Banco, así como en su confiabilidad. Sin embargo, uno de los puntos que no se probó fue la integración de aplicaciones que tengan el mismo objetivo como es el administrar y operar los diferentes productos y servicios que se ofrecen.

Dentro del *DW* , y a diferencia de las aplicaciones operacionales, la forma de organizar la información es por tópicos (grandes temas), más que por las mismas aplicaciones. Esto significa que el modelo de un *DW* sugiere que las entidades que lo componen sean genéricas y por lo tanto sean cargadas por distintas aplicaciones.

Para explicar de mejor forma el punto anterior, veamos el siguiente ejemplo en la figura 4.1

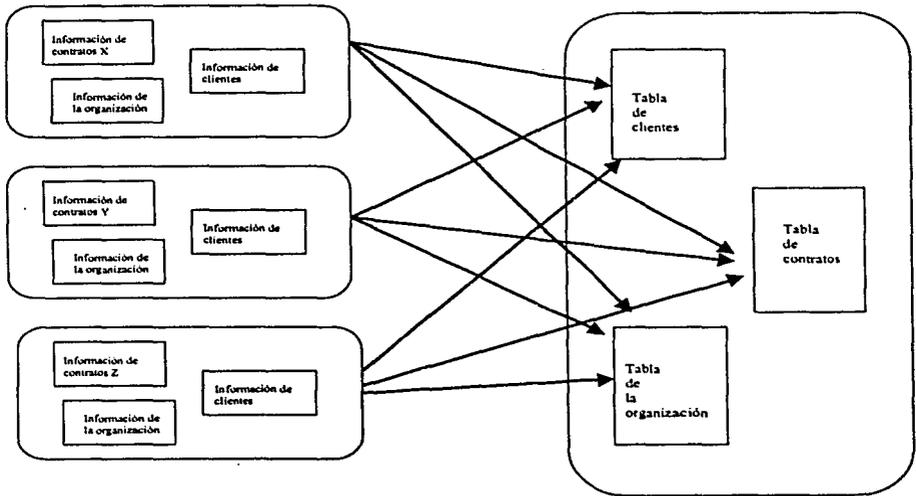


Figura 4.1 Agrupación de información por tópicos

Dentro del diseño del *DW* existen las entidades por tópico como cliente, contratos, direcciones, etc., pero para cargar estas entidades es muy probable que se utilicen entidades de diferentes aplicaciones operacionales.

En este caso, se optó por empezar el *DW* prácticamente con dos aplicaciones, una que es común para todo el banco, " Datos de Clientes" , y una que maneja un tipo de producto, "Cheques". No se tiene la experiencia de integrar más de una aplicación de manejo transaccional dentro del *DW*.

Dentro de este punto es muy importante que se vaya construyendo un modelo de datos corporativo que le dé cabida al menos al 80% de las aplicaciones transaccionales, ya que de lo contrario podría generar retrabajos y/o ajustes en el modelo de datos del *DW*.

Para crear un modelo de datos corporativo, es importante crearlo desde el punto de vista del negocio, además de que se lograría homogeneizar el lenguaje en la organización y serviría de base para nuevas aplicaciones operacionales.

El modelo de datos corporativo ofrece ventajas significativas en la parte de integración de datos de diversas aplicaciones. Es decir, integra diferentes puntos de vista del negocio sobre los datos y los representa en un solo modelo de datos, es por esto que se recomienda para las siguientes fases del proyecto.

De entrada esto no parece ser fácil, pues puede resultar un modelo que no es compatible con los sistemas actuales, pero vale la pena buscar su definición.

4.3 Funciones

Analizaremos cómo se aplicaron las funciones de Adquisición, Administración , Operación, Acceso, Navegación y Distribución en el proyecto del *DW* del Banco.

4.3.1 Adquisición

Esta función fue una de las más desarrolladas en el Banco durante el proyecto. Tomando en cuenta los sistemas operacionales críticos del Banco, se definió el tipo de fuentes de las cuales se alimentó el *DW*. En base a esto, se adquirió la herramienta *Prism* para adquisiciones de datos después de evaluarla con otras herramientas del mercado.

El trabajo que se hizo en este punto fue importante, se realizaron los procedimientos de extracción, mapeo, conversión y transportación apoyados con esta herramienta y los resultados fueron satisfactorios reduciendo notablemente el tiempo de desarrollo respecto a lo estimado de no haber usado esta herramienta.

4.3.2 Administración

Como vimos en el caso, la forma que el Banco trabajó el proyecto del *DW* fue a partir de una iniciativa del área de Arquitectura. Para ello, se definió un grupo especial y dedicado a definir y crear el ambiente *DW* incluyendo la información base con apoyo del usuario y ejecutivo de cuenta responsable.

Al momento del desarrollo de esta fase, no se tenía definido en la organización quien debería dar continuidad a las funciones del *DW*. Se sugiere que para los siguientes pasos o proyectos, se defina claramente la organización que continuará.

Este aspecto puede convertirse en un problema potencial, ya que mientras más tiempo pase y desarrollos haya, será más difícil pasar la responsabilidad y el conocimiento a la organización definitiva.

No es nada sencillo definir la organización dentro del Banco, pues como comentamos, la organización de Sistemas está estructurada por ejecutivos de

cuenta quienes atienden a un Negocio. Poner en manos de algún ejecutivo el desarrollo del *DW* puede provocar limitaciones, ya que el *DW* tiene un carácter corporativo y si se hace de esta forma se puede correr el riesgo de perder la visión global llevando a competir en prioridades los proyectos de construcción del *DW* con los de Negocio.

Si la situación actual se mantiene por más tiempo, será muy difícil pasar este proyecto al área responsable, ya que mientras más grande y complejo sea el *DW*, se tendrá que hacer un esfuerzo mayor para instalarlo dentro del Banco. Los procesos que debemos tomar en cuenta dentro de este cambio son:

- Capacitación del concepto.
- Capacitación de herramientas.
- Capacitación en los sistemas que alimentan el *DW*.
- Conocimiento de la información en el *DW*.
- Conocimiento en la organización del *DW* (Modelo de Datos).
- Relaciones con las áreas usuarias y de sistemas.
- Soporte del segundo nivel a los procesos del *DW*.

Contar con la organización definida es importante pues es ésta la que administrará las bases de datos del *DW*. En el caso de las colecciones, es claro que quedaría a cargo de los ejecutivos de cuenta de Sistemas que apoyan a los Negocios, ya que éstos son los usuarios que requieren las colecciones departamentales o individuales.

Entre los riesgos más importantes al no contar oportunamente con el grupo definitivo de soporte al *DW* tendremos:

- No contar con el personal adecuado para el mantenimiento al DW del Banco lo que repercutiría en la calidad del proyecto.
- No responder adecuadamente a las necesidades del negocio en cuanto a aplicaciones de análisis y de toma de decisiones.
- Perder definición en las responsabilidades dentro del ambiente DW con lo que sería difícil dar seguimiento a la función completa del DW.
- Perder la credibilidad del DW por parte del usuario al no contar con un apoyo adecuado a sus necesidades de información. Si esto pasa perderíamos también el apoyo de los patrocinadores del proyecto

La función de administración es corporativa y debe tener una visión completa de la organización sin marginar algunos sectores de la institución, dándole un carácter global. Se pueden encontrar más patrocinadores y se pueden alcanzar mayores beneficios.

4.3.3 Operación

En el aspecto de operación se utilizaron los criterios definidos en el Banco para la operación de cualquier aplicación. Esto incluyó actividades desde planear la puesta a producción, el desarrollo de procedimientos y su instalación.

La forma en la que se instaló esta función dio buenos resultados. Es posible que se puedan encontrar mejoras en la parte de operación logrando responder rápidamente a cambios y mejoras o nuevas funciones a lo instalado, considerando que una operación de este tipo. Tiene demandas de operación

diferentes y con tiempos de respuesta diferentes a los que requiere las aplicaciones operacionales.

Por ejemplo, las aplicaciones operacionales, necesitan de un estricto control de cambios, uso de CPU, disco, etc. y las aplicaciones para toma de decisiones tienen un contenido de mayor agilidad para incorporar nuevos cambios y mayor tiempo y disco asignado para procesos, sobre todo los no planeados.

A este tema podemos ligar la infraestructura utilizada, que fue la misma que actualmente tienen las aplicaciones operacionales con sus restricciones de tiempo y control de acceso, planeación de la producción, etc. Sería conveniente analizar más a fondo si la infraestructura utilizada es la más indicada para el tipo de aplicaciones que se tienen.

4.3.4 Acceso

En este punto, se llevaron a cabo varios cambios con respecto a la forma tradicional de desarrollo del Banco, ya que una de las premisas fue el de no desarrollar sin el apoyo de herramientas automatizadas (hasta donde fuera posible). Este cambio permitió que se adquirieran herramientas que accesan información y permiten hacer algún tipo de análisis y presentación.

Para esta función, se optó por evaluar y adquirir una herramienta de consulta de información (*Query Tool*) de tipo gráfico, que generalmente corren en equipos personales y se conectan a la base de datos donde se encuentra la información. La herramienta fue el *Business Objects*.

Esta decisión provocó buena aceptación tanto para el personal de sistemas como para el negocio.

Las ventajas que presenta esta decisión son:

- Uso sencillo para el usuario por ser amigable.
- No es necesario saber programar.
- El tiempo para configurar los accesos es poco.
- Permite al usuario acceder la información en términos que él conoce.
- Reduce requerimientos de reportes y desarrollos de extracción de información de única vez para las áreas de sistemas.

Es conveniente que para otro tipo de requerimientos más especializado como *marketing*, *rentabilidad*, etc., se evalúen otras herramientas y así poder darle independencia al usuario en el cumplimiento de sus metas.

4.3.5 Navegación

Este punto es muy ligado al de metadatos. En el momento que se concluyó la primera fase, el Banco no tenía un sistema claramente definido de Navegación. En ese momento no implicó mayor problema ya que solo existían dos aplicaciones en el *DW*. Pero es importante que se defina y desarrolle antes de que se integren más aplicaciones.

Esta función ofrece ayuda y autonomía al usuario final, ya que permite conocer qué información está contenida en el *DW* y en qué condiciones está almacenada, volviéndose crítica en la medida que el *DW* crece.

4.3.6 Distribución

Esta función no fue desarrollada en el caso del Banco ya que los requerimientos planteados originalmente no planteaban envío de información a otras localidades. Sin embargo, ésta puede ofrecer grandes oportunidades en el desarrollo del *DW*, por lo que se debe investigar sobre herramientas y formas de lograrlo.

4.4 Topología

La topología utilizada en el caso fue la de un *DW* único central, donde la información base y las colecciones están almacenadas en el mismo sitio (ubicación y equipo).

Dentro de la Dirección de Arquitectura Tecnológica del Banco, actualmente se analiza si las colecciones pueden migrarse a una plataforma del tipo de servidores, validando que sea lo más conveniente para los usuarios finales.

La figura 4.2 muestra la forma actual y la topología a corto plazo que está evaluando el Banco.

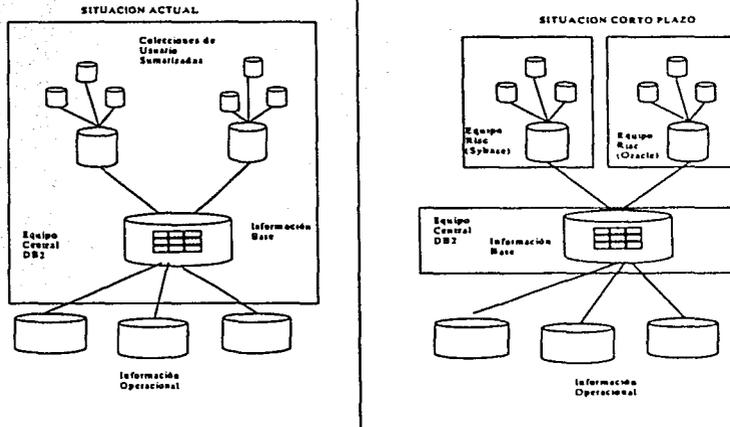


Figura 4.2 Topología del DW

4.5 Pasos para la elaboración de un DW

De acuerdo al capítulo tres, en donde se describen algunos pasos importantes a seguir en la construcción de un DW, podemos mencionar lo siguiente:

- **Identificar las necesidades del usuario final.**

Como mencionamos al principio de éste capítulo, se identificó plenamente los requerimientos y necesidades del usuario.

- **Búsqueda de la fuente más apropiada para alimentar el DW.**

Durante el proyecto se identificó claramente cuáles eran las fuentes más apropiadas para que de ahí se obtuvieran los datos para el DW. Ejemplos de ello fue la definición clara de las aplicaciones de "Cheques" y "Base de Datos de Clientes".

- **Identificar las reglas de Negocio claras.**

Dentro del proceso del Modelo Lógico, se identificaron claramente las definiciones de cada dato que se requería y su relación con otros, de tal forma que permite dar el primer paso a una definición de un modelo de datos corporativo.

- **Generar las bases de datos físicas.**

Dentro del apéndice se puede apreciar el Modelo Físico definido para este proyecto. No hay que perder de vista que el caso de la base de datos del *DW* es un modelo relacional normalizado y que para proyectos futuros deberá soportar (con algunos ajustes) otras aplicaciones.

- **Generar los procedimientos para llevar la información de la operación del *DW*.**

En este punto, se utilizó una herramienta que permitió generar todos los procedimientos y programas necesarios para la población del *DW* (*Prism*).

- **Construir el acceso al *DW*.**

Paso muy importante, ya que sin él, los usuarios no podrían validar y ver satisfechas sus necesidades. En este punto se contruyeron dos tipos de acceso:

- Por medio de una herramienta de consulta de información (*BO*).
- Contruyendo pantallas de acceso directo

- **Implantación de los Mecanismos de administración, operación, soporte y mantenimiento.**

En este tema fueron implementados de acuerdo a los estándares y mecanismos establecidos en el banco para operar aplicaciones.

- **Equipo de trabajo y tiempo de desarrollo.**

En este punto, el grupo fue definido para dedicarse al proyecto con un número de 12 personas y llevo un tiempo de 10 meses (considerando los dos meses de planeación y capacitación en una metodología adoptada), acordes con algunas sugerencias hechas por especialistas en el tema.

4.6 Recomendaciones y propuestas

Hasta este momento, hemos criticado la implementación efectuada del caso, comparándola con los conceptos y definiciones del *DW*. Se han descrito las actividades que a mi juicio se efectuaron correctamente y las actividades y/o funciones que requieren de ser mejoradas o desarrolladas más a fondo.

Con el objeto de hacer un resumen y puntualizar los aspectos que nos parecieron más relevantes, esta sección la hemos dividido en dos partes:

- La primera presentará un resumen de las razones por las cuales considero que el trabajo desarrollado fue satisfactorio.
- La segunda parte expondrá algunas propuestas para aquellas actividades y/o funciones que, desde mi punto de vista, hay que mejorar en las futuras implementaciones.

4.6.1 Resumen de beneficios

El *DW* implementado en el Banco mostró resultados satisfactorios para la organización porque:

- El impacto de la primera fase sobre el negocio fue positivo. Actualmente están accesándolo un promedio de 1300 personas vía terminal asíncrona y 30 usuarios analistas que usan la información.
- El usuario está conforme y confía en el concepto, ya que lograron satisfacer sus requerimientos de corto plazo y, cuentan con una herramienta que les permite satisfacer sus necesidades de extracción de información sin la intervención del área de Sistemas.
- Se partió de una arquitectura de información claramente definida
- El trabajo presentado ofrece una propuesta muy estudiada y formal de la construcción de un *DW*
- Muestra una aplicación de la propuesta en el Banco donde ha sido bien recibida
- Para el Banco, ha propiciado que los esfuerzos continúen en el desarrollo de este almacén de datos con el apoyo de usuarios y de otras áreas de Sistemas

Por otro lado, es importante considerar que esta tecnología toma fuerza cada vez más, ya que propone una solución real a las empresas que ven en los sistemas de análisis un factor crítico en su desempeño. Adicionalmente, el trabajo presentado es de los esfuerzos serios desarrollados en México, que se han llevado a un uso extenso.

Esta tecnología está en proceso de maduración y existen varias compañías de hardware, software y consultoría que pueden interpretarla de diferentes formas. Lo importante es tomar provecho de esas propuestas para

lograr los mejores resultados para las organizaciones, situación que sin duda el Banco consideró desde el momento que definió como solución a su problemática la construcción de un almacén único de información.

El trabajo presentado ofreció una propuesta muy estudiada y formal del *DW* y además muestra su aplicación en una organización grande y compleja y que sin duda fue bien recibida en el Banco.

En el momento de concluir la primera fase, es muy prematuro pensar en las ventajas competitivas que le generará el *DW* completo a la organización, pero es un hecho que estos beneficios serán más importantes conforme se incorpore más información.

4.6.2 Detalle de Recomendaciones y Propuestas

Sin embargo, durante la implementación del caso, existieron algunas fallas que, consideramos importante mencionar para dar algunas recomendaciones y propuestas de solución, como aportación para que se tomen en cuenta en las siguientes fases.

Todas las fallas y recomendaciones fueron mencionadas a lo largo de la crítica efectuada. En esta sección se profundizará en cada una de ellas.

4.6.2.1 Propuesta para la Administración del *DW*

Se debe definir lo antes posible el equipo de trabajo que administrará y dará soporte al *DW*. Este es uno de los puntos que más áreas de oportunidades tiene en el caso. En la primera etapa prácticamente contó con dos aplicaciones y esto hizo que la falta de definiciones no impactara de algunos puntos como:

- Definir los equipos de trabajo que darán soporte al *DW*
- Definir el esquema para poder diferenciar costos de los participantes en el *DW*. Este punto es importante ya que, si son muchos los usuarios y muchas las aplicaciones que lo alimentan, la forma de determinar los costos se vuelve compleja.
- Definir prioridades para la entrada o modificación de información al *DW*. Sugerimos que esto se realice con un comité de usuarios para revisar fases del proyecto y prioridades. El *DW*, por su naturaleza centralizada, tiende a ser un cuello de botella, por lo que es importante planear cada fase en conjunto con los usuarios del negocio.

Como mencionamos en las secciones anteriores, la forma de organización del Banco sugiere que por cada unidad de negocio exista un ejecutivo de cuenta de la parte de sistemas. Esto hace que no sea claro dónde ubicar la función de la administración. No olvidemos que el *DW* tiene un alcance corporativo por lo que el grupo que lo administre tiene la responsabilidad de satisfacer necesidades de todo el negocio. Siempre ante cualquier necesidad, aunque sea de un solo departamento, se debe pensar en toda la organización.

Dentro del esquema del *DW* podemos separar las funciones, y cuando un área de negocio necesita información, el ejecutivo de cuenta de sistemas debe atender dicho requerimiento, elaborar las colecciones necesarias y dándoles las herramientas adecuadas para acceder la información. Pero la parte de la base, de donde se alimentan las colecciones, debe estar regido por un grupo centralizado con visión corporativa. La figura 4.3 ilustra este punto

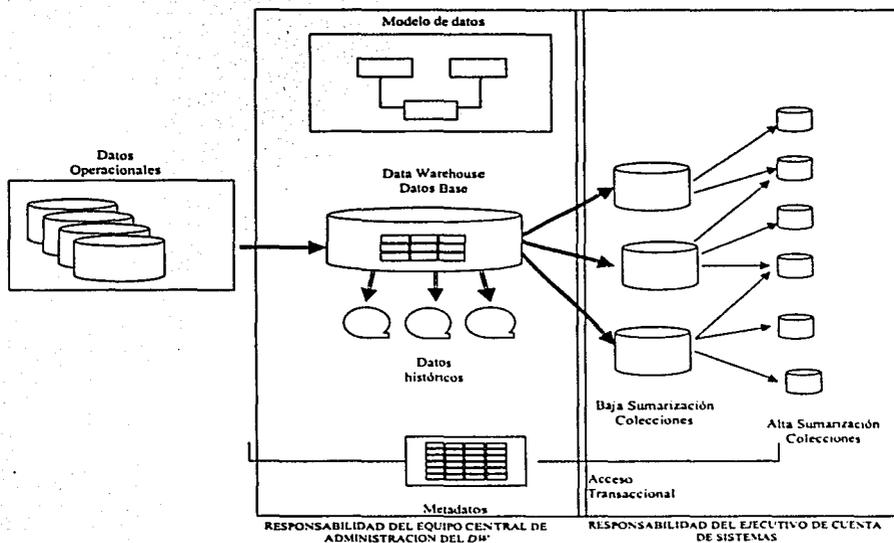


Figura 4.3 Responsabilidades de Sistemas en el DW

Veamos en forma general las actividades que se necesitan para la administración del *DW* por parte del grupo centralizado:

- Análisis y desarrollo de nuevas funciones o incorporaciones de información al *DW*:
 - Recibir el requerimiento junto con el ejecutivo de cuenta.
 - Planear el desarrollo de la incorporación a las bases.
 - Modificar el modelo de datos del *DW*.
 - Coordinar posibles modificaciones e impactos.

- Desarrollar programas y procedimientos para la nueva información.
- Apoyar en las herramientas de acceso para el usuario final.
- Apoyar en la elaboración de colecciones de usuario.
- Mantenimiento a la operación en producción del *DW*:
 - Desarrollar y optimizar procedimientos de mantenimiento al *DW*.
 - Dar soporte de segundo nivel a los procesos del *DW*.
 - Monitorear los procesos involucrados al *DW* para detectar oportunidades.

Sugerimos un grupo permanente para la administración del *DW* y es el siguiente:

- Un Gerente de proyecto.
- Dos líderes.
- Cuatro analistas de datos.
- Dos administrados de Bases de Datos.

Este primer equipo de trabajo se puede pensar que es grande para iniciar con la función pero, es importante considerar que, el equipo requiere de una capacitación especializada y que por el tiempo que ésta consume, es conveniente iniciar con un grupo de tamaño razonable. Este equipo después podrá ser tutor de otros si así se requiriera. Adicionalmente es necesario considerar sustitutos en caso de ausencias.

Este grupo podrá crecer en la medida que crezca la información contenida en el *DW*. Será responsable de coordinar los siguientes desarrollos o nuevas incorporaciones y podrá apoyarse de proveedores externos de analistas y programación, por ser funciones finitas y perfectamente acotadas.

La razón por la que sugerimos un grupo externo para nuevos desarrollos, es porque cada proceso de adicionar información al *DW* es una actividad finita y variable en cuanto al número de recursos y tiempo.

La actividad de maquilar el desarrollo o nuevas incorporaciones no justifican el contar con un equipo fijo de personal, es mejor contratar el personal necesario durante el tiempo requerido en el proyecto de integración de información.

Este grupo central de *DW* debe tener una comunicación directa con los otros grupos de sistemas involucrados, que son, básicamente, los ejecutivos de cuenta que le dan la cara al usuario y el administrador de datos corporativo.

Sobre la ubicación en la organización de la administración del *DW*, la figura 4.4 sirve para explicar mejor la propuesta

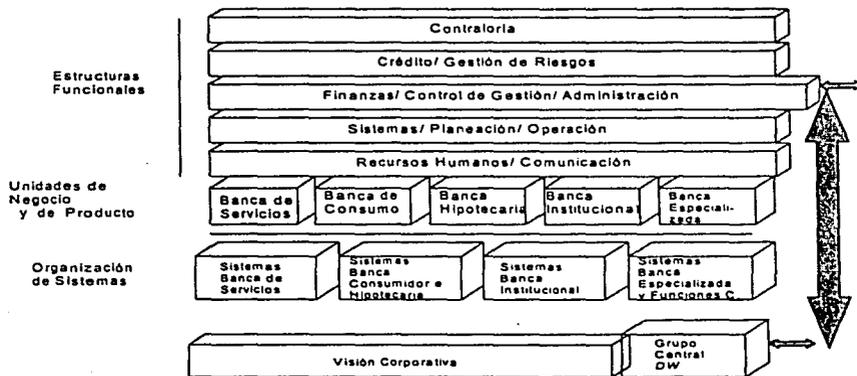


Figura 4.4 Administración del *DW*

El banco está dividido por bancas y áreas funcionales. El área de Finanzas, Control de Gestión y Administración es una fuerte candidata para administrar las necesidades de todos los negocios, y que estén relacionadas con requerimientos para sistemas de análisis. Entre las funciones del área de gestión están definir el cómo se calculará el estado de resultado del banco y por ende de cada banca, obtener la información, analizarla y presentarla a la dirección. También cuando una banca necesita incorporar nuevos datos para el cálculo de su margen financiero, esto se realiza analizando y en su caso, autorizando por esta área. Es por ello que podría fungir como la coordinadora de necesidades y usuarios del *DW*. Por otro lado, y dada importancia de la administración, sugerimos que el área que se establezca para la dicha función, debe depender directamente de la dirección general de Sistemas. Con esto se garantiza darle continuidad y todo el apoyo que requiere en la organización.

4.6.2.2 Propuesta para el Desarrollo de Metadatos

Como comentamos, no hay una propuesta clara y completa para manejar los metadatos. Estos juegan un papel muy importante dentro del *DW*, ya que representan el directorio de la información existente en el *DW* y datos referentes a esta información. Es por ello, que es necesario pensar que la utilidad de los metadatos crece en la medida que se integren nuevos datos al *DW* o se genere mayor historia.

Por lo anterior, es necesario reflexionar en la construcción de infraestructura para los metadatos buscando mediano y largo plazo.

Este tema puede causar confusión y puede hacerse un elemento muy complicado. Muchos autores exponen un esquema complejo para el manejo de los metadatos.

Se recomienda que se empiece por definir qué necesitamos acerca de los metadatos, es decir, qué tipo de metadatos se quieren tener en la organización y a partir de esto, desarrollar la estrategia propia.

Para el caso del banco, se propone que:

- la ubicación de los metadatos sea la misma plataforma que los datos del *DW*,
- con el mismo manejador de bases de datos y
- diseñar una pequeña base de datos con esta información.

Se sugiere que se empiece con el desarrollo de:

- El inventario de tablas físicas base del *DW* que incluya nombre, alias, descripción, fecha de última actualización de la tabla.
- Inventario de columnas (campos) por cada tabla donde se muestre el nombre de la columna, descripción, tipo de dato, fórmulas de cálculo del dato (si aplica), fuente de donde proviene el dato.
- Datos de los responsables del *DW* central y de las aplicaciones operacionales que alimentan al *DW*.

Se propone que también se empiece con un acceso de datos por medio de la herramienta de consulta de información ya implementada para los usuarios (*Business Object*), por su facilidad y conocimiento.

4.6.2.3 Propuesta para la Integración del Modelo de Datos del *DW*

En virtud de que durante la primera fase del *DW* se creó un modelo basado en tópicos, pero no considera el soportar a toda la organización, es necesario desarrollar un modelo de datos corporativo.

Si logramos desarrollar un modelo de datos corporativo, en primera instancia a un nivel alto en donde se definan claramente los grandes tópicos de datos, desde el punto de vista de negocio, serviría para ir guiando los subsecuentes desarrollos del *DW*.

En el mercado existen algunos modelos corporativos, como por ejemplo el de IBM llamado Modelo de datos del sistema financiero (*FSDM*), que tienen como objetivo regir la comunicación entre negocios y sistemas. Otros proveedores lo que venden es la construcción de un modelo de negocios que finalmente guiará la construcción del *DW*. Para ambos casos es costoso su adquisición. Es por ello que se sugiere que la empresa construya su propio modelo.

Para construir su propio modelo de datos se recomienda que se efectúe con la participación de todos los negocios (un representante a nivel Dirección), en sesiones de trabajo donde lo que hay que obtener es:

- Las preguntas de negocio más importantes no resueltas por la información
- Las grandes estrategias comunes de los negocios.

Con las preguntas se identifica a qué nivel se requiere la información, los grandes tipos de información y frecuencias.

Con las estrategias comunes se identifica qué conceptos y tipos de análisis se requieren. Por ejemplo: análisis de mercado, rentabilidad por cliente, cálculo de riesgo en una serie de tiempo, análisis del comportamiento del cliente para segmentarlo, etc.

Con lo anterior se puede armar un modelo corporativo, conceptual y de alto nivel.

La definición del modelo corporativo conceptual debe efectuarse antes de iniciar la siguiente fase y utilizar ésta como prueba del modelo.

Una vez desarrollado el modelo y probado con la fase 2 del proyecto de *DW*, es conveniente incorporar un artefacto temporal que permita extraer información de ambos modelos de datos (Fase 1 y Fase 2) para que sea transparente para los usuarios. Por último, se eliminaría el artefacto junto con la migración de datos de la fase 1 al nuevo modelo corporativo. La figura 4.5 representa la secuencia que se propone.

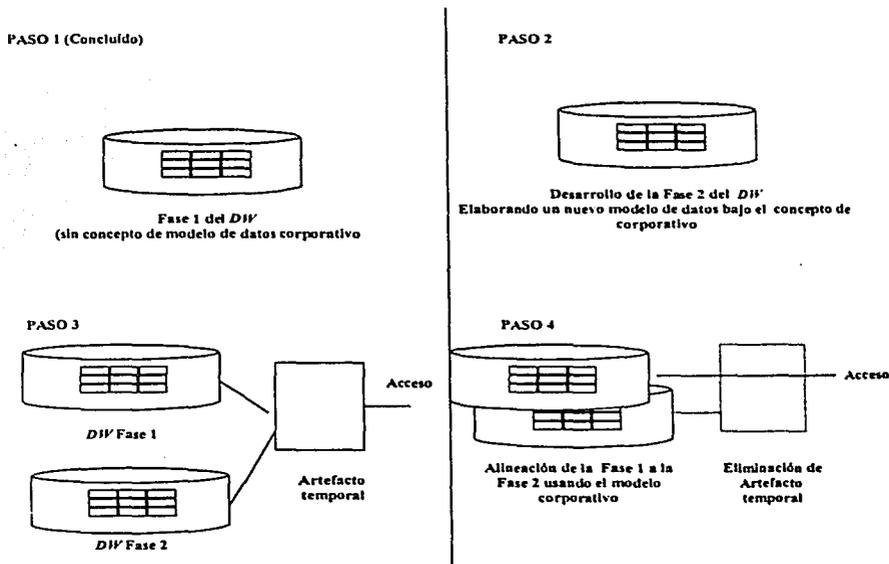


Figura 4.5 Integración del modelo de datos corporativo para próximas fases del *DW*

4.6.2.4 Propuesta cambio de plataforma del *DW*

Otra recomendación es el plantear na arquitectura para el *DW* en diferente plataforma. El plan a corto plazo que se tiene de pasar las colecciones departamentales y generar las de usuario en plataformas fuera del equipo central es adecuada.

Podemos pensar en extraer también del equipo central no sólo las colecciones sino la información base. Esto significa que el *DW* se genere en una plataforma que presente mayor flexibilidad para acceso, para la construcción y administración, trabajando como un servidor de información. Para ello, en el mercado existen los equipos de proceso paralelo que ofrecen ventajas importantes para las aplicaciones de consulta y análisis como lo es el *DW*.

Las máquinas paralelas ofrecen ventajas al mejorar el rendimiento de dos formas: la primera es reduciendo el tiempo de acceso a grandes volúmenes de información y la segunda reduciendo el tiempo de espera en transacciones pequeñas.

Ya que el *DW* almacena grandes cantidades de información y los accesos que debe soportar son de barridos completos a altos volúmenes, las máquinas de proceso paralelo son más convenientes para este tipo de necesidades.

En la figura 4.6, se puede apreciar la diferencia entre hacer consultas en un ambiente de un procesador y uno de proceso paralelo.

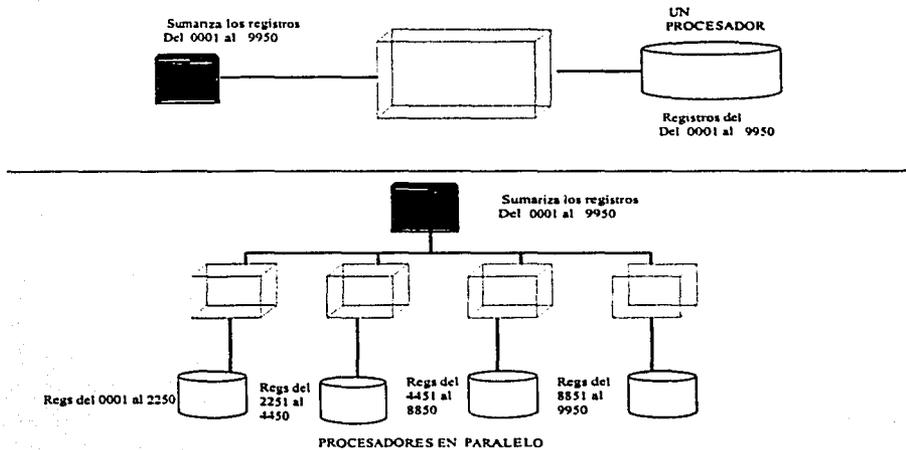


Figura 4.6 Entorno de procesadores en paralelo

Dentro de este ejemplo, un acceso podría ser dividido en cuatro procesadores en paralelo con lo que se reduciría enormemente el tiempo de proceso.

Los equipos de proceso paralelo se componen de diversos procesadores y tienen la facilidad de que pueden crecer fácilmente.

Por otro lado, los recursos necesarios en un equipo paralelo son más económicos que un equipo central uniprocador. Sin embargo, se complica el control del proceso en esquemas de varios procesadores, por lo que se debe poner más énfasis en la forma de diseñar las aplicaciones.

Dentro del entorno paralelo, se debe contemplar que puede existir paralelismo a nivel del equipo (hardware), del sistema operativo, del manejador de base de datos y de la aplicación. El *DW* tiene grandes oportunidades en cuanto a su rendimiento existiendo en un entorno de este tipo, pero siempre cuidando el diseño del entorno.

En la parte de manejadores de bases de datos, la mayoría de los proveedores más importantes están lanzando versiones de sus manejadores para entornos paralelos, con lo cual se puede construir un ambiente adecuado para el *DW* paralelo.

Se sugiere que se investigue este tipo de tecnología en cuanto a máquinas y manejadores paralelos y se incluya como un factor importante de análisis el tamaño de instalaciones con este tipo de tecnologías en arquitecturas de *DW*.

Es muy importante considerar como un criterio obligatorio el número de instalaciones exitosas en el mercado de este tipo. De hecho, es conveniente hablar directamente con las compañías que cuenten con estas instalaciones y que sean semejantes en cuanto a tamaño y volúmenes para compartir sus experiencias y considerarlas en el proceso de selección.

Otro criterio muy importante es el nivel de servicio que proporcionan los proveedores de las máquinas paralelas. Dependiendo del soporte o capacitación, se puede decidir manejar en el banco por máquina externa.

La arquitectura que se propone considera los planes a corto plazo del banco en donde las colecciones las migrará a otra plataforma; se muestra en la figura 4.7

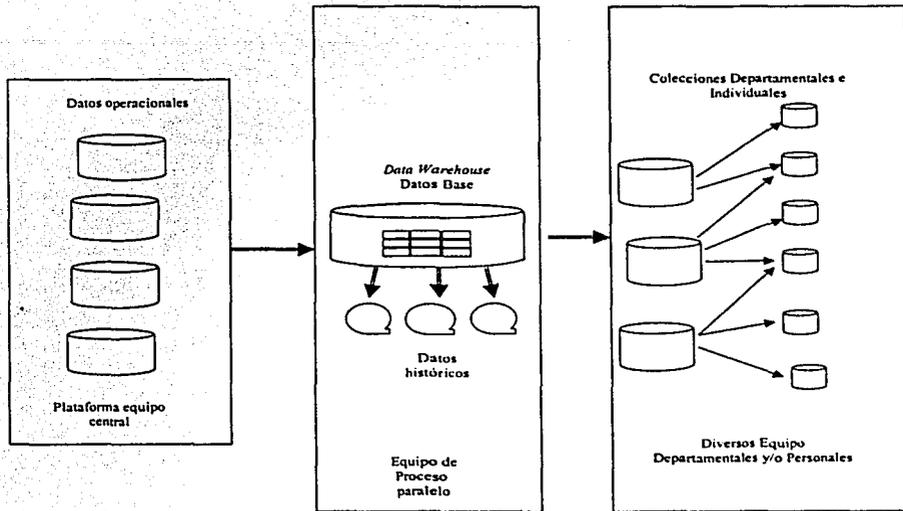


Figura 4.7 Arquitectura DW usando equipo paralelo

Dentro de esta propuesta, la información operacional se separa del DW, de tal forma que el DW se ubica en un sitio diferente al equipo central. Esta nueva plataforma puede ser un equipo paralelo y a partir de ella, se pueden crear colecciones de usuario final en otras plataformas más cercanas al usuario cuando esto convenga. También es factible conectar accesos directamente al DW.

4.6.2.5 Otras recomendaciones

A continuación se presentan algunas sugerencias de temas que no se tocaron ampliamente en la instalación del caso, pero que consideramos relevantes

para las siguientes fases del proyecto. Algunas de ellas requieren de mayor investigación, pero no por ello dejaremos de mencionarlas.

4.6.2.5.1 Planeación de la capacidad

Tenemos que ser cuidadosos con tener y seguir el plan de la capacidad del *DW*. Es muy evidente que al almacenar información histórica de varias aplicaciones y con un nivel alto de granularidad, el *DW* tiende a crecer exponencialmente. El punto aquí es asegurarse que se está cargando la información necesaria para que el usuario pueda efectuar sus análisis.

Una de las funciones importantes del equipo de soporte al *DW* es la de planear y medir el uso de espacio dentro del almacén. Este equipo deberá monitorear junto con el área de planeación de la capacidad, la forma en que el *DW* está necesitando recursos de almacenamiento y de procesamiento.

> Recursos de almacenamiento

Se debe cuidar tanto el espacio que ocupa el almacén como el necesario para las áreas de trabajo y dispositivos de información histórica. Este tema es uno de los objetivos críticos del grupo que administrará el *DW* y en cada fase del proyecto se deberá planear el espacio que ocupará. Las funciones al respecto que deberá llevar el grupo responsable del *DW* son:

- Planeación del espacio en cada fase del *DW*.

En base al modelo físico del *DW* de cada fase, calcular el espacio necesario para las bases de datos y áreas de trabajo, cintas y/o discos de respaldo y para almacenamiento de datos históricos. Es recomendable hacer un pronóstico de cómo se van a ocupar estos recursos en el tiempo.

- Solicitud de recursos al área de planeación de la capacidad.

En base al estudio realizado se podrá solicitar el espacio necesario para la operación del *DW*. Esto debe ser planeado en función al proceso establecido en la organización para la adquisición de dispositivos.

- Monitoreo de espacio ocupado por el *DW*.

Como acción conjunta entre el grupo responsable del *DW* y planeación de la capacidad corporativos, se deberá revisar periódicamente (se recomienda periodos bimestrales) el comportamiento del uso de recursos de espacio. Con esto podremos hacer alguna acción preventiva o ajustar los cálculos de planeación de espacio del *DW*.

➤ Recursos de procesamiento

Relacionado con este tema, debemos tomar en cuenta que en el *DW* típicamente se realizan consultas que pueden tomar mucho tiempo en resolverse, y éstas no pueden ser del todo planeadas. El usuario debe hacer investigaciones con la información que se encuentra almacenada en el *DW* y dependiendo de sus hallazgos podrá profundizar o no. Por esto se sugiere el uso de monitores conectados al manejador de base de datos para poder detectar cómo se están llevando los accesos al *DW*, y tomar alguna decisión de diseño para reducir el procesamiento. Por ejemplo, podemos detectar que se están realizando frecuentemente accesos que combinan información de dos entidades, y puede provocar una fuerte demanda de recursos de proceso. La decisión que se podría tomar en el ejemplo es la de construir una nueva entidad que contenga la información combinada con las dos entidades y, con esto, apoyar a bajar la demanda de recursos.

Monitorear los accesos toma relevancia en la medida que crezca la información en el *DW* y el número de usuarios, por lo que recomendamos que desde esta primera etapa se utilicen monitores.

Durante el desarrollo e implantación de la primera fase del caso, no se consideró la instalación de algún esquema formal de monitoreo de consumo de recursos y tampoco la planeación de la capacidad se efectuó considerando las características propias del *DW*, por lo que antes de iniciar la siguiente fase debe tomarse el tema e integrarlo como uno de los pasos más a cubrir en cualquier proyecto y operación de sistemas.

4.6.2.5.2 Manejo de archivos históricos

En cuanto al manejo de archivos fuera de línea, no se ha definido la forma en la que los datos pasarán a otros dispositivos cuando ya no sea necesario contar con ellos en línea. Se debe definir cada cuando se realizará esta actividad, cómo se llevará a cabo y dónde se almacenará la información. Adicionalmente, también se debe clarificar cómo se accederá, sobre todo porque puede ser que esté en dispositivos más lentos o económicos.

Aunque el *DW* tiene capacidad de almacenar grandes volúmenes de información histórica, es conveniente poner un límite ya que el almacenamiento en discos tiene un costo alto cuando se habla de grandes cantidades. La información que tiene el *DW* puede estar y no estar en línea. Para la información fuera de línea podría estar en dispositivos mucho más económicos como cintas magnéticas.

Para esta función de manejo de información histórica, es importante la interacción con el usuario para acordar y determinar cuánta información histórica debe estar activa (la que es usada con mayor frecuencia) y qué otra parte podría llamarse a dispositivos fuera de línea.

Proponemos para el manejo de información histórica del caso un esquema como el que se muestra en la figura 4.8

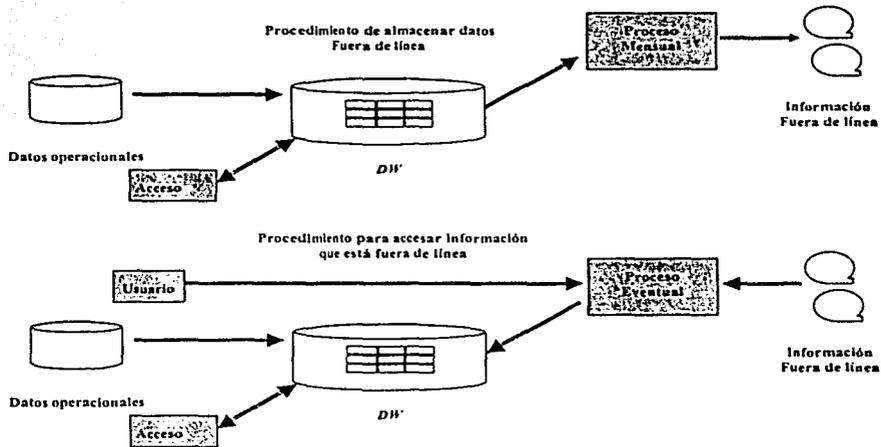


Figura 4.8 Esquema de manejo de datos históricos

De acuerdo a la frecuencia de almacenamiento definida en el caso, los datos referentes a las transacciones y saldos se guardarán diariamente. Se requiere un proceso mensual que extraiga todos los datos del mes y los lleve a guardar a cintas magnéticas. Posteriormente para acceder a la historia, ante un requerimiento del usuario, se ejecutaría un proceso eventual, usando procedimientos que copien de las cintas, la información a la base del DW. Esto significa que cuando se efectúe la planeación de la capacidad del DW, se debe considerar en al menos 15% adicional, previniendo el uso de información histórica. Por otro lado, cuando hablamos de historia, las definiciones de las entidades podrían cambiar, por lo que es necesario guardar en los metadatos las definiciones de las estructuras que están en las cintas magnéticas. Cuando sea fin

de mes, los procesos moverán la información que ha pasado el límite del tiempo, por lo que se depurará automáticamente del ambiente línea.

Los pasos que podemos aplicar en el caso del banco son:

- Definir con el usuario el período de tiempo en el que usará con mayor frecuencia la información, es decir, cuánto tiempo se necesita en línea.
- Definir el período de la información histórica fuera de línea.
- Elaborar procedimientos mensuales para mover datos de línea a cintas y depuración.
- Elaborar procedimientos eventuales de copia de cintas a información en línea.
- Definir esquema para que el usuario solicite se copie información fuera de línea.

4.6.2.5.3 Seguridad

Otro punto que no estuvo considerado en el caso fue el aspecto de seguridad. Es lógico pensar que debemos tener esquemas de seguridad implementado para el *DW* ya que contiene información de toda la organización. La seguridad en el *DW* puede ser compleja partiendo de que la información puede ser restringida para algunos usuarios y para otros no, pero de cualquier manera es un tema importante a abordar para generar confianza en el almacén que maneja el principal activo de la organización.

La seguridad en el *DW* presenta un dilema, ya que por un lado estamos creando un almacén que permita a quien necesite, hacer análisis utilizando la

información que se encuentre almacenada y por otro lado ésta puede ser usada en contra de la organización por lo que se debe guardar un esquema de seguridad.

Para el caso del Banco el esquema que se propone tiene los siguientes lineamientos:

- Definir un comité de usuarios dueños de la información que contiene el *DW*.
- El acceso a esa información se hará por medio de herramientas del tipo consultas de información (*Query Tools*), sistemas de soporte a las decisiones o sistemas ejecutivos (*EIS*) y éstas deben contar con esquemas de seguridad.
- Los administradores de la seguridad de las herramientas deben ser usuarios que pertenecen al comité de usuarios.
- La seguridad de la información se hará combinando los esquemas que ofrece el manejador de bases de datos y las herramientas de acceso.
- Cuando se necesite información del *DW* el comité de usuarios autorizará los privilegios de acceso por medio de la herramienta.

La figura 4.9 muestra un esquema de seguridad que podría aplicar para el *DW*

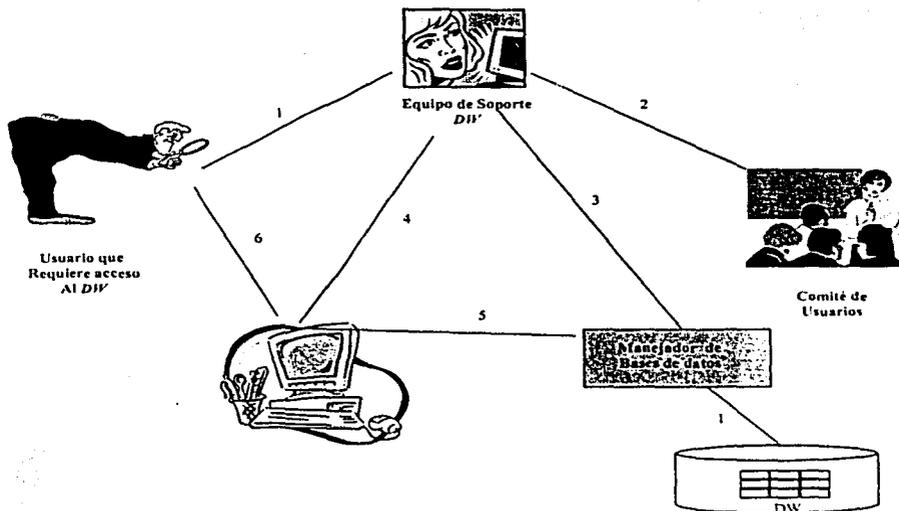


Figura 4.9 Esquema de seguridad DW Bancomer

Para lograr este entorno se necesita definir:

- El comité de usuarios.

Los usuarios dueños de la información presentan las restricciones de uso dentro del DW para que se formule el esquema y clasificación de seguridad de los datos. También se designan representantes del comité quien aprobará el acceso a la información.

- Mecanismo del flujo de autorización de acceso.

Cuando un usuario solicita acceso al DW, el equipo de soporte recibe dicha solicitud y emite un correo electrónico a los responsables involucrados del comité para su autorización.

- Mecanismo de acceso a los datos.

De ser aprobado, el equipo de soporte del *DW* verifica que la herramienta con la que el usuario va a acceder, tenga la autoridad adecuada.

- Mecanismo de acceso con la herramienta por el usuario.

El equipo de *DW* solicita que se le permita uso de la herramienta al usuario con los privilegios definidos en el acuerdo del comité.

- Revisión bimestral del uso del *DW* según monitores.

Periódicamente se tendrá una reunión donde se revisará, entre otros puntos, la forma en que se está utilizando el *DW* (quién accesa, frecuencia, etc.).

Capítulo 5 Conclusiones del estudio

Por último se presentan algunas conclusiones sobre los resultados del presente trabajo que a nuestra consideración nos parecieron las más evidentes.

El trabajo desarrollado generó beneficios importantes, entre los que se encuentran:

- Ha permitido aplicar los conocimientos sobre el concepto de *DW* en un caso práctico, permitiendo profundizar en éste.
- Al ser un caso práctico, ha dado la oportunidad de estudiarlo y analizarlo desde diferentes prespectivas y así lograr hacer las recomendaciones y propuestas aquí expuestas.
- Muestra la importancia que tiene el concepto de *DW* sobre todo en instituciones grandes con sistemas antiguos, con carencias de integración de información y con una necesidad latente de conocimiento de sus clientes.
- Ofrece un documento con experiencia práctica para todos aquellos que esten interesados en el tema.

Se recomienda que en lo futuro se realicen estudios posteriores en temas relacionados con el concepto de *DW* como :

- Investigación sobre el concepto de Almacén de datos operacionales (*ODS*) que en algunos libros como los de Inmon, aparece como una extensión del *DW* y respondiendo a decisiones operativas

- Investigación de aplicaciones de explotación del DW basados en:
 - Sistemas Ejecutivos (*EIS*)
 - Sistemas de soporte a decisiones (*DSS*)
 - Minería de datos (*Data Mining*)
 - Sistemas para la Administración de la relación con el cliente (*CRM*)

Bibliografía

- | | |
|-----------------|--|
| IBM, 1993 | IBM, Information Warehouse Architecture I, Abril 1993. |
| Barry Devlin | Data Warehouse from Architecture to Implementation, Addison-Wesley 1996. |
| Ralph Kimball | The Data Warehouse Toolkit, John Wiley & Sons, 1996. |
| Claudia Imhoff | Building the Operational Data Store, John Wiley & Sons, 1996. |
| Sholom M. Weiss | Predictive Data Mining, Morgan Kaufmann, 1998. |
| Gartner Group | Symposium ITXPO, 1996,97,98. |
| DB/EXPO, 1995 | DB/EXPO, Real IT Solution, Brochure, San San Francisco. |

Hugh Ryan

Building an enterprise information architecture, Infoworld, marzo, 1993.

K. Strange

Data Marts do Not Replace the Data Warehouse, Gartner Group, 1994.

Bill Inmon

Building an Information Architecture for the 90's, Conferencia, 1994.

Henry Lucas

Managing Information Services, McMillan, 1989.

Vidette Poe

Database Programming & Design, Julio 1993.

Anexos:

Anexo 1.- Diseño Conceptual del Segmento de Clientes en el *DW*

Anexo 2.- Diseño Conceptual del segmento de Cheques en el *DW*

Anexo 3.- Ejemplo de Diseño Lógico del Modelo de Datos

Anexo 4.- Diseño Físico del segmento de Clientes en el *DW*

Anexo 5.- Diseño Físico del segmento de Cheques en el *DW*

Descripción de símbolos

Tipos de Relaciones

Relación UNO



Significa que cada valor de llave primaria de una tabla debe aparecer necesaria y obligatoriamente en la tabla a donde llega la línea de relación pero con un máximo de una sola ocurrencia.

Relación UNO o CERO



Significa que cada valor de llave primaria de una tabla puede o no aparecer en la tabla a donde llega la línea de relación pero con un máximo de una sola ocurrencia.

Relación UNO o MUCHOS



Significa que cada valor de llave primaria de una tabla debe aparecer necesaria y obligatoriamente en la tabla a donde llega la línea de relación y sin límite máximo de ocurrencias.

Relación CERO, UNO o MUCHOS



Significa que cada valor de llave primaria de una tabla puede en la tabla a donde llega la línea de relación, esto es, no hay límites mínimo ni máximo de ocurrencias.

Relación Identificadora



Significa que la tabla donde se hereda la llave primaria de otra tabla como llave foránea, es una tabla hija o subtipo de la tabla de donde procede dicha llave, por esta razón la llave foránea forma parte de la llave primaria de la tabla hija.

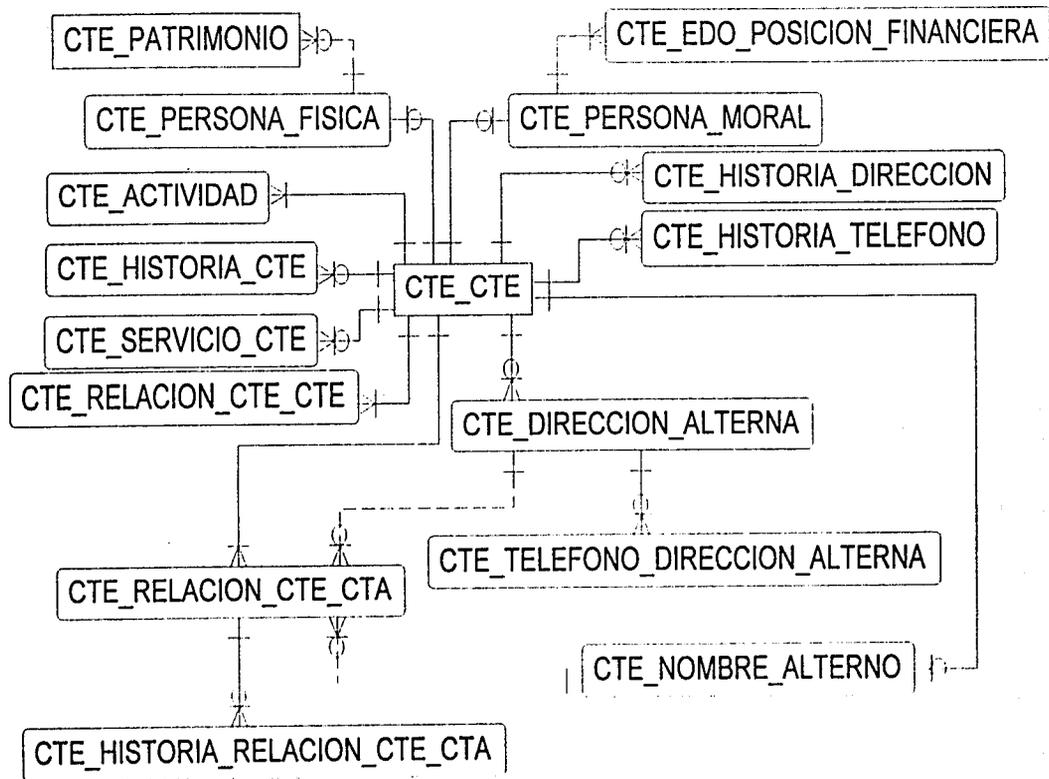
Relación No Identificadora



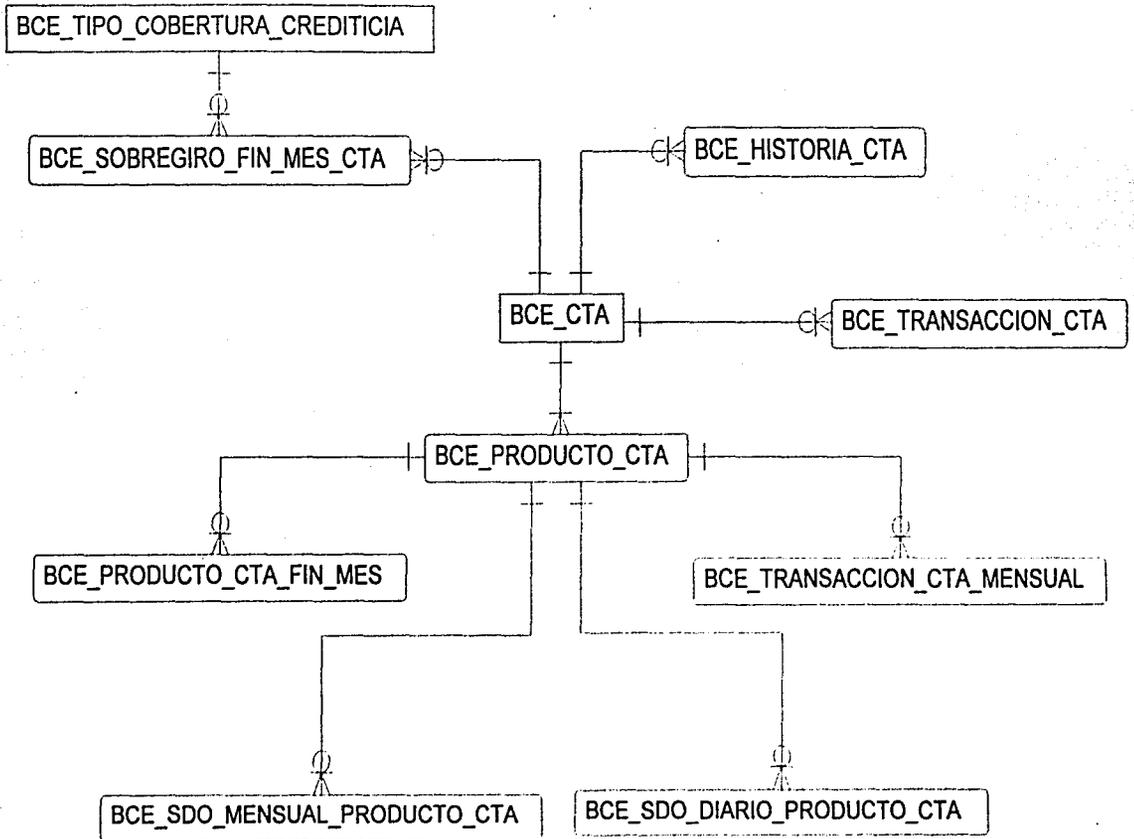
Este es el tipo de relación más simple donde sólo se indica una relación entre dos tablas. En este caso la llave primaria que se hereda de otra tabla como llave foránea es sólo un atributo más en la tabla relacionada (tabla donde se hereda la llave).

NOTA ADICIONAL: Los atributos que aparecen al principio de cada entidad y que están separados del resto por una línea horizontal forman la llave primaria de las mismas.

Anexo 1.- Diseño Conceptual del segmento de Clientes en el Data Warehouse



Anexo 2.- Diseño Conceptual del segmento de Cheques en el Data Warehouse



Anexo 3.- Ejemplo del Modelo Lógico

1. Diagrama general de descomposición del modelo lógico
2. Diagrama de entidad Acuerdo – AC (8)
3. Descripción de la entidad Acuerdo (33)
4. Diagrama de entidad Producto - PD (29)
5. Descripción de entidad Producto (48)
6. Diagrama de entidad Participante – PA (27)
7. Descripción de entidad participante (47- 49)

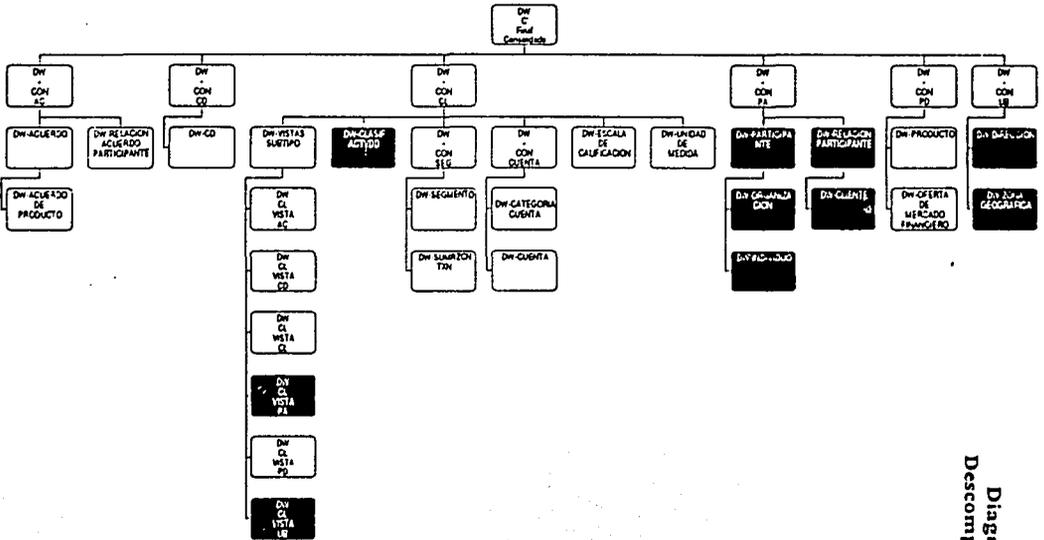
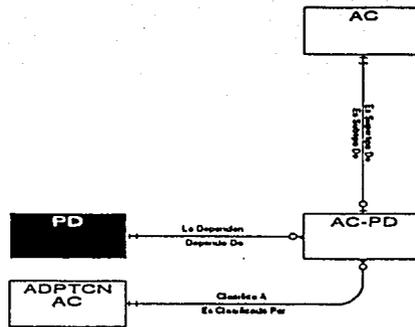


Diagrama de Descomposición



DEFINICIONES DEL NIVEL C' FINAL

Subject Area: DW C' Final Consolidado

MODELO DE DATOS C' FINAL CONSOLIDADO DEL INFORMATION WAREHOUSE

Incluye todas las Aplicaciones y sus respectivos Objetos de Negocio que residirán en el Almacén de Información:
 1) Modelo de Datos para la Aplicación "Bancomer, Gestión Estadística" (BGE) El Sistema de Gestión Estadística (BGE) realiza el seguimiento cuantitativo no contable como fuente única de información para el control de los recursos, productos y mercado objetivo de la Institución. Seguimiento: Registro histórico de sucesos Recursos: Unidades de Negocio, Funcionarios Control: Puntos de Evaluación, tomar acciones para modificar el comportamiento Mercado Objetivo: Clientes, segmentos, competidores, áreas geográficas Productos: Captación, Colocación, Fiduuciario, etc.

Entidad: AC

Una Entidad Fundamental y una Entidad Focal. Un ACUERDO representa a un convenio, ya sea potencial o real, que implica a dos o más PARTICIPANTES, que proporciona y afirma las reglas y obligaciones asociadas con la venta, intercambio o provisión de bienes y servicios. Por ejemplo, el ACUERDO N° 123 (un acuerdo específico de Certificado de Depósito entre la Institución Financiera y Carlos Fernández); el ACUERDO N° 346 (un convenio de testamento específico entre la Institución Financiera y Carlos Fernández, que incluye a Lourdes Fernández como la beneficiaria pero no como participante en el acuerdo).

Atributo: IDFR AC

Un IDENTIFICADOR DE ACUERDO identifica en forma única a un caso de ACUERDO.

Atributo: DESC

Una DESCRIPCIÓN identifica a una explicación textual o comentarios en forma libre sobre un caso de ACUERDO.

Entidad: AC-PD

Un Subtipo de ACUERDO. Un ACUERDO DE PRODUCTO es un ACUERDO que regula el uso, la venta o el intercambio de PRODUCTOS entre PARTICIPANTES. Por ejemplo, el ACUERDO DE PRODUCTO N° 347 es para establecer la relación entre un ACUERDO de Cuenta y un PRODUCTO.

Atributo: IDFR AC

Un IDENTIFICADOR DE ACUERDO identifica en forma única a un caso de ACUERDO DE PRODUCTO

Entidad: ADPTCN AC

Un subtipo de CLASIFICACION. Una ADAPTACION DE UN ACUERDO es una CLASIFICACION que clasifica a un ACUERDO dependiendo si ha sido adaptado especialmente o establecido con características estándares. Por ejemplo, Acuerdo Adaptado y Acuerdo Estándar.

Atributo: IDFR CL

Un IDENTIFICADOR DE CLASIFICACION identifica en forma única a un caso de CLASIFICACION.

Atributo: TIPO DE CLASE

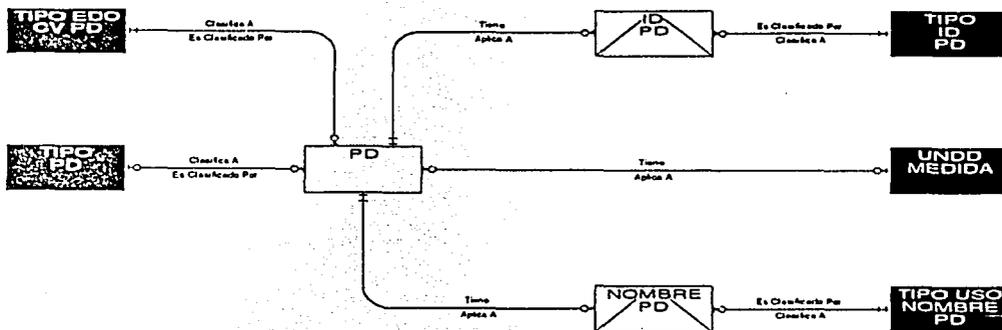
Un TIPO DE CLASIFICACION clasifica a una ADAPTACION DEL ACUERDO en virtud de si es estándar o adaptado

Dominio:

AC ADAPTADO, AC ESTANDAR

Entidad: AFCN INDV

Una entidad asociativa entre INDIVIDUO y TIPO DE AFICION DEL INDIVIDUO. Una AFICION DEL INDIVIDUO identifica una de las aficciones específicas de un INDIVIDUO. Por ejemplo, el INDIVIDUO N° 2453 (Marlo Lozano) tiene el TIPO DE AFICION DEL INDIVIDUO N° 674 (Fútbol) con fecha efectiva Julio 18 de 1986. Las ocurrencias históricas se mantienen.



• Entidad: PD

Definición: Una entidad fundamental y una entidad focal. Un PRODUCTO identifica bienes y servicios que pueden ser ofrecidos, vendidos o comprados por la Institución Financiera, su competencia y otros PARTICIPANTES, o en los que la Institución Financiera tiene interés durante el curso normal de su actividad de negocios con una UNIDAD de MEDIDA. Por ejemplo, el PRODUCTO N° 220 (Préstamo Hipotecario Residencial de Bancomer), PRODUCTO N° 98 (Cuenta de Cheques Personal), PRODUCTO N° 440 (Servicio de Negociación de Valores).

Atributo: IDFR PD

Un IDENTIFICADOR DE PRODUCTO identifica en forma única a un caso de PRODUCTO.

Atributo: DESC

Una DESCRIPCION identifica a una explicación textual o comentarios en forma libre sobre un caso de PRODUCTO.

Entidad: PD CONTNDO SEG

Definición: Una Entidad Asociativa entre PRODUCTO y SEGMENTO. Una RELACION PRODUCTO CONTENIDO EN SEGMENTO identifica al conjunto específico de productos que se encuentran agrupados en un SEGMENTO. Por ejemplo, el PRODUCTO N° 888 (contrato de cuenta maestra) está contenido en el SEGMENTO N° 3 (Depositos Vista).

Atributo: IDFR PD CONTNDO SEG

Un IDENTIFICADOR DE PRODUCTO CONTENIDO EN SEGMENTO identifica a un caso de PRODUCTO que se encuentra en un SEGMENTO específico.

Entidad: PLAN INDV

Definición: Una entidad asociativa entre INDIVIDUO y TIPO DE PLAN DEL INDIVIDUO. Un PLAN DEL INDIVIDUO identifica a uno de los planes específicos de un INDIVIDUO. Por ejemplo, el INDIVIDUO N° 777 (Jorge López) tiene el TIPO DE PLAN DEL INDIVIDUO N° 24 (Realizar un viaje de negocios) con fecha efectiva Agosto 4 de 1995. Las ocurrencias históricas se mantienen.

Atributo: IDFR PLAN INDV

Un IDENTIFICADOR DE PLAN DEL INDIVIDUO identifica en forma única a un caso de PLAN DEL INDIVIDUO.

Atributo: FECHA EFCTVA

Una FECHA EFECTIVA identifica a la fecha calendario en la que un PLAN DEL INDIVIDUO se vuelve válido.

Entidad: PLAN ORG

Definición: Una entidad asociativa entre ORGANIZACION y TIPO DE PLAN DE LA ORGANIZACION. Un PLAN DE LA ORGANIZACION identifica a uno de los planes específicos de una ORGANIZACION. Por ejemplo, la ORGANIZACION N° 2313 (Miscelánea La Lupita) tiene el TIPO DE PLAN DE LA ORGANIZACION N° 8 (Diversificar su mercado) con fecha efectiva Mayo 13 de 1996. Las ocurrencias históricas se mantienen.

Atributo: IDFR PLAN ORG

Un IDENTIFICADOR DE PLAN DE LA ORGANIZACION identifica en forma única a un caso de PLAN DE LA ORGANIZACION.

Atributo: FECHA EFCTVA

Una FECHA EFECTIVA identifica a la fecha calendario en la que un PLAN DE LA ORGANIZACION se vuelve válido.

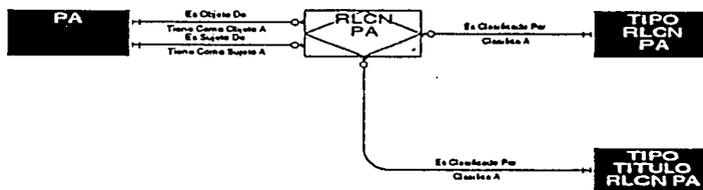
Entidad: RLCN AC

Definición: Una Entidad Asociativa entre dos casos de ACUERDO y TIPO DE RELACION ACUERDO, y una Entidad Focal. Una RELACION ACUERDO identifica a un caso de un ACUERDO relacionado con otro ACUERDO y la naturaleza de cada una de esas relaciones. Por ejemplo, el ACUERDO N° 123 (acuerdo de designación de beneficiarios) Está Incluido En el ACUERDO N° 345 (acuerdo de fideicomiso A). Las ocurrencias históricas se mantienen.

Atributo: IDFR RLCN AC

Un IDENTIFICADOR DE RELACION ACUERDO identifica en forma única a un caso de RELACION ACUERDO.

RELACION PARTICIPANTE



(Abogado) con fecha efectiva Junio 2 de 1975 y también tiene el TIPO DE OCUPACION DEL INDIVIDUO N° 48 (Contador) con fecha efectiva Mayo 11 de 1989. Las ocurrencias históricas se mantienen.

Atributo: IDFR OCPCN INDV

Un IDENTIFICADOR DE OCUPACION DEL INDIVIDUO identifica en forma única a un caso de OCUPACION DEL INDIVIDUO.

Atributo: FECHA EFCTVA

Una FECHA EFECTIVA identifica a la fecha calendario en la que una OCUPACION DEL INDIVIDUO se vuelve válida.

Entidad: OFERTA DIVISAS

Definición: Un subtipo de OFERTA DEL MERCADO FINANCIERO. Una OFERTA DE DIVISAS es una OFERTA DEL MERCADO FINANCIERO que implica a un instrumento financiero en la forma de divisas. Por ejemplo, dólares de los EU, dólares australianos, yen japonés.

Atributo: IDFR OFERTA DIVISAS

Un IDENTIFICADOR DE OFERTA DE DIVISAS identifica en forma única a un caso de OFERTA DE DIVISAS

Entidad: OMF

Definición: Una OFERTA DE MERCADO FINANCIERO es un Objeto del Negocio que describe a un PRODUCTO que es cualquier instrumento financiero o ACUERDO para un instrumento financiero, que está disponible en el mercado financiero a través de una institución financiera u otros conductos. Por ejemplo, divisas, commodities, acciones, bonos, etc. La OFERTA DE MERCADO FINANCIERO representa a un catálogo de instrumentos financieros y acuerdos de instrumentos financieros disponibles. La compra y venta de las ofertas del mercado financiero (o sus partes) se representa como un Servicio de Trading. Un caso de OFERTA DE MERCADO FINANCIERO que ha sido comprada se representa como un RECURSO. Por ejemplo, una oferta de acciones de la compañía IBM emitidas el 10 de Mayo de 1995; una oferta de commodities para jugo de naranja; una oferta de divisas en dólares norteamericanos; una oferta de derechos de un acuerdo de derechos minerales.

Atributo: IDFR OMF

Un IDENTIFICADOR DE OFERTA DE MERCADO FINANCIERO identifica en forma única a un caso de OFERTA DE MERCADO FINANCIERO.

Entidad: ORG

Definición: Un subtipo de PARTICIPANTE y una entidad focal. Una ORGANIZACION es un PARTICIPANTE de naturaleza colectiva, que generalmente está formado por varios individuos o grupos enlazados por finalidades y actividades en común, en un contexto operativo o legal. Por ejemplo, la ORGANIZACION N° 224 (Banco A); la ORGANIZACION N° 523 (Corporación B); la ORGANIZACION N° 926 (Asociación C); la ORGANIZACION N° 553 (Club D).

Atributo: IDFR PA

Un IDENTIFICADOR DE PARTICIPANTE identifica en forma única a un caso de ORGANIZACION.

Atributo: NUMERO EMPLD

Un NUMERO DE EMPLEADOS identifica la cantidad de empleados que tiene una ORGANIZACION.

Atributo: NUMERO OFCNS

Un NUMERO DE OFICINAS identifica la cantidad de oficinas que tiene una ORGANIZACION.

Entidad: PA

Definición: Una entidad fundamental y una entidad focal. Un PARTICIPANTE representa a un individuo u organización sobre el cual el Grupo Financiero Bancomer quiere mantener información. Por ejemplo, el PARTICIPANTE N° 124 (Daniel Hernández); el PARTICIPANTE N° 432 (Institución Financiera XYZ).

Atributo: IDFR PA

Un IDENTIFICADOR DE PARTICIPANTE identifica en forma única a un caso de PARTICIPANTE.

Atributo: FECHA EFCTVA
Una FECHA EFECTIVA Identifica a la fecha calendario en la que una RELACION ACUERDO se vuelve válida.

Atributo: FECHA EXP
Una FECHA DE EXPIRACION Identifica a la fecha calendario en la que una RELACION ACUERDO se da por terminada o deja de ser válida.

Entidad: RLCN AC PA

Definición: Una Entidad Asociativa entre ACUERDO y PARTICIPANTE y una entidad focal. Una RELACION ACUERDO PARTICIPANTE Identifica una relación específica entre un PARTICIPANTE y un ACUERDO y su tipo. Por ejemplo, el PARTICIPANTE N° 123 (Oscar Torres) tiene una RELACION ACUERDO PARTICIPANTE con el ACUERDO N° 6002 (Valores Plazo Fijo) del TIPO DE RELACION ACUERDO PARTICIPANTE N° 46 (Cliente Propio) con fecha efectiva Noviembre 27 de 1994.

Atributo: IDFR RLCN AC PA
Un IDENTIFICADOR DE RELACION ACUERDO PARTICIPANTE Identifica en forma única a un caso de RELACION ACUERDO PARTICIPANTE.

Atributo: FECHA EFCTVA
Una FECHA EFECTIVA Identifica a la fecha calendario en la que una RELACION ACUERDO PARTICIPANTE se vuelve válida.

Atributo: FECHA EXP
Una FECHA DE EXPIRACION Identifica a la fecha calendario en la que una RELACION ACUERDO PARTICIPANTE deja de ser válida.

Entidad: RLCN PA

Definición: Una entidad asociativa entre dos casos de PARTICIPANTE y una entidad focal. Una RELACION ENTRE PARTICIPANTES Identifica una relación específica entre dos PARTICIPANTES y su tipo. Por ejemplo, el PARTICIPANTE N° 4632 (Alfredo Mejía) tiene una RELACION ENTRE PARTICIPANTES con el PARTICIPANTE N° 2392 (Institución Financiera ABC) del TIPO DE RELACION ENTRE PARTICIPANTES N° 6 (El PARTICIPANTE Es Cliente Del PARTICIPANTE) con fecha efectiva Agosto 17 de 1994 y también tiene una RELACION ENTRE PARTICIPANTES con el PARTICIPANTE N° 118 (Banco Z) del TIPO DE RELACION ENTRE PARTICIPANTES N° 8 (El PARTICIPANTE Es Empleado Del PARTICIPANTE) con fecha efectiva Marzo 23 de 1977 y con el TIPO DE TITULO DE LA RELACION ENTRE PARTICIPANTES N° 111 (Director General Adjunto). Las ocurrencias históricas se mantienen.

Atributo: IDFR RLCN PA
Un IDENTIFICADOR DE RELACION ENTRE PARTICIPANTES Identifica en forma única a un caso de RELACION ENTRE PARTICIPANTES.

Atributo: FECHA EFCTVA
Una FECHA EFECTIVA Identifica a la fecha calendario en la que una RELACION ENTRE PARTICIPANTES se vuelve válida.

Entidad: SALDO CUENTA

Definición: Una entidad atributiva de la CUENTA y una entidad focal. Un SALDO DE LA CUENTA Identifica a un caso de un saldo monetario o no monetario asociado con una CUENTA. Por ejemplo, un SALDO DE LA CUENTA N° 9873-465 Identifica al 'Valor Acarreado', para la cuenta de cheques de Pedro Gómez (ACUERDO N° 343-43, al que le hace seguimiento la CUENTA N° 28733) es identificado en el SALDO DE LA CUENTA N° 87487256. Las ocurrencias históricas se mantienen. El requerimiento de datos representado por esta entidad se puede derivar pero ha sido incluido debido a su aplicabilidad a nivel de toda la empresa. Cada caso de CUENTA tendrá por lo menos un SALDO DE LA CUENTA asociado a él. Sin embargo, para CUENTAS recientemente establecidas puede que todavía no haya un saldo. Por estarazón, existe una relación opcional entre la CUENTA y el SALDO DE LA CUENTA. Si la intención es de poner el saldo en "ceros" para las nuevas CUENTAS, la relación puede modelarse como obligatoria.

Atributo: IDFR SDO CUENTA
Un IDENTIFICADOR DE SALDO DE CUENTA Identifica en forma única un caso de SALDO DE CUENTA.

Atributo: FECHA EFCTVA
Una FECHA EFECTIVA Identifica a la fecha calendario en la que un SALDO DE LA CUENTA se vuelve válido.

Anexo 5.- Diseño Físico del segmento de Cheques en el Data Warehouse

